UNIVERSITY OF BERGAMO

School of Doctoral Studies

Doctoral Degree in 'Transcultural Studies in the Humanities'

36th Cycle

SSD: L-LIN/10 Letteratura Inglese, L-LIN/12 Lingua Inglese

Digital Sorrows:

the spaces of terrorist violence through the lens

of Computational Literary Studies

Supervisors:

Prof. Rossana Bonadei (University of Bergamo)

Prof. Stefania Maria Maci (University of Bergamo)

Doctoral Thesis

Simone ABBIATI

Matricola n. 1073080

ACADEMIC YEAR 2022/2023

Abstract

This thesis delves into the convergence of literature and computer science, probing how algorithms can interact with nuanced literary texts. Concentrating on fiction related to The Troubles, the ethno-nationalist conflict in Northern Ireland, it strives to establish a theoretical framework for computational literary criticism (CLC) and subsequently applies it practically. Stressing the paramount role of human interpretation, it advocates for collaboration between computational and non-computational literary studies. Through case studies, it highlights the constraints of algorithms in grasping subtle meanings, and asserts that while AI excels in data extraction, human interpretation remains vital, particularly in ethical and political contexts within diverse computational social sciences domains.

Chapter III 128

Appendixes 156

Bibliography   229

Abstract

This thesis delves into the convergence of literature and computer science, probing how algorithms can interact with nuanced literary texts. Concentrating on fiction related to The Troubles, the ethno-nationalist conflict in Northern Ireland, it strives to establish a theoretical framework for computational literary criticism (CLC) and subsequently applies it practically. Stressing the paramount role of human interpretation, it advocates for collaboration between computational and non-computational literary studies. Through case studies, it highlights the constraints of algorithms in grasping subtle meanings, and asserts that while AI excels in data extraction, human interpretation remains vital, particularly in ethical and political contexts within diverse computational social sciences domains.

Introduction

This thesis explores the intersection of two seemingly irreconcilable worlds. On one hand, it delves into the realm of literature, characterized by the subjectivity of interpretation (Bleich 2019), where intertextuality (Kristeva 1980) and a plurality of meanings prevail. On the other hand, it ventures into the field of computer sciences, governed by a type of computational thinking (Wing 2008) that mandates the conversion of everything into binary codes, leaving little room for ambiguity.

How can these two worlds coexist? How can an algorithm, with its inherent need for very strong structuring, read and process something that is based on subjectivity? And, above all, why teach a computer to read literature if the essence of literature is to provoke ethical and educational reflection on the self and others, essentially serving as a call for ethics (Todorov 2007)?

> Literature can do a lot. It can reach out to us when we are deeply depressed, guide us towards the human beings around us, help us better understand the world, and assist us in living. It doesn't intend to be a way to heal the spirit; nonetheless, as a revelation of the world, it can also, as we go along, profoundly transform us. (Todorov 2007, p. 52)

To address questions like these, the thesis necessitated acquiring skills from both fields, especially computer science. This journey began with fundamental text mining skills and culminated in utilizing large language models that have been developed and popularized in recent months, such as BERT (Devlin 2019), Lama (Touvron, 2023) or Mistral (Jiang, 2023) – a proficiency the author initially lacked due to their background in spatial semiotics studies.

Secondly, upon acquiring the appropriate tools, one must ponder the means of fostering a dialogue between these two domains. Addressing the question, «How can an algorithm comprehend a literary text?» – which is synonymous with asking, «How does an algorithm distinguish itself from humans in interpreting a literary text?» – another pivotal motivation guiding the composition of this thesis is an interest in literature and

literary criticism that is politically engaged. This perspective, rather than isolating itself in an ivory tower and speculating almost 'teleologically' (Moretti 2000, p. 57) on a few canonical texts, critically evaluates the present or the very recent past through the medium of literature.

By virtue of this second driving force, a literature that explores recent pain was chosen as a case study: fiction representing ethno-nationalist conflict in Northern Ireland, known as The Troubles (Darby 1995). Although the conflict was formally resolved by the Good Friday Agreement of 1998 (Tannam 2001), the tension continued to have profound repercussions. Recently, it returned to the news due to Brexit (Gormley-Heenan & Aughey 2017), a political shift that reignited discussions about the border between Northern Ireland and the United Kingdom, as the latter was no longer part of the European Union.

The decision to opt for literature recounting The Troubles, beyond a personal interest in creating a political space for reflection, arises from the recognition that if an algorithm is fed a type of literature written to sublimate, reconcile, debate, and provoke readers regarding a still-painful historical discord, the disparity between algorithms and human subjectivity becomes even more apparent. The need to address how to guide an algorithm becomes more urgent due to the sensitivity of the matter. An algorithm lacks ethical sensitivity and empathy, or, more importantly, cannot take a political stance. Essentially, with the data it extracts from the text, it can do nothing more than present information. It is only through a human subject who can interpret the communicative context that the algorithm's output can contribute to reestablishing the ethical function that literature holds.

Also, the debate surrounding the collaboration or antagonism between algorithms and human subjectivity has become even more pronounced in recent months, as computer science, particularly artificial intelligence, continues to gain momentum. The ongoing development of technologies such as large language models (Naveed et al. 2023), some of which are alleged to possess superhuman intelligence (Bubeck et al. 2023), prompts society to question the potential for machines to acquire not necessarily consciousness (Floridi 2023) but a certain level of intelligence that could enable them to

replace humans in various tasks (Russell 2021), including those related to the military, such as drones and even nuclear deterrence (Kissinger et al. 2021).

The central question is whether there will be anything exclusively reserved for human prerogative, a question that was the focal point of the entire fourth Reith Lectures in 2021 on 'Artificial Intelligence,' where Professor Stuart Russell pondered a post-Artificial General Intelligence world where the toil of work would be supplanted by leisure activities requiring new practices and passions.

The question, «How can an algorithm 'read' literature that depicts recent grief?» is fundamentally theoretical. Therefore, we need to address it from a theoretical standpoint. However, the primary challenge encountered in this thesis is the absence of a comprehensive theory in the field known as computational literary criticism (CLC), leading to difficulties in examining it from a theoretical perspective. CLC spans various disciplines and is a part of the broader and diverse domain of digital humanities (DH). Without a general theory of reference, it becomes impossible to respond to a theoretical question that seeks to address the issue at a broader level.

For this reason, in the first chapter, we attempted to formulate an initial outline of a comprehensive theory for CLC. After assessing the problem, identifying the existing gap, and recognizing the necessity for a theory in the field of computational literary studies, two references were employed to shape the theoretical framework.

The first reference is Propp's *Morphology of the Folktale* (1968). Indeed, when the semiotician wrote his classic, he was in a state similar to that in which computational literary studies find themselves now. Just as was the case with folktales in Propp's time, numerous studies nowadays claim to be part of CLC, while it remains challenging to address this field due to the diverse characteristics of each individual study. A lack of studies attempting to elucidate the commonalities among different scientific communities led us to formulate a comprehensive tentative theory. In this light, Propp's methodology seemed fundamental to us: first, we initiated the identification of a representative corpus of studies that could be defined as computational literary criticism; then, we delved into the extensive analysis of a small sample of papers to attempt to elucidate theoretical commonalities. As will be seen from the first chapter,

the thesis argues how computational literary studies consider five conceptual steps, each time respecting differences in geographic origin or methodologies used.

However, presenting and discussing this initial theoretical structuring of the field of Computational Literary Studies can be particularly challenging to follow. Therefore, it was deemed necessary to include a visual component. We attempted to find a way to visually represent the five theoretical steps that computational literary criticism studies undertake when analyzing one or more texts. While exploring a mode of spatiotemporal representation that also considers the sequential order of the various conceptual steps, we looked to theoretical physics. This leads us to the second reference in the first chapter: Hermann Minkowski.

Minkowski (1908), indeed, builds upon Einstein's theory of the inherent connection between time and space. He suggests a schematization that enables the visualization of the three dimensions of space combined with the dimension of time. The graph, by highlighting certain aspects of Einstein's space-time concept, depicts in three dimensions what would exist in four dimensions. In adapting this visualization to our tentative comprehensive theory for CLC, the first chapter introduces a visualization that is akin to Minkowski's but also exhibits idiosyncrasies.

The aim of this visualization is not to add further complexity to the thesis, but rather the opposite. In fact, one of the thesis's most pivotal arguments is that the human element in interpretation holds paramount importance in computational literary studies. This significance is not diminished by the presence of an algorithm; on the contrary, the thesis asserts that the human aspect is the most crucial, as machines need to be guided, given that biases cannot simply be cast aside (Bode 2020). When contemplating the theoretical steps that scholars undertake during CLC, it becomes apparent that the algorithmic aspect is just one part of the process, while the majority of it remains within the realm of human prerogative.

The first chapter addresses the question of how an algorithm can theoretically engage with literature. It establishes a conceptual framework by outlining the five steps involved in these interpretive processes. And if the majority of these steps belong to the

interpreting subject, the risk of objectifying interpretation is mitigated – a concern occasionally raised in the context of Digital Humanities (Dobson 2019).

With this assumption in mind, we move on to chapters two and three, delving into a case study that revolves around fiction recounting the troubles in Northern Ireland. The objective is to tackle the same question that was initially approached from a theoretical standpoint in chapter one: «Where does the human element of interpretation come into play when an algorithm 'reads' a literary text?». This involves navigating the intricacies of determining what information to process and how to do so, particularly when dealing with sensitive data. This process begins with a comprehensive examination of a broader corpus and subsequently narrows down the focus to a singular literary work, employing a concentric approach.

In line with Westphal (2011), we have chosen to prioritize the narratological category of space as a crucial entry point into the texts. This decision is based on the concept that a narratological category serves as an interpretive tool, similar to others, but acknowledging that, in the context of a political case closely linked to territorialization,[1] such a category appears particularly relevant.

Chapter Two begins with a compilation of novels that delve into the geographical landscapes of Northern Ireland, seeking to depict the widespread sectarian violence in the region. This compilation consists of 41 novels published between 1968, marking the canonical onset of the Troubles, and the present day. It initially provides a historical overview of the Northern Irish situation, subsequently clarifying the spatial characteristics under examination and outlining the methodologies employed to analyze the texts.

Methodologically, while endeavoring to instruct machines to process the corpus in a respectful manner, we have opted to prioritize the narratives of individuals who have been victims of sectarian violence. It is essential to clarify that by 'victims,' we are not

---

[1] 'Territorialization' refers to the process by which individuals, groups, or institutions establish, claim, or exert control over a specific geographic area. It involves the demarcation and organization of space, often tied to social, cultural, political, or economic factors. Territorialization can be a complex and dynamic phenomenon, shaping the interactions and relationships within a given space. In case of interest, please refer to Vanier (2009).

exclusively referring to individuals who lost their lives due to the violence inflicted by terrorist groups. Rather, we also include those who indirectly encountered the violence as relatives of the deceased.

To understand how literature uniquely portrays these spaces and fosters an idiosyncratic form of reflection, distinct from the reflections offered by the actual testimonies of victims, two separate corpora were established. The first corpus comprises the genuine testimonies of victims of sectarian violence, while the second corpus encompasses the 41 novels centered on The Troubles. In the first corpus, efforts were made to discern the patterns through which grief is articulated, utilizing large language models such as GloVe (Pennington et al. 2014). In the second corpus, operating under the assumption that spatiality in literary texts carries semiotic significance (Lotman 1975), we chose to explore how the expression of grief, mirroring patterns found in victim's testimonies, intersects with the portrayal of fictional spatiality.

Consequently, this chapter not only provides fresh insights into how literature addressing the Northern Irish conflict approaches spatiality and deviates from the spatial structuring evident in victim's accounts but also illustrates how, in practice, much of the computational literary criticism process relies on human guidance. Humans, ultimately entrusted with the responsibility of adopting a synthetic perspective, manage to contemplate how the intersection of patterns expressing pain and other forms of spatiality fits into an overarching semiotic system.

Lastly, the third and final chapter focuses on a single novel – *Resurrection Man* by Eoin McNamee – in an attempt to understand whether a computational literary criticism process assigns a different role to the human interpretive element when the corpus consists solely of a single literary text. In this case, as McNamee's novel centers on the infamous case of the Shankill Butchers (Dillon 1999), which serves as the inspiration for McNamee's work, an attempt is made to elucidate the particularities of the literary piece by selecting two corpora. The first corpus is journalistic in nature, focusing on the news case of the Shankill Butchers, while the second corpus encompasses only the novel itself.

In the context of Chapter Three, artificial intelligence algorithms, such as Named Entity Recognition (Nadeau & Sekine 2007), were utilized to extract spatial landmarks and itineraries from these two corpora. The underlying belief is that, when integrated into literary works, the cognitive aspects of mentally reconstructed maps gain symbolic significance (Ferguson & Hegarty 1974).

This symbolic value, particularly in literature addressing contentious political issues that demand empathy and political awareness, is exactly the kind of meaning that the algorithm finds challenging to capture. In this context, the third chapter starts by examining the discussion of the Shunkill butchers' events within journalistic debates and then explores literature, navigating what we would like to term as 'absences of fiction': those subtle and implicit meanings that the interpreting subject can grasp by collaborating with the data supplied by the algorithm. Conversely, the algorithm cannot discern these meanings because they are not explicitly stated and, as a result, are not processable.

The result of this thesis is an argument for collaboration between computational literary studies and non-computational literary theory. Indeed, while algorithms excel in data extraction, the act of interpretation remains primarily within the domain of humans. Scholars engaging with non-computational theories not only have a significant role but also occupy the most crucial space in this field of study. Furthermore, as a point extending beyond the literary realm, this thesis reiterates that artificial *intelligence*, despite its semantically fascinating nature, does not imply the replacement of interpretive processes or the objectification of interpretation. This assertion holds true not only in literature, where ethical considerations may play a prominent role, but also in all the other fields of computational social sciences. In these diverse domains, algorithms, rather than assuming subjectivity, must always be guided by a human subject with a distinct ethical, political, and interpretive stance.

Chapter I

1.1 Lexical Ambiguities and Historical Perspectives

> 3.323   In everyday language it very frequently happens that the same word has different modes of signification—and so belongs to different symbols—or that two words that have different modes of signification are employed in propositions in what is superficially the same way. […]
>
> 3.324   In this way the most fundamental confusions are easily produced (the whole of philosophy is full of them).

Wittgenstein (1933, p. 18)

In his *Tractat*us Logico-Philosophicus Wittgenstein aims to is to provide a systematic account of the relationship between language, thought, and reality. Wittgenstein argues that language and the world are structured in a similar way, and that the logical structure of language mirrors the logical structure of reality. According to him, the purpose of language is to represent reality, and the meaning of language is determined by its correspondence to the world. In such a context, he posits that the building blocks of language are «atomic propositions», which correspond to states of affairs in the world. These atomic propositions can be combined to form more complex propositions, and the truth value of these propositions is determined by their correspondence to reality (Wittgenstein 1933).

Although we accept that «in everyday language it very frequently happens that the same word has different modes of signification», when expressions point to vague signifieds (Saussure 1966), problems arise. This is certainly the case with the term 'Computational Literary Criticism' and its many related terms, including 'Distant Reading' coined by Moretti (2005, 2013), 'Macroanalysis' introduced by Jockers (2013), 'Algorithmic criticism' presented by Ramsay (2003), 'Culturomics' as put forth by Michel (2011), 'Cultural analytics' introduced by Manovich (2020), 'Telescopic

reading' as employed by Algee-Hewitt (forthcoming) and the more general term 'Digital Humanities'[2]. Together, these terms have become an expansive umbrella used to encompass any methodology that uses computation to analyze also literary texts in various ways and to different degrees. Although these terms may not all carry the same connotation,[3] they are frequently used interchangeably within the field of Digital Humanities.

As it is relatively straightforward to determine which papers and studies fall under the purview of Computational Literary Criticism (CLC) based on a set of methodological perspectives, when it comes to a comprehensive theory, the term encompasses such a broad range of theoretical approaches that it is challenging to unite them under a singular framework. While other well-known literary theories have foundational texts, such as Freud's *The Interpretation of Dreams* for Psychoanalytic theory, Cleanth Brooks' *The Well Wrought Urn: Studies in the Structure of Poetry* for New Criticism, or Foucault's *The Order of Things* for post-structuralism, CLC have emerged from a domain – Digital Humanities – that is methodological by definition. Rather than being rooted in a specific theory, Computational Literary Criticism has

---

[2] In discussing the challenge of defining the scope of Digital Humanities, a useful reference can be found in the three-volume collection *Debates in the Digital Humanities*, edited by Gold (2013, 2016, 2019). The essays in this collection address key debates and questions, including the relationship between technology and traditional humanities methods, the role of digital tools in scholarly research and communication, and the social, ethical, and political implications of digital humanities work. These essays also grapple with the thorny issue of defining what exactly constitutes the Digital Humanities.

[3] The term 'Distant Reading' was coined to describe a method of literary analysis that involves analyzing vast corpora of texts, either with or without computational tools, in contrast to close reading, which involves analyzing individual texts in detail (Moretti 2005, 2013). 'Macroanalysis' refers to a similar approach to literary analysis but with a specific focus on scale, rather than explicitly opposing a critical tradition (Jockers 2013). While 'Algorithmic criticism' encompasses a comparable approach which places a greater emphasis on the iterative nature of this type of literary analysis (Ramsay 2003). 'Cultural analytics' describes a broader field of research that combines computational methods with cultural analysis to investigate cultural phenomena in a broad sense, including literature, art, film, and music (Michel 2011).

spread across a multitude of methodological approaches that researchers have only later reflected upon.

Although notable recent attempts have been made to frame a general framework and to address the lack of theory (Kleymann, 2022), there are still few studies confronting the problem of devising a comprehensive theory. It is not surprising that Moretti, one of the most prominent theorists in the field, has called multiple times for a general theoretical definition various times (Moretti, 2017, p. 213; Moretti, 2018, p. 113; Moretti, 2020, p. 12).

In fact, computational literary scholars have explored the concept of 'distant reading,' centring on the idea that computation involves a sense of detachment that arises from the fact that computers process (often large) corpora rather than individual literary works. However, the original concept of 'distant reading', as introduced by Moretti, differed from its current usage. Moretti used the term to describe a reading method that began with the identification of shared characteristics among literary works (Moretti 2005) that could be represented in an abstract manner, without the aid of a computer. His definition focused on abstraction and on visualization. Nonetheless, in contemporary times, most scholars employ the expression 'distant reading' as a synonymous reference to computer-assisted analysis. This development represents a significant deviation from Moretti's initial notion, not exempt from theoretical implications, and unfortunately, this distinction has gone almost unnoticed within the field.

With the exceptions of notable papers addressing terminological shifts and historical change of 'distant reading', such as Underwood (2017), in general terms we witness to digital humanists discussing the equivalence of close and distant reading (Bode 2018) or asserting that distant reading is not distant (Drucker 2017), implying that the differences between close and distant reading are less significant than their similarities. How did we arrive at this point when Moretti, who was famously dubbed «the professor urging his colleagues to stop reading books» by The New York Times, emphasized the stark contrast between the two literary approaches of distant and close reading (Schuessler

2017)? How did (literary) digital humanists come to speak of equivalence and non-distance, considering the initial opposition?

In this quite controversial context, this chapter aims to contribute to the discussion by attempting to redefine the term Computational Literary Criticism (CLC) and frame its significance – that is to say, sketching a first comprehensive theory for CLC.

One could ask, why do we need a theory in the first place, when Anderson (2008) invokes the end of theory and argues that correlation without causation is enough in the era of social networks?[4] Indeed, Anderson interestingly states that in the age of big data, correlation should be enough: we do not need to look for the 'whys' because correlative patterns based on billions and billions of examples are enough to predict general social behaviors (Anderson 2008). This situation eventually allows social media to profit well from digital surveillance, extracting material that enables them to track us even before birth (Barassi 2020). We indeed think that causation is still essential to formulate a theory, and theory is that foundation that helps defend the validity and the coherence of an approach, as formulated by Murphy and Douglas (1985).

The field of CLC has been quite critiqued from its beginning to its recent evolutions. Da (2018) argues that CLC lacks a solid theoretical foundation and focuses too much on technicalities. Others have criticized the field for neglecting cultural or gender perspectives (Liu 2012, Klein 2020). Some claim that it fails to capture the intricacies and nuances of literature (Fish 2012, Conrad 2014, Bottiroli 2018), while others accuse

---

[4] «This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves» (Anderson, 2008, p. 1): in a society inundated with vast quantities of data and fueled by cutting-edge computational methods, conventional theoretical frameworks in the social sciences may lose their significance. Anderson hypothesises a shift away from depending on established theories to explain human behavior, advocating for the adoption of data-driven methodologies that enable us to meticulously observe, track, and measure behaviors with unprecedented precision. In his opinion, due to the amount of data, correlation could be enough to identify human behaviors without lingering on theorizing a causation reason.

it of risking objectification by separating methodological and interpretive processes by reductively paralleling the social sciences with the humanities (Dobson 2019). Some have even claimed that the Digital Humanities approach «leaves out that contradictory and negating quality of what is normally called 'thinking'» (Brennan 2017). However, navigating this diversity of criticism can be challenging without a comprehensive theory, as it is arduous to identify a singular paradigm that is being scrutinized. Among others, we argue that there is a need for a theory. But how might we begin to develop such a theory?

θεωρός, «spectator». As it is often the case, etymology can provide us with a clue. Aristotle believed that knowledge could be acquired through practical experience and observation of the natural world. He placed great emphasis on empirical observation and argued that the senses were the starting point for all knowledge (Aristotle 1933). By carefully observing the world around us, we can identify patterns and regularities, and through this process of induction, we can develop general principles and theories. Therefore, instead of devising a prescriptive and aprioristic theory, the Aristotelian approach would be to take on the role of an observer. We must examine works and research conducted in the field and, from there, extract a form, imagine a structure.

The studies that would require our attention are in fact numerous. Despite facing strong criticism, the research field of CLC has firmly established itself and is continuing to expand. CLC experiments conducted in accordance with diverse principles is being published by esteemed publishing houses (MIT Press Cambridge University Press, Palgrave, Taylor & Francis, Springer, among many others), fosters the creation of new solely digital humanities-focused journals (such as *The journal of computational literary studies*, created in 2023), showcases its novelty in methods employed when published in established journals of other field, such as *New Literary History* or *Poetics Today*. Up to December 2023, there are several thousands of works in the field. Attempting to account for all of them would result in a confusing pursuit of perfectionism. As such, we argue that it could still be effective to narrow the scope of our research using three criteria: time-span, representativeness and geographical diversity.

Time-span could be limited to the post-Moretti era (if we might), as the field of Computational Literary Criticism gained significant resonance after the introduction of 'distant reading' and its formulaic definition in Moretti's famous essay *Conjectures on World Literature*. Consequently, we can limit ourselves to the last two decades, confident that recent works share common ground with their predecessors and do not appear as a lone rose in the desert.

On the other hand, the issue of representativeness presents further challenges. One possible approach to this task would be to start by examining papers and monographs that have undergone peer review in the most prestigious journals within the field, with the prestige of these journals being gauged by considering the citations of scholars who serve on the editorial board: *Digital Humanities Quarterly*, *Digital Scholarship in the Humanities*, *Le champ numérique*, *The Journal of Cultural Analytics*, and *The Journal of Computational Literary Studies*.

The list of CLC journals we have provided could be disputed in terms of its importance, with some scholars possibly contending other significant studies have been carried out outside of or have not been reviewed in these journals. This is indeed true, but if our objective is to adopt a paradigmatic perspective rather than a syntagmatic perspective, we could refute this possible criticism by stating that the studies published in the chosen journals stem from an intrinsic acknowledgement of other computational literary traditions.

Many philosophers have indeed discussed the correlation between research and awareness, consigning Aristotelian observation to a new complexity. Michel Foucault, for instance, asserted that knowledge is shaped by power dynamics and historical contingencies. According to Foucault, scientists must be aware of the discursive frameworks that mould their work to understand the boundaries and prospects of their research (Foucault 1976). In the case at hand, some significant influences could shape CLC itself, rather than omitting influential works. Kuhn (1962) contended that scientific research is not simply the accumulation of facts, but rather the outcome of a specific paradigm or worldview that shapes how scientists approach their work. Scientists are not impartial observers of nature, but rather are influenced by their cultural, historical,

and theoretical contexts). Similarly, Bourdieu (2004) also claimed that scientists are not impartial observers but are situated within a specific field of power dynamics and cultural capital, and must be aware of the social and cultural context of their work to generate meaningful and effective research. Subscribing the idea that science functions within a discursive framework, we can presume that starting from the work published in esteemed journals considered the official voices of the widest organization of Digital Humanities (the *Alliance of Digital Humanities Organization*s (ADHO)[5] – except for *the Journal of Cultural Analytics* and *The Journal of Computational Literary Studies*, whose importance we believe is justified by the fact that its editorial board includes some of the most highly cited scholars in the field – would assure us that those papers and monographs embody other works that might otherwise escape our attention. By adopting this perspective, we can view these resources as a starting point that not only avoids omitting vital research issues but also helps us access branches that we cannot directly reach.

By applying these first two criteria - time-span and representativeness - we have selected approximately 250 critical works[6]. From this extensive list, we further narrow it down to fifty studies, choosing them on the basis of the third criterion that we consider crucial: geographical diversity.

Castell (2000) argues that in today's interconnected society definable as a space of flows, where global economy is characterized by the almost instantaneous flux and exchange of information, capital, and cultural communication that order and condition both consumption and production, knowledge is exchanged freely regardless of geographical location. The consideration of this third filtering criterion may not thus be deemed necessary. Though cultural capital is not directly tied to geographic distance, there are numerous social and cultural factors related to geographic location that can

---

[5] See: https://adho.org/publications/ [retrived on 31 July 2023]

[6] The full list of this archive is provided in the appendix of this chapter, with a geographical reference referring to the country of employment of each another. In this way the reader could check the geographical provenance.

influence the acquisition and distribution of cultural capital, especially in the case of CLC.

Individuals hailing from regions plagued by poverty or lacking access to educational and cultural resources are often deprived of opportunities to foster cultural capital compared to those who grow up in wealthy areas with more abundant resources. Likewise, those dwelling in rural areas or small towns may have little exposure to cultural institutions like museums, theaters, and art galleries, which are pivotal sources of cultural capital (Harvey 1973). Also, geographic location intersects with other factors such as race, ethnicity, and social class that can influence the accumulation and distribution of cultural capital. For instance, individuals from marginalized communities may face additional barriers to accessing educational and cultural resources, which can limit their ability to accumulate cultural capital (Lareau 2003).

More importantly, in CLC, digital infrastructure indeed plays a pivotal role, encompassing foundational electrical infrastructure that needs to be in place, as well as cutting-edge High-Performance Computing capabilities. It is no coincidence that Stanford's LitLab, situated in the heart of Silicon Valley, has garnered significant acclaim for its work in CLC. Undoubtedly, the rapid technological advancements in this region, facilitated by the dynamic interplay and competition amongst the Big Five[7] and the staggeringly fast start-up environment (Kenney 2000, Etzkowitz 2022), have contributed to this success.

After considering these factors, we believe that filtering by geographical location could prove advantageous in balancing cultural disparities and highlighting theoretical differences. This filter would also prevent the unjust attribution of innovative research solely to those with greater cultural power, influenced by external factors rather than the clarity of theoretical reflection itself. If our goal is to abstract a universal theory and give significance to the substantial, theoretically refined voices that have emerged regardless of the provenance from prestigious institutions, it is mandatory to consider

---

[7] 'GAFAM' is an acronym commonly used to refer to the five largest technology companies in the Western IT, which are Amazon, Apple, Facebook, Google (Alphabet), and Microsoft. For instance, please refer to de Bustos & Izquierdo-Castillo (2019), or to Human & Cech (2021).

that geographical location could help address these inequalities. Our objective is here to recognize the value of *reflective abilities* within digital scholarship, where the infrastructure allows for it.

Following these criteria, we now can count on our foundation: an archive of CLC works to be observed with an Aristotelian approach. The question is: what to do then with them?

1.2. At the crossroad of disciplines: methods and materials

We have identified a deficiency in the theoretical foundation of CLC, discussed the importance of establishing a solid theoretical framework, and deliberated on an initial pool of studies to extract common conceptual steps. Moving forward, we aim to highlight the two primary references that underpin our endeavor to outline a comprehensive theory: Propp and his contributions to folktales, and Minkowski for his space-time theory. Despite belonging to disparate disciplines – the former in narratology and the latter in theoretical physics – we will now explore how these two references can serve as the groundwork for our conceptual framework.

1.2.1 Propp's narratological approach

> How is one to explain the similarity of the tale about the frog queen in Russia, Germany, France, India, in America among the Indians, and in New Zealand, when the contact of peoples cannot be proven historically? This resemblance cannot be explained if we have wrong conceptions of its character […]. We see, then, that very much depends upon the study of forms. We shall not refuse to take upon ourselves the crude, analytical, somewhat laborious task which is further complicated by the fact that it is undertaken from the viewpoint of abstract, formal problems. Such crude, 'uninteresting' work of this kind is a way to generalize 'interesting' constructions'.
>
> <div align="right">Propp (1968, p. 14)</div>

Propp's intent in writing *Morphology of the Folktale* was to distill a set of fundamental principles upon which folktales are built given the absence of a comprehensive framework for understanding folktales. In order to tackle this issue, Propp adopted a structuralist approach to decipher the shared elements and roles found in traditional Russian folktales. His analysis was based on the notion that there exists an inherent structure, or 'morphology,' that underlies all folk narratives. Consequently, he identified recurring narrative functions and character archetypes within these tales. He illustrated the need for this analytical approach by drawing an analogy with language:

> Is it possible to speak about the life of a language without knowing anything about the parts of speech, i.e., about certain groups of words arranged according to the laws of their changes? A living language is a concrete fact—grammar is its abstract substratum (Propp 1968, p. 15).

Folktales thus necessitate a morphology to serve as a useful tool for in-depth analysis and exploration of their content. Similarly, to gain a comprehensive understanding of language evolution, researchers require a framework of grammar.

When we initially delved into the *Morphology of the Folktale*, we observed that it started from the fact that the field of folktale analysis had seen numerous studies but lacked a comprehensive theory. This state of affairs actually mirrored the current situation in CLC. Similar to what was happening with folktale studies at the time, there are now thousands of academic papers worldwide falling under the umbrella term of 'distant reading.' Each of these papers employ diverse methodologies and have different scopes and objectives, yet the absence of a unifying theoretical framework persists. Consequently, it can sometimes be challenging to grasp the true scientific significance of these studies or understand why they share similar terminology. This led us to wonder: Is there a method for abstracting the common steps employed in CLC? Is it feasible, similar to Propp's approach, to construct a grammar or a theoretical framework for this field?

However, the citation with which we begin this paragraph contains a problematic excerpt. It highlights the importance of structure but also asks, «How is one to explain

the similarity of the tale when the contact of peoples cannot be proven historically?» Propp defends the validity of this approach by stating that folktales demonstrate a similar structure even though the authors are not proven to have any contact. This implies that there must be an underlying structure that emerges regardless of the origin of the folktale.

For our thesis, this point poses a problem. As we have already pointed out, in a globalized world, Digital Humanities communities have had more than a few interactions. We cannot claim that unrelated papers from communities that have never had contact share conceptual steps, and argue that these conceptual steps are proven to be foundational solely because the DH communities have not had any contact.

On the contrary, CLC assumed a global resonance starting from the Anglophone environment, or to put it differently, it was the cultural power of this academia that strongly nominalized a concept that already existed, and then disseminated those nominalizations to other countries. Indeed, in those papers we analyzed, many of the indicated seminal references are from Moretti's theoretical work or to other Anglophone studies such as Jerome McGann's work at the Institute for Advanced Technology in the Humanities (McGann 2004b) at University of Virginia in the late 1990s, with some notable exceptions such as Roberto Busa's work and his *Index Thomisticus* (IBM Data Processing Division 1973). As Isabel Galina pointed out relatively recently (Galina 2014), the dominance of English in the domain of Digital Humanities adds another layer of complexity to the reasons why CLC seem to be more abundant in the Anglophone world.

Putting these intricacies aside, we argue that it is still useful to study how the same concept has changed accordingly to the environment it has traveled through, regardless of where it was initially established.

We have, therefore, selected fifty works based on the criteria outlined in the previous subchapter (time-span, representativeness, geographical diversity).[8] For each of these studies, we have identified five fundamental conceptual steps that we believe might constitute the core of a general theory. These steps are shared across different studies to varying degrees. According to Propp's work, where the functions of *dramatic personae* are initially introduced as a general category, followed by a specific account of how they manifested in reference folktales, the present paragraphs will follow a similar structure: we will first introduce each conceptual step and then provide a concise description of how it was executed in the respective work. Each paragraph will eventually include a broader comment on that conceptual step.

Since we will be dealing with 50 reference papers, each requiring an introduction of five conceptual steps, comprehending these intricacies may pose a challenge. In order to facilitate a better understanding of our theory, we have opted to heavily incorporate visualisation. As we need to find method of visually representing the five steps shared among all the reference papers, allowing for both comprehension and the ability to comment on their order and significance, we sought a visualisation that could consider both spatial and temporal aspects. Consequently, we turned to physics and discovered that Minkowskian spacetime theory (1908) could offer a valuable framework. We decided to adapt his theory and utilise his space-time visualisation to represent the five steps of CLC. To grasp the initial outline of his theory and comprehend how we modified it, it is essential to momentarily shift our focus and delve into the theoretical foundation of Minkowskian spacetime. Subsequently, we will discuss the aspects we have incorporated or omitted from this theory.

---

[8] The list of the papers is provided in the appendix at the end of the chapter. The selected geographical provenance, listed in alphabetical order, includes: Algeria, Argentina, Australia, Austria, Belgium, Brazil, Cameroon, Canada, Chile, China, Colombia, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Hong Kong, India, Indonesia, Iran, Ireland, Israel, Italy, Japan, Kazakhstan, Kyrgyzstan, Latvia, Lebanon, Malaysia, Mexico, Netherlands, New Zealand, Niger, California, Peru, Poland, Portugal, Russia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Turkey, United Arab Emirates, the United Kingdom, and the United States of America.

1.2.2. A Metaphor from Physics: Adapting Minkowski's Spacetime to CLC

In the late 1800s, theoretical physics was debating the inconsistency between Newton's Laws of Motion and Maxwell's equations, which describe electricity and electromagnetism (Stachel 1998, pp. 10-14). The discrepancy centered around the speed of light. Maxwell demonstrated that the speed of light is a wave propagating in space. The changing magnetic field results in a changing electric field, ultimately leading to a self-propagating magnetic field. Moreover, Maxwell's theory predicted the speed of light to be approximately 300,000 kilometers per second. This raised a question: what would be the measured speed of light if the observer was moving with the Earth's spin? According to Newton, the moving observer should measure a different speed than someone who is stationary, but Maxwell's equation indicated that light has only one speed, which was named 'C.'

Maxwell's equations did not account for a speed of light that deviated from the prediction of the equations. At that time, nearly all scientists (including Maxwell himself) believed that Newton's laws were correct, and they considered Maxwell's equations to be incomplete. This conundrum persisted for almost a decade until Albert Einstein made his breakthrough: by revising Newton's laws of motion, in 1905, Einstein postulated that the speed of light remained constant regardless of the observer's reference frame (Einstein 1905), thus confirming Maxwell's equations. Einstein demonstrated that observers moving at different speeds would perceive the distance and time of two events differently, revealing that time and space are relative to the observer. This idea, the core of the theory of relativity, suggests that the measurement of time and space depends on the person taking the measurement. The crucial point is that the speed of light remains constant for everyone. Herman Minkowski, Einstein's former professor, later realised that the theory of relativity was about the geometrical relationship between space and time. In 1908, he famously stated:

> Henceforth space by itself, and time by itself, are doomed to fade away
> into mere shadows, and only a kind of union of the two will preserve an
> independent reality. (Minkowski 1909, p. 1)

In such a way, Minkowski introduced the concept of a four-dimensional spacetime, where three dimensions represent space and the fourth represents time. In this framework, events are described by four coordinates—three for space and one for time. In other words, Minkowski's insight was that the union of space and time into a single, four-dimensional continuum was more fundamental than treating them separately. He argued that the distinction between space and time, as separate entities, was artificial and that they are intricately interconnected in a unified structure, thus laying the mathematical foundation for Einstein's general theory of relativity, which describes gravity as the curvature of spacetime.

Four dimensions describe an event in the universe: three for space and one for time. But how can we visualize those four dimensions if we are accustomed to thinking in three dimensions? To visually simplify the concept of spacetime, Minkowski, instead of proposing a spacetime diagram in a four-dimensional representation, proposed a three-dimensional spacetime diagram, comprising two spatial dimensions and one time dimension from where he eliminated one spatial dimension.

This spacetime diagram had time depicted on the vertical axis and the two dimensions of space on the horizontal axes (Stein 1968). However, there was a significant challenge since time is measured in seconds, which is different from space. To tackle this problem, Minkowski suggested that time should be represented as a length by multiplying it with the speed of light. For instance, if the speed of light is 300 million meters per second, then a meter is equal to three hundred millionths of a second. By expressing the time dimension in terms of the speed of light multiplied by time, or 'CT', Minkowski created a visually consistent representation with three coordinates (Fig. 1.1).
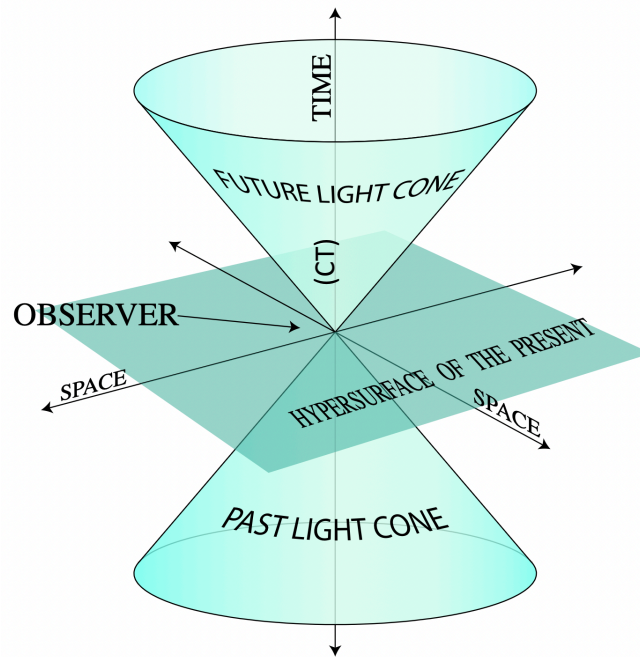
Figure 1.1. The concept of dividing Minkowski spacetime into four distinct sets originates from Sard (1970). These sets include the light cone, the absolute future, the absolute past, and a region referred to as 'elsewhere.'

Imagine that a light is emitted at a specific location in the expanse of spacetime, denoted as 'event A.' As time elapses, this light propagates in every direction, forming a cone-shaped structure known as the 'light cone.' Coined by Minkowski himself, the light cone encompasses all potential future events in spacetime that could be reached by the light emitted from event A, thus including all the events observable by a person placed in the spacetime position of event A.

Light cones can be utilized to represent any event in spacetime, visually demonstrating all the possible future events that could be observed from a specific point in space-time. Additionally, an inverted cone, known as the 'past light cone,' can display past events in spacetime that are connected to a particular event A. Any point located outside these two light cones would be considered causally disconnected from event A, meaning they cannot be reached by or influence event A .

Let's consider labeling the initial gravitational singularity preceding the universe's birth as 'event A.' and event A represents the moment when the universe came into existence. For the sake of simplicity, we will assume that Minkowski's spacetime remains constant, and event A represents the moment when the universe came into existence.[9] If event A were to occur during the spacetime singularity that preceded the birth of space and time, it would not be possible to represent the past light cone. This is because before that singularity neither space nor time existed (Russell 1925). Moreover, as we have previously established, events that hypothetically lie outside the light cones cannot be reached by event A. Therefore, if event A was the first moment in spacetime to ever exist, it follows that every event that has ever occurred in the universe should lie within that cone (fig 1.2):
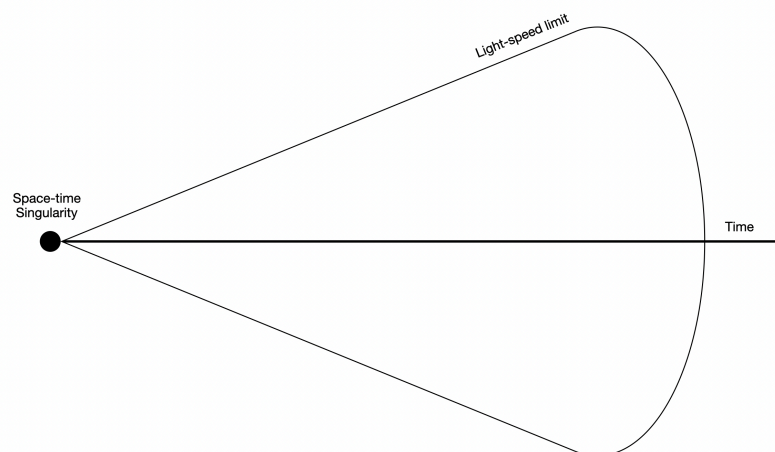
Figure 1.2. A Minkowsky-inspired visualization where the reference event (in other words, the so-called 'event A' where the observer is located) represents the space-time singularity, which marks the first moment of the universe. The only existing cone would be the future light-cone, encompassing all the events reachable by event A, which is to say all the possible events.

---

[9] Minkowski's spacetime representation serves as an abstract tool in theoretical physics, visualizing the relationship between space and time in a single graph, but it has its limitations as it focuses on specific aspects of Relativity and does not account for the bending of spacetime caused by gravitational singularities. Indeed, placing event A directly into the spacetime singularity preceding the Big Bang would require non-Euclidean geometry (Hawking & Penrose 1970, 2010)

If we accept the premise that this abstract visualization could encompass all the events in spacetime that occurred after the spacetime singularity (which we can visualize as points included in the future cone of our Mirkoskian visualization), then every act of CLC could be represented on it, as it is also a set of events in spacetime. Regardless of complexity and specificity, any example of CLC could be represented – albeit with a great deal of abstraction – on our spacetime, considering it as a set of events in a possible universe.

As the reader will hopefully understand, we chose to utilize an adapted version of Minkowski's spacetime representation to comprehensively visualize all the conceptual steps involved in virtually every act of Computational Literary Criticism (CLC). This adaptation enables the creation of a schema that systematically illustrates and comments on all potential acts of CLC. By employing this visual framework, it becomes evident that the step of interpretation is always present, while the application of software constitutes just one of the five critical steps. This schema effectively challenges the perception that CLC is merely an objective and soulless application of software. In fact, four other steps prominently highlight the human prerogative, and even the software application step is meticulously executed by human scholars. Therefore, this visualization underscores the integral role of human involvement throughout the entire CLC process, emphasizing the interpretative and subjective dimensions inherent in the field of Digital Humanities and Computational Literary Criticism.

We are now fully prepared to delve into the aforementioned five conceptual steps that we believe are fundamental within the realm of CLC.

1.4. Visualizing the act of Computational Literary Criticism

After a comprehensive analysis of our elected fifty reference studies, we have identified five key conceptual steps that literary digital humanists follow when conducting computational literary analysis. We will explain these steps in an analogous manner, first by introducing each step and visualizing it on our introduced adapted Minkowskian space-time. Secondly, we will discuss how each of the reference papers

addresses the step. Finally, we will attempt to identify higher hierarchical categories by grouping similar examples together to discuss general patterns.

To navigate through this section effectively, we recommend that readers have Appendix 3 readily available. Given the complexity of presenting an in-depth analysis of the fifty selected studies within the main text, we opted to include a table in the appendix. This table delves into the comprehensive examination of each study, highlighting the application of the five conceptual steps. It's important to note that this table is the outcome of an interpretive process rather than a mere compilation: the initial structure of the analysis did not follow a five-folded format; the five conceptual steps organically evolved during the study, and after completing the individual analyses for each paper, we made the decision to reorganize them, presenting the five conceptual steps for each study in a more structured manner.

As regards visualisation, for each conceptual step presented, we will add a differently colored section to our Minkowskian visualization. This involves representing a mathematical set of points (or events) in spacetime that share the same time coordinate but differ in their spatial position. The purpose behind this approach is to account for instances occurring within a short period, where events of the same type are closely spaced in time.

To clarify, let's consider the step involving the selection of a research question. For example, we may initially decide to investigate the dialogical voice of Elizabeth Bennet in *Pride and Prejudice* to explore the aspects concerning the role of women in 19th-century England. Moreover, upon studying some critical works, we might choose to broaden our focus and examine the entire character system, as the protagonist's voice becomes better understood when analyzed alongside the interactions of other characters in the narrative universe of the time. Such a decision could happen spontaneously within a short time span or even after months of contemplation. Despite the different timeframes, both moments share a similarity: they involve selecting a literary-related aspect to explore. These two points or events are closely related in time and can be represented together in the same section (fig 1.3).
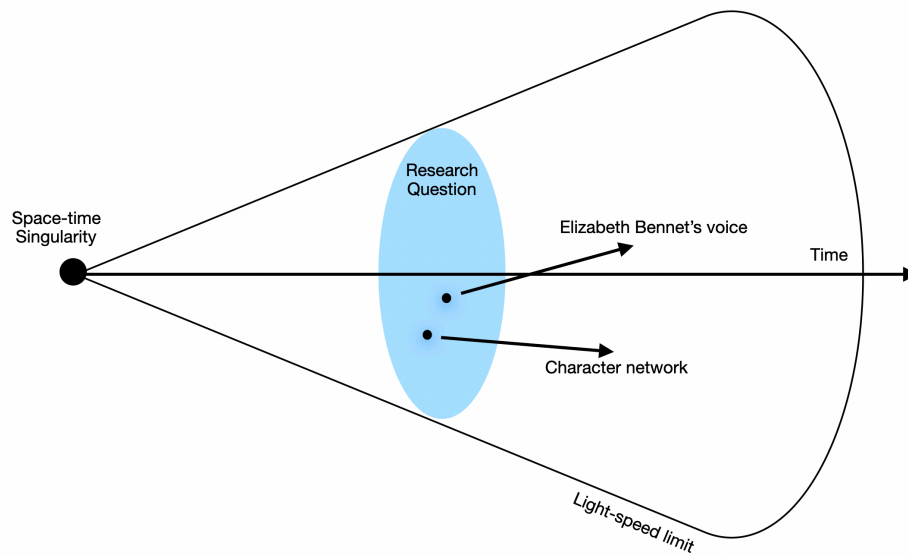
Figure 1.3. The visualisation of the steps involved in the selection of research questions on our Minkowskian space-time.

Mathematically speaking, the section should not be considered a Euclidean plane because the research questions addressed by literary analysis are not selected simultaneously. There may be a time gap of several days or more before this foundational part of scientific research is completed, specifically before the analysis begins. Consequently, this section should be viewed as a surface encompassing points/ events in spacetime with different spatial and slightly varying time coordinates. Just like we handled other mathematical complexities, for the purpose of representation, we approximate the time distance between events related to the same step as zero to simplify representation.

Now that we have described the visualization used, we need to discuss the method of presenting the data that we will adopt in this chapter. After selecting fifty reference papers from which to extract the conceptual steps fundamental to employing CLC, we include an explanation of where each specific step occurs among these fifty studies in appendix 3. Now, let's proceed to the first theoretical step in CLC.

1.4.1. Selection of (an) abstract concept(s)
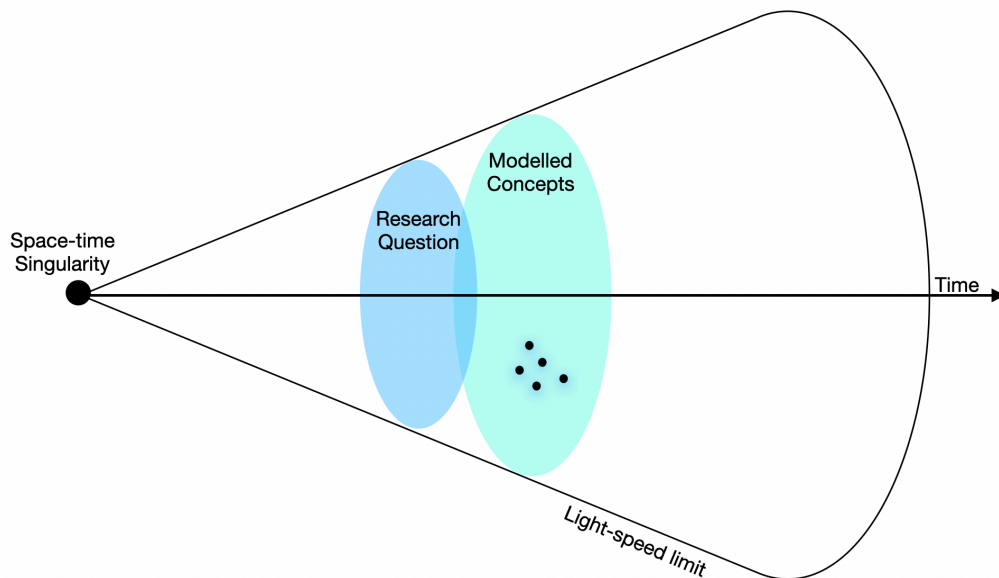
Figure 4. The visualisation of the step of category selection in Computational Literary Criticism.

In CLC, the initial step involves selecting one or more precise literary-related concepts or categories. This step is visually represented as a light blue section on our spacetime diagram (see fig. 1.4). Every critical analysis must consider foundational categories to work upon; however, in the context of Computational Literary Criticism, this stage becomes distinctive.

Here, the undefined and expansive literary realm intersects with the binary code of computers, capable of processing only zeros and ones, resulting in more than a banal interplay. As computers categorize things by labeling them with a yes/no answer, researchers face the challenge of fitting abstract concepts into strict binary labels, often requiring them to make significant interpretive choices.

This process gives rise to what has been termed 'computational thinking' (Wing 2008) – a problem-solving approach that draws upon principles and techniques used in computer sciences. It entails deconstructing intricate problems into smaller, manageable tasks and solving them by employing a sequence of logical steps and algorithms. In a nutshell, digital humanists working on literature examine fiction as an interplay between categories, endeavoring to dismantle them as one would untangle silicon cables.

Based on the selected exemplary studies, we could encounter various re-formulations of this step, involving the selection of different concepts to investigate or retrieve. Those concepts or categories can be grouped in the following way:

1. **Categories fundamental for a singular literary author**: concepts of 'space', 'doubt' and 'form' in Italo Calvino's literary production (Falcetto 2019); the 'Joyceannism' of the latter Flann O'Brien works (Queiroz et al. 2015); personification and metamorphosis in Ovid (McCarty 2003); the aesthetic awareness of the poet in Wordsworth and Yeats's poetry (Kusi 2022); the figure of the 'portal' in Michael Joyce, an American author of electronic literature (Ciccoricco 2021); Rosalía de Castro's bilingualism (Méndez 2020); the symbolic value of houseware nouns in the novel *De Sobremesa* by José Asunción Silva (Grillo Gálvez 2019);

2. **Categories specific to a literary genre**: timelines, theatrical space, relationships between dramatic characters in dramaturgical texts (Varela 2021); the representation of the wolf in the context of Portuguese contemporary literature (Queiroz et al. 2015); the representation of the temple in Jewish fiction (Brill et al. 2020); poetic prosody in Urdu-Hindi poetry (Rizvi et al. 2022); occurrence of technology-related topics (mention of AI, internet, social networks, robotics, among others) in contemporary fiction relating the post-digital age (Menéndez de la Cuesta González 2021); relationship between characters at the 'adegan' (scene) level in Javanese 'shadow puppetry' (Rokib et al. 2018); string similarity in finnic oral folk poetry (Janicki et al. 2023); different per media (print/digital) in

contemporary Czech poetry (Škrabal & Piorecký 2022); the symbolic meaning of love in Saadi's lyric poems (Mahmoudi & Abbasalizadeh 2019); conceptual metaphors related to love in mystical literature (Li et al. 2021); the relationships between text and pictures in digital poetry (Waliya 2018);

3. **Categories important for specific literary theories (regardless of the genre)**: the narrative voice concept (Howell et al 2014); relationships among literary genres (Kapitan & Wills 2023); the concept of scene as theorized by Genette (Zehe et al. 2021); plot schema in detective novels (Murai 2020); the «inexplicable of literature», conceived as its figurative meaning (Fan 2018); the concept of digital social reading and the interaction between user and literary works (Rebora et al. 2021); the concept of multiple narratological voices (Skorinkin & Orekhov 2023); relationships between characters and characters network (Varela 2021, Van Zaanen et al. 2020, Gervas 2009); concept of character (Fischer et al. 2018); lifespan of genres, influence of prestige and its relation to gender (Underwood 2019); plot structures (chronology, focalisation, characters' functions, among other things) (Gervas 2009);

4. **Categories concerning the field of humanities broadly conceived (linguistics, book history, history, economics, geography, codicology, cognitive studies, gender studies, among others)**: lexico-grammatical features to identify stylistic variations (Tian & Liu 2023, Janicki et al. 2023); linguistic diversity measured with type/token ratio, and most frequently used words (Varela 2021); keyword-in-contex and foreignisms (Škrabal & Piorecký 2022); rhymes and prosodic rhythms (Rizvi et al. 2022); the conceptualisation of conflict (Smeets et al. 2021); concepts from book economics, such as salary differences in book-selling jobs (Dealberto et al. 2022); the relationship between nationality and the publishing success (Bode 2018); geographic information such as toponymies, name of regions and countries (Hazlon et al. 2021, Heuser et al. 2016, Queiroz et al. 2015, Wrisley 2017); key semantic domains (anatomy and psychology, religion and the supernatural, etc.) (Can & Cangir 2022); frequency of morphological categories

(adverbs, nouns etc.) (Mourad 2021); differences in manuscripts' lectiones (Asanov et al. 2021); source and target domains of metaphors (Berglund et al. 2022); existential loneliness in different cultures (Rakymzhan et al. 2022); the construction of national consciousness and the formation of the concept of the nation-state in Greece (Koidaki et al. 2021); representation of gender, focusing on older people (Geybels 2023); the lack of Arab and Middle Eastern electronic literature coverage in academic research (Hosny 2018).

Several observations can be made regarding this first step of Computational Literary Criticism (CLC). First and foremost, the choice of concepts to investigate in Computational Literary Criticism (CLC) is closely tied to the theoretical approach and the researchers' interests. For instance, if one subscribes to the belief that literature is, and should remain, entirely detached from extra-literary factors, as argued by Benedetto Croce when he emphasized the autonomy of intuition, independent of historical context (Croce 1922), the selected concept for digital investigation would be limited to literature itself and its intrinsic forms. This approach would result in CLC becoming primarily a morphological study or a focused analysis of the structure of literary traditions.

Conversely, some researchers align themselves with different theoretical paradigms, such as cognitive studies, as exemplified by Fludernik (2002). This perspective heavily draws from Discourse Analysis and Cognitive Psychology, emphasizing the significance of narrative itself as a means of discursively making sense of various phenomena. If a digital humanist were to adopt and align with Fludernik's approach, it would choose, as abstract concepts, ideas that go beyond the intrinsic forms of literature. Similarly, researchers influenced by Marxist theory, rooted, for example, in Eagleton's approach (Eagleton 2006), focus on exploring literature's relationship with history, the complex interplay between 'form' and 'content' in literature, questions regarding literature's political engagement, and the role of production and technology. When dealing with algorithms, these scholars adopt a similar viewpoint and address these same concepts.

In other words, it is crucial not to equate the entire CLC framework with just one of its subfields and, based on that, reject the entire field, as some have done (Bottiroli

2018).[10] It is not surprising that if an approach to literary criticism is found disagreeable, regardless of the utilization of algorithms, it will also be found objectionable even if it serves as the foundation for a literary analysis conducted with algorithms. This means that each computational literary approach can be subject to criticism individually, without necessitating the outright rejection of the entire field of CLC due to misalignment with one's literary approach preferences.

To rephrase our previous statement, reflecting on the underlying theories upon which digital literary methodologies are constructed we should not focus on establishing the correct underlying methodology because, as has been said, there is none. Similarly, such reflection should avoid framing the investigation of computer-aided methodologies as a clear-cut conflict between 'traditional' and 'new' methodologies. As Julia Kristeva eloquently explored in her renowned work *Desire in Language*, «the science of literature is an always infinite discourse, an always open enunciation of a search for the laws of the practice known as literature» (Kristeva 1980, p. 95). Neither new methods nor emerging elements within the literary context are capable of stagnation or abruptly coming to a halt.

As a second observation regarding the first step of CLC, the selection of a theoretical approach, which influences the choice of concepts for investigation, involves

---

[10] «In the context of literary studies, the shortsighted love looking into the distance (*distant reading*): they squint, staring at the era to which the work belongs, and they make that blurred synthesis, which the text has become, into the emblem of a historical period.» (Bottiroli 2018, p 14). Bottiroli distances himself from distant reading by exclusively framing it within the Marxist tradition of computer-aided criticism, which is a result of the Italian reception of distant reading as closely associated with Franco Moretti and his distinct approach to computer-aided literary criticism. Bottiroli's perspective on literary theory diverges significantly from Moretti's, as evident in his work 'What is the Theory of Literature' (Bottiroli 2006), and in our view, this dissimilarity leads to the outright rejection of the entire field of distant reading by Bottiroli. However, it is crucial to acknowledge that Bottiroli may have overlooked the potential for integrating and synergizing his own theory, rooted in Lacan and Derrida, among others, with algorithms. Such an approach was for instance explored by Dickstein (2018) in his analysis of the connection between 'arithmocentrism' and the phallus, leading to a contemplation of possible ways in which computers might simulate the analytical experience.

discussions about the distinction between hard and soft sciences. Computational Literary Studies lie at the intersection of the humanities and computation. As a result, much of the research in CLC involves experimenting with formulas derived from mathematics (Ashley 2022) or economics (Sinykin and others 2019), among others, to apply them effectively to literary studies. Interestingly, several researchers begin by categorizing literature using classical literary categories before transitioning to computational methods to aid in extracting those categories from texts, focusing on what Berry (2011) discusses as «computationality,» which refers to the extent to which a system or process can be described, modeled, or understood in terms of computation. As a consequence, the incorporation of concepts from both hard sciences and soft sciences prompts a debate about what truly qualifies as literature. If we regard literature merely as another form of data, then employing complex mathematical formulas would present no issue. However, if we still view literature as a peculiar text, governed by its own set of rules, then utilizing formulas from other sciences without considering its literary particularities becomes problematic.

Thirdly, based on the investigated categories, the field presents itself as highly interdisciplinary, drawing concepts from a wide range of fields, including linguistics and codicology, in a more seamless manner compared to traditional literary approaches. The perception of interdisciplinarity as a strength or weakness depends on the academic culture and the dynamics between tradition and innovation (Frodeman & at, 2017). However, CLC tends to exhibit greater interdisciplinary qualities than other literary approaches, possibly due to its foundation in two disciplines from the outset – literary studies and computer sciences. This foundation allows CLC to fearlessly integrate different disciplines and explore with creativity across boundaries. Needless to say, the interdisciplinary nature of CLC is further exemplified by the collaborative nature of its papers, often written by multiple authors (McCarty 2012).

Fourthly, in most cases, CLC operates at the textual level, as the concepts being retrieved are primarily expressed through words. However, as we will explore in the following theoretical steps, there are a few exceptions to this pattern. This still raises doubts about whether CLC solely refers to Computational Linguistics, or Corpus

Linguistics, applied to literary texts. On the other hand, we firmly believe that CLC can also be defined as a form of literary criticism, especially when the objective is to collect data to reflect on literature or utilize it as a tool to gain insights into society. That's why we have included linguistic categories in the selected concepts. Within the framework of CLC, these categories are not examined in isolation; instead, they serve as symbolic means for something else.

Fifthly, when categories extracted from one or a few literary works are reshaped through computational thinking, it becomes crucial to formally define a concept. This process, referred to as operationalizing (Moretti 2013), holds inherent value because shaping a concept requires precise formulation. As one can imagine, the process of formalization and abstraction carries the risk of reductionism, and thinkers have always debated whether a specifically formalized ontology or concept can adequately symbolize something else.

Last but not least, the selected categories are always determined by humans. Even when utilizing the most algorithmically complex algorithms, as we will delve into in a later theoretical step introduced in this chapter, the human prompt remains fundamental. This holds true even in the case of so-called exploratory analyses or stylometric analyses, where cutting-edge algorithms automatically choose what to retrieve – often prioritizing the most linguistically frequent structures (Neal & et al. 2017). Attempting to eliminate the human prompt is, simply put, not feasible. This concept, while not extensively explored in algorithmic literary studies, is a well-known phenomenon in all other sciences related to automation, often referred to as 'algorithmic bias' (Kordzadeh & Ghasemaghaei, 2022). Indeed, while «algorithmic systems are used to automate decision-making processes or assist human decision-making by providing decision-makers with algorithmically generated information such as classification results and predictive analyses [...] black-boxed algorithms may pose ethical risks at different levels of organizations and society» (p. 388). Owing to the inherent biases present in all machine learning algorithms, researchers must carefully consider their societal implications. Likewise, when these very algorithms are employed in the realm of

literature, they bring along the same biases, necessitating thorough investigation and critical analysis.

In fact, according to the Oxford English Dictionary, an algorithm is a documented series of steps that leads to the transformation of data. These steps are conceived and taught to computers by humans, even when the software resulting from a programmer's work is applied to new data by a different user. From the simplest algorithms that follow basic rules to deep learning algorithms that rely on complex mathematical assumptions, the initial impetus always comes from a human. This is further emphasized by the fact that recent developments in computer sciences have focused on addressing societal problems stemming from the inherent assumptions that algorithms rely upon – biases and assumptions that programmers inadvertently incorporate into their work (Žliobaitė 2017; Nelson 2019; Ntoutsi 2020; Pessach & Shmueli 2023). These assumptions can trigger a butterfly effect, leading to situations so problematic that they necessitate accountability (Shah 2018; Horneber & Laumer 2023). We saw a prime example of this with the emergence of social networks that exhibited biases against black women (Buolamwini & Gebru 2018). When so-called artificial intelligences discriminate, who is responsible for the way these algorithms were authored? Copious research on ethical issues in AI highlights the undeniable fact that algorithms, from their inception, are reliant on human assumptions that must be uncovered. This principle holds true for all algorithms and AI-driven processes, and it remains equally relevant in Computational Literary Studies, where concepts are either directly selected by humans or chosen by machines based on human-made assumptions.

1.4.2. Literary objects to be investigated

After selecting a concept to explore, Computational Literary Criticism (CLC) involves choosing one or more literary object from which to extract and investigate relevant structures or concepts or recurrences. This step goes beyond analyzing only textual materials; it also includes other literature-related objects. Hence, as a crucial

step, we refer to this phase as the 'selection of literary objects for investigation', rather than 'corpus construction', as the word 'corpus' usually refers only to textual data.

In fact, an important characteristic of CLC lies here. We argue that within the domains of Computational Human Sciences and Data Mining, researchers often explore literary-related subjects, but this does not necessarily qualify their work as CLC. Certainly, if the object under investigation is connected to literature, it satisfies a 'necessary and sufficient' condition for the research to fall under the umbrella of CLC. However, this characteristic alone is insufficient. To be fully recognized as CLC research, the main purpose of the work should primarily concern literary subjects, rather than using fiction as a textual example. Many researchers use fiction as their object of study with the aim of gaining non literary insights – such as psychological (Mohseni, M. 2021) or linguistic insights (Vatri, A., & McGillivray, B. 2018), among many others. These aims are valid, but we argue that if the primary objective is not literary, then the study's alignment is different from the aim CLC has as a field of literary criticism. The definition of CLC as an interpretive act carried out with digital approaches, focusing on literary gains and to contribute to literary sciences, is crucial if we attempt to establish a general theory for CLC. This, in turn, leads to the development of a new set of tools for its exploration.

Moreover, since there can be one or multiple literary objects under investigation - ranging from a single novel to a extensive corpora, from a single paratextual object to an entire manuscript archive - this challenges the notion that Computational Literary Studies revolve solely around digital archives. While archival reflection is undoubtedly important (we will touch on this briefly at the end of the paragraph), when focusing on a single work, extensive archival considerations may be less pertinent. However, this does not diminish the fact that the research can still be regarded as an excellent example of CLC.

As a result, the section representing this step in our space-time can encompass anywhere from one to millions of points (see Figure 1.5) for each artistic event, whether it is a text or another type of media, selected within a limited time span and being examined in the current study.
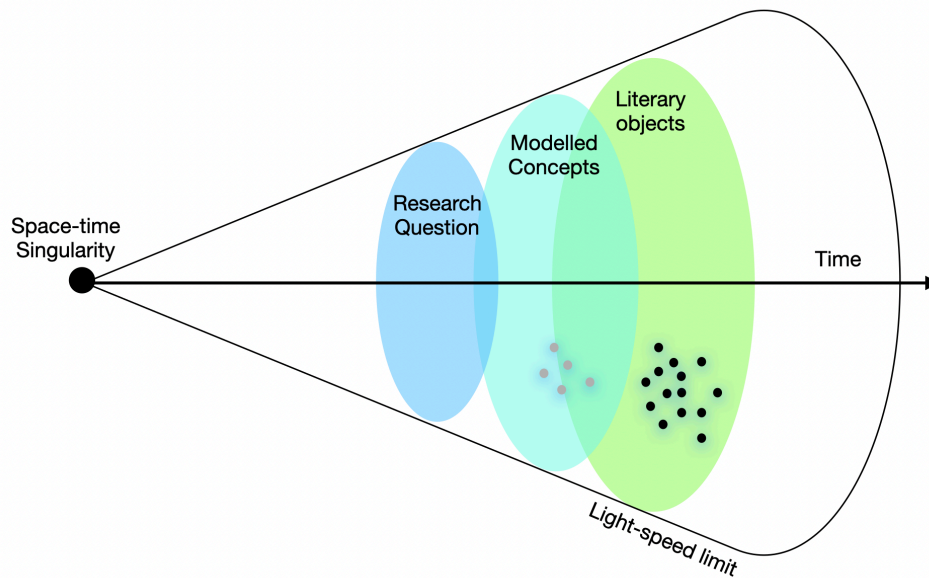
Figure 1.5. The visualisation of the step of Literary Objects selection in Computational Literary Criticism.

As we revisit the fifty works together to identify literary-related elements under investigation from our fifty reference works representing fifty different countries, we discover various typologies of objects being examined. These typologies can be categorized as follows:

1) **Singular literary work**: a single novel (Ciccoricco 2021, Grillo Gálvez 2019); an exemplary work of travel narrative (Uenishi et al. 2022); a dramatic text (Masías et al. 2017); a digital poem (Waliya 2018).

2) **Author-structured corpus**: Calvino's works (Falcetto 2019); Molière's comedies (Dealberto et al. 2022); poems by Pessoa (Skorinkin & Orekhov 2023); Amy Tan's novels (Tian & Liu 2023); Shakespeare's plays (Brill et al. 2020); works from Wordsworth and Yates (Kusi 2022); four Michael Joyce digital works (Ciccoricco 2021); poetic work from Rosalía de Castro (Méndez 2020); Saadi's lyrical poems (Mahmoudi 2019); the works of Charles Dickens

(Brandão & Frota 2017); the works by Bustan of Sa'adi and Divan of Hafiz (Li et al. 2021);

3) **Genre/tradition-structured corpus**: motion picture screenplays (Roberts-Smith et al 2013); 170 Dutch novels (Smeets et al. 2021); works of Irish Modernism (O'Sullivan et al. 2018); German-language dime novels (Zehe et al. 2021); comic stories (Murai 2020); Babylonian Talmud (Brill et al. 2020); works from different genres within the Medieval French tradition (Wrisley 2017); two novels dealing with the post-digital age (Menéndez de la Cuesta González 2021); 487 poems about the Great War (Can & Cangır 2022); 80 poems from 47 poets (Kim et al. 2021); three Afrikaans novels and dramas (Van Zaanen et al. 2020); 14 novels on existential loneliness (Rakymzhan et al. 2022); 91 modern Greek novels (Koidaki et al. 2021); 41 contemporary Dutch novels (Geybels 2023); the Electronic Literature Collection (Hosny 2018); 465 German-language dramas (Fischer et al. 2018); computationally-generated stories (Gervas 2009);

4) **Vast literary archives**: databases concerning Old Norse-Icelandic manuscripts with around 16,000 manuscripts (Kapitan & Wills 2023); 423 literary works including a variety of genres (Queiroz et al. 2015); novels, novellas, and short stories published in early Australian newspapers (Bode 2018); works from the Indonesian classic period including poetic and prosaic genres (Rokib et al. 2018); HathiTrust, Chicago Text Lab Corpus, and ECCO Corpus (Underwood 2019); 90,000 oral folk poems (Janicki et al. 2023); 442,000 contemporary Czech poems (Škrabal & Piorecký 2022); the Brown Corpus (Brandão & Frota 2017); the Swedish Language Bank (Borin et al. 2014); 400 poetry books (Berglund et al. 2022);

5) **Paratexts**: descriptions of manuscripts (Kapitan & Wills 2023); comments in the margin of stories (Rebora et al 2021); theatrical program booklets (Varela 2021).

6) **Objects related to literature**: questionnaires on literature (Rebora et al 2021); comments about reading platforms (Rebora et al 2021); book reviews (Rebora et al 2021); locations where a dramatic text was being performed (Varela 2021);

7) **Literary remediations**: movement of actors within drama (Varela 2021), literary translations (Asanov et al. 2021).

Several observations are here mandatory. First, the possibility for CLC to be applied to both individual works and vast corpora goes beyond the concept of 'distant reading' proposed by Moretti. While 'distant reading' was originally designed to address works that were not only overlooked but also impossible to read within a given canon (the 'great unread', as described by Cohen in 1999), CLC can effectively analyze individual literary works as well. In fact, Eve (2019) discussed the possibility of effectively investigating a single work by combining both telescopic and microscopic perspectives and by conducting a series of close reading exercises through computational methods. His approach neither disconnects the reader from the text nor separates the findings from mainstream literary criticism. Eve argues that by focusing on Mitchell's *Cloud Atlas*, computation can assist in analyzing earlier versions of the book, identifying semantic and syntactic alterations, and exploring genre and authorship through computational stylometry to uncover features that may not be readily discernible to the human eye. If we agree that critical works such as Eve's (2019) exemplify CLC, then critiquing the archive-focused approach of CLC should not result in the dismissal of the entire field. The question of proximity and distance in relation to literary texts sometimes appears to be crucial in assessing CLC. However, we believe that this framework needs further exploration and discussion.

In fact, Felski (2015, p. 52) poses the question: «Should we be close readers or distant readers? Dive in or draw back? [...] Critics contemplate the implications of proximity versus distance and ponder the merits of surface and depth». Felski

emphasizes the use of spatial metaphors in the literary debate to underscore the proximity or distance from a single literary work, essentially highlighting the choice between immersing oneself in a single literary work or exploring a broader corpus. Though, Dobson (2019) argues that, in order to move beyond the contemplation mentioned by Felski, humanists require a heightened methodological awareness (p. 53). He convincingly asserts that recognizing the significance of different viewpoints, along with their supporting evidence and potential arguments, is essential. As Dobson points out, the incorporation of novel computational techniques is pivotal in this discourse. These tools are poised to become an integral part of the digital researcher's toolkit, broadening our horizons and prompting relevant inquiries that traditional methodologies might overlook. However, in addition to an awareness of methods, understanding the potential scale of the investigated corpus is crucial for effectively evaluating computational methods. Recognizing that Computational Literary Criticism (CLC) can be both applied to single works and to massive corpora is essential in clarifying the debate.

A second observation we can make regarding the second step of CLC is that while distinguishing between genre/tradition-structured corpora and vast literary archives may initially appear unimportant, from a theoretical perspective, it is crucial. Differentiating between research focused on a limited, manageable corpus and those seeking to embrace the largest possible dataset, commonly known as big data (De Mauro et al. 2016), highlights a conceptual difference. In fact, if CLC is handling an extensive corpus, we enter an entirely different realm of literary reflection. When dealing with big data, the corpus/archive becomes an object that can only be processed, as intended by De Mauro's definition: «Big Data is the information asset characterized by such a high volume, velocity, and variety that it requires specific technology and analytical methods for its transformation into value» (p. 127). While reading fiction in snippets, conducting a full, in-depth study on a singular item becomes almost impossible, much like any other type of big data analysis (Gandomi & Haider 2014). This results in a detachment from individual literary works that might eliminate the interaction between a reader and a text, which Wolfgang Iser (1979) posited as the crucial aspect of the act of reading and

the formation of an aesthetic object that must be subjected to dynamic and unique interpretation. In other words, when selecting a literary object for investigation using a computational methodology, it necessarily follows that one must choose a type of analysis, along with its relevant questions and challenges.

Thirdly, the works included in our reference corpus span from ancient to contemporary texts, thereby demonstrating that there is no correlation between the contemporaneity of CLC approaches/tools and the texts under consideration. Although many Big Digital Projects involving CLC often prioritize classics such as Dante (Kumar 2021), Shakespeare (Estill 2019), Cervantes (Bia & Pedreño 2001), among others, partly for revitalization purposes and financial reasons, this does not preclude the examination of strictly contemporary works.

Fourthly, the selection of the corpus relights inquiries about heritage, memory construction, and canon formation, topics that have already been the subject of intensive discussion by scholars like Derrida (1995) and Foucault (1966). Derrida, in particular, begins his exploration by delving into the etymology of 'archive,' tracing it back to the Greek word 'ἀρχή,' which means 'principle' or 'commencement.' This etymology underscores the foundational nature of information storage: to archive is not just to record an event but to actively construct it, thereby endowing it with historical significance. This process is inherently political, as the act of archiving is intrinsically tied to historiography. Derrida draws a parallel between the Greek archons, who wielded authority and controlled knowledge, and the contemporary custodians of archives, stating that «the archons are first and foremost the guardians of documents» (1995: 9). In the modern era, the willingness to engage in archival practices, coupled with the remarkable capabilities offered by the digital revolution, especially in terms of storage and processing power, which continues to grow at an exponential rate (Hilbert & López 2011), reignites the vision of a Borgesian 'Library of Babel.' However, the distinction lies in the vast landscape of big data, where we amass data on an unprecedented scale. Here, the true power lies not in storing but in choosing what to erase, as deleting equals the power of indirectly choosing what remains and,

therefore, what creates the 'data' for any kind of computational analysis – as cleverly pointed out by Floridi:

> In history, the problem was what to save: which laws or names were going to be baked in clay or carved in stone, which texts were going to be handwritten on papyrus or vellum, which news items were worth printing on paper. In hyper-history, saving is the default option. *The problem becomes what to erase.* Since storage is insufficient, something must be deleted, rewritten, or never be recorded in the first place (Floridi 2014, p. 21).

Foucault, on the other hand, draws upon Freud's ideas to critique the work of historians and archivists. He connects this critique not only to the study of history but also to a broader examination of the human subject. Just as historians are not detached, self-transparent observers passively recording past events, the human subject is not a completely independent entity merely observing the realm of memory. Archiving is not a detached, mirror-like process of documentation; instead, it represents an *ordering of things* that results in the empowerment of a particular perspective - the dominance of a particular discourse - over others.

In other words, when working with author/genre/tradition-specific corpora or vast archives (meaning, when processing more than one literary work) for literary purposes, it's crucial to remember that nothing should be chosen for computational analysis unless it has been properly preserved and digitised. Therefore, the (digital) archiving process, especially in our age where these practices are closely linked to specific technological infrastructures, is both significantly influenced by cultural power dynamics and is an act on cultural power in itself, as the selection (inclusion or elimination) creates the foundation of analyses. Corpora and archives will inevitably exhibit biases, and we cannot simply eliminate these biases. The very act of selecting what to store is shaped by cultural and political factors (Bode 2020). Instead, we should acknowledge the structural impossibility of achieving complete impartiality from the beginning, address these issues, and be aware that the scope of our observations is constrained by the completeness of our archives.

One final observation we can make regarding this stage of CLC is that the analyzed literary objects encompass both textual and non-textual elements. This is particularly relevant not only for literary genres involving a visual component, such as graphic or illustrated novels, but also for multimodal literature in general. Multimodal literature, by its very definition, is hyper-textual and can include videos, images, sound, code, among other elements. These aspects need to be discussed within the framework of «multimodal cognitive poetics», as outlined by Gibbons (2012, p. 24).

As we can see, computation can be applied to various types of literary objects and their diverse media aspects, which ultimately requires the utilization of tools not only from Computational Linguistics and Corpus Linguistics but also from sound, video, and image processing.

1.4.3. Software application

The third step that characterizes the process of CLC involves the application of a computational tool. We represent this step visually in our space-time model by adding a yellow section (figure 7), which follows the preceding steps of concept operationalizing (light blue section) and the selection of the literary object to investigate (green section).
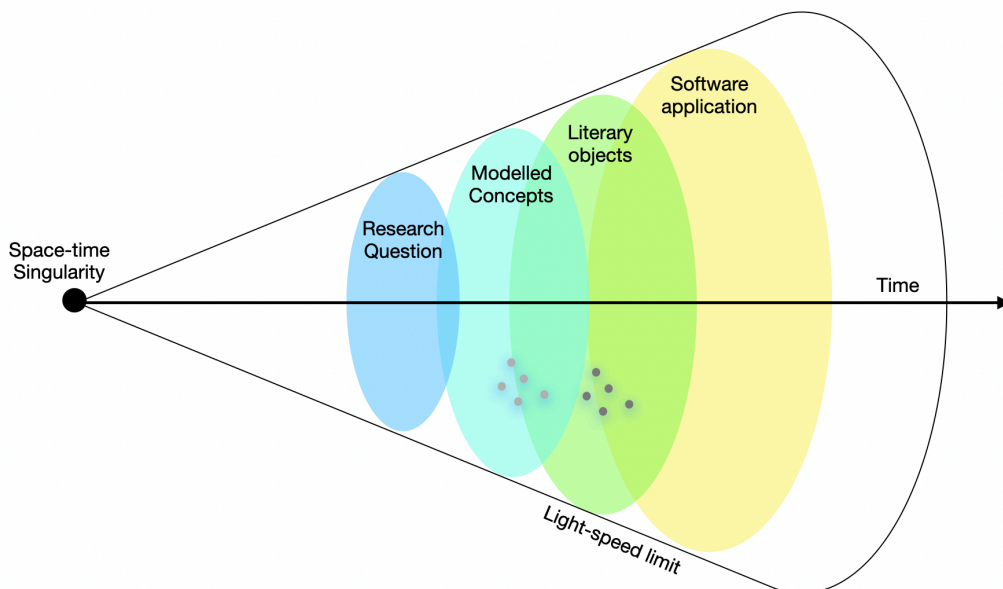
Figure 1.6. Visualisation of the step of Software application in Computational Literary Criticism.

From our analysis of the fifty referenced critical works, it becomes evident that, in most instances, software and tools are selected according to the research questions and the texts to be investigated. Moreover, software applications exhibit a wide spectrum of complexity, ranging from straightforward word retrieval to the utilization of highly advanced AI-oriented software.

Through an examination of these fifty studies, we can discern how each one leverages software applications. The resulting conclusions are as follows: it is evident that the role of software applications holds particular significance in CLC studies when compared to previous stages in the research process. These applications come into play at various stages, each with its unique level of complexity, and could be categorised according to the type of object being processed (whether it be the literary object before the actual analysis, the literary objects themselves, the database storing the literary objects, or the way of visualizing the literary insights):

1) **For data curation/preparation:** catalogues curation (Kapitan & Wills 2023); Optical Character Recognition (OCR) (Rizvi et al. 2022, Škrabal & Piorecký 2022, Van Zaanen et al. 2020, Hazlon et al. 2021, Geybels 2023); General preprocessing (lemmatization, stop word removal, among others) (almost all the studies);

2) **For literary object analysis** (text, image, sound, etc.), with two approaches:
   A. Human-oriented: We can define as human-oriented those software applications in which the emphasis is placed on the framework or ontology rather than on the efficacy of the algorithm. These strategies often involve manual annotation, utilizing various markup languages (e.g., XML, HTML), or platforms (Howell et al. 2014, Smeets et al. 2021, Falcetto 2019, Rizvi et al. 2022, Grillo Gálvez 2019, Škrabal & Piorecký 2022, Mourad 2021, Geybels 2023), with the possibility of feeding a machine learning model with examples only at a later stage (Zehe et al. 2021, Underwood 2019) or involving collaborative markup via a digital platform (Rizvi et al. 2022,

Menéndez de la Cuesta González 2021, Asanov et al. 2021, Rakymzhan et al. 2022).

B. Machine-oriented: On the other hand, we can define as machine-oriented those software applications where the focus is on the algorithm and that do not dwell much on devising an ontological framework. Examples include general statistical analyses (type-token/ratio, keywords in contexts, among others) (Varela 2021, Méndez 2020, Can & Cangır 2022, Li et. al 2021, Koidaki et al. 2021, Hosny 2018, Fischer et al. 2018); automatic parsing (Roberts-Smith et al 2013, Smeets et al. 2021, Uenishi et al. 2022, Kim et al. 2021, Geybels 2023) or Named Entity Recognition (Uenishi et al. 2022, Ciccoricco 2021, Van Zaanen et al. 2020, Heuser et al. 2016); network analysis to investigate relationships between investigated literary objects (Rebora et al 2021, Bode 2018, Masías et al. 2017, Fischer et al. 2018); keyword/lemma search based on an ontology (Queiroz et al. 2015, Rokib et al. 2018, Mourad 2021); unsupervised Artificial Intelligence approaches, such as Topic Modeling or Decision Trees, or stylometry (Zehe et al. 2021, Bode 2018, Varela 2021, Underwood 2019, O'Sullivan et al. 2018, Rebora et al 2021, Skorinkin & Orekhov 2023, Kusi 2022, Brandão & Frota 2017, Koidaki et al. 2021); sentiment analysis (Rebora et al 2021, Heuser et al. 2016); computing mathematical-linguistic differences after transforming words into vectors (Brill et al. 2020, Janicki et al. 2023, Berglund et al. 2022); gazetteer's approach to finding locations (Wrisley 2017); visual analytics (Varela 2021); supervised language generation based on a narratological hypothesis (Gervas 2009).

3) **At the database level:** Machine learning processes on human-extracted data (Smeets et al. 2021, Brandão & Frota 2017); machine learning processes on automatically extracted data (Masías et al. 2017); relational databases (Fan 2018, McCarty 2003); mathematical operations performed on database-structured data, such as dimensionality reduction (Murai 2020, Varela 2021,

Škrabal & Piorecký 2022);

4) **For visualization:** Network visualization (with Gephi or other tools) (Kapitan & Wills 2023, Ciccoricco 2021, Van Zaanen et al. 2020); creative visualizations drawing from information design or other sciences (Falcetto 2019, Dealberto et al. 2022); georeferencing locations with GIS or cartographic systems (Uenishi et al. 2022, Wrisley 2017, Varela 2021, Méndez 2020, Borin et al. 2014, Hazlon et al. 2021); using classical types of visualization software to visualize the data (graphs, charts, etc.) (Bode 2018, Van Zaanen et al. 2020, Grillo Gálvez 2019, Menéndez de la Cuesta González 202113, Li et. al 2021, Hosny 2018);

To begin with, machines can play a crucial role in the digitization and organization of materials for analysis. Their primary objectives are to preserve these materials and facilitate data handling through algorithms. However, this process also highlights a significant imbalance in terms of digital infrastructure and power distribution. Recently, Automatic Optical Character Recognition (OCR) has become predominantly reliant on Artificial Intelligence (AI) software. These software solutions are extensively trained on millions of examples and can establish connections between character representations and their digital counterparts (Chaudhuri 2017). It's worth noting that languages with extensive access to digital infrastructure and computing power have reached a point where their software has been trained on vast amounts of data, making them very efficient (Mittal & Garg, 2020). In contrast, endangered languages struggle to find sufficient materials for training their software, perpetuating an asymmetry. The lack of language data results in the inability to effectively train machine learning models (Nathan 2011; Conathan 2011), and consequently, in the inability to computationally process cultural artifacts.

The same principle applies to digitizing various material objects: the more funding available, the better the quality of machine digitization, as a better machine can achieve higher graphical resolution with expensive tools, and having more people involved in a

digitization team can significantly expedite the process. Clearly, these issues involve a set of ethical considerations that have always been important in the archiving process. However, they need to be reevaluated in light of new digitization activities and technologies, new funding models, the possibility of sharing and manipulating digital content, and online engagement with heritage items in a global digital environment. In this context, a single digitized object can theoretically be shared and reproduced without limits, exponentially amplifying cultural disparities in power (Manžuch 2017).

Secondly, literary objects can undergo analysis, either emphasizing or contrasting with human capabilities. Human-oriented methods often rely on digital tools or platforms for markup and annotation. In these cases, human researchers perform literary analysis, while computational processes handle tagging and storage. This setup creates a self-sustaining «autopoietic system» with interdependent components and textual fields (McGann, 2004a). Alternatively, machine-oriented approaches harness the potential and iterative efficiency provided by algorithms. This enables the processing of tens of millions of literary objects with a level of scope, speed, and precision that would otherwise be unattainable. Such approaches leverage scalability, as defined by Weinstock & Goodenough (2006, p. 3), as «the ability to handle increased workloads by consistently applying a cost-effective strategy to expand a system's capacity». With the same algorithm, the process is scaled up, and a large number of similar cultural objects can be processed simultaneously. However, this potential advantage of these machine-oriented approaches may come at the cost of forfeiting many of the adaptable qualities of human judgment when dealing with cultural objects, which involve a type of reasoning that adapts more flexibly to various contexts.

Thirdly, from a technical perspective, the tools utilized can either be built from the ground up or can involve the use of existing software. In both cases, the central focus remains consistently on the human user or programmer. If digital humanists takes on the role of software development, they bear responsibility for every computational step and steer the algorithms along a specific path. This first option appears theoretically advisable, as it demands a comprehensive grasp of the subject. However, owing to the

steep learning curve of Natural Language Processing (NLP),[11] it can often be more time-efficient to make effective use of pre-existing software, aligning with Sample's (2013) suggestion that Digital Humanities should prioritize «sharing» over «building,» framing the DH discourse with an emphasis on inclusiveness rather than exclusion based on coding skills. Nonetheless, even when digital humanists choose to employ pre-existing software tools, they must still make decisions regarding which literary elements and concepts to explore. They then fine-tune and experiment with the software to produce and retrieve specific results. In other words, there always exists a layer of theoretical-interpretive reflection, even when dealing with seemingly autonomous algorithms such as stylometric approaches or scripts based on Large Language Models (LLMs).[12] Recognizing this aspect is crucial for dispelling concerns about the potential dehumanization of interpretation, a concern frequently raised in the context of CLC (Allington et al. 2016).

In fact, the debate about the dehumanization of interpretation, resulting in objectification, fails to take into consideration that even the hypothesis of presenting a literary analysis as defensible solely by claiming that the data is self-explanatory lacks a solid foundation on two fronts. Firstly, assertions that argue the data on which the analyses were carried out are 'raw data' overlook essential considerations such as data

---

[11] Natural Language Processing (NLP) refers to the field of computer science that focuses on the interaction between computers and human languages. The primary goal of NLP is to enable machines to understand, interpret, or generate human language in a way that is both meaningful and contextually relevant.

[12] Large Language Models (LLMs) refer to advanced computational models, particularly those built upon neural network architectures, that have been trained on vast datasets containing diverse examples of human language (Naveed et al. 2023). These models exhibit an unprecedented ability to understand and generate text in a manner that closely mirrors human language patterns, and can be valuable tools for discourse analysis due to their ability to understand and generate human-like text (Törnberg 2023). While large language models offer powerful capabilities for discourse analysis, it is essential to approach their use with critical thinking and a consideration of potential biases in the training data. Interpretation should involve a nuanced understanding of both the strengths and limitations of these models, ensuring that the insights derived are contextually relevant and meaningful.

reporting infrastructures, human errors, and numerous other factors. The term 'data,' which derives from its etymology and frequently emphasizes the concept of 'something given,' always involves a human element (Gitelman 2013).[13] Secondly, even with the most advanced algorithms, tracing patterns back is not inherently valuable. Mathematical functions designed to identify linguistic structures will naturally identify precise linguistic structures because that is their programmed function. However, without an interesting interpretation, statistics alone are not self-evident. This is why Moretti highlights, in the title of his last monograph, the risk of a 'falso movimento' (Moretti 2022) – an Italian expression denoting a circular motion – then offering throughout the book a critical assessment of the approach to computational criticism, an approach he himself had a hand in developing.

Another significant observation that consistently emerges at this stage, as it did in previous phases, is the imperative need for interdisciplinary collaboration in the field of CLC. This need becomes particularly evident when addressing highly complex software, such as cutting-edge Artificial Intelligence algorithms. In such cases, the required expertise extends beyond the literary domain and involves collaboration with computer scientists, as advanced techniques often necessitate a firm understanding of linear algebra, statistics, and web design.

In fact, the distinction drawn by Piotrowski (2018) between 'theoretical digital humanities' and 'applied digital humanities' is quite apt and can be of significant value here. Much like other computer sciences, 'theoretical digital humanities' can be defined as the 'research and development of means and methods for constructing formal models in the humanities.' (Piotrowski 2018, p. 2) It focuses on operationalizing and shaping computational models of cultural phenomena. On the other hand, 'applied digital humanities' can be likened to applied computational linguistics. Theoretical digital

---

[13] Gitelman's book explores the historical and cultural aspects of data in episodes spanning from the pre-digital to digital eras. It emphasizes that data is not merely a raw resource but also a *cultural artifact* that necessitates generation, protection, and interpretation. The essays delve into various topics, including the reciprocal influence between data and different fields, from the processes of data collection and utilization to the intellectual history of data, financial modeling, the origins of databases, online surveillance, and scientific data curation.

humanities often exhibit a closer alignment with the hard sciences, while applied digital humanities are more intimately connected to the humanities due to their involvement in using computational models and tools to assess new cultural artefacts. The distinction made by Piotrowski reflects the fact that researchers in this field acknowledge varying requirements for collaboration and degrees of computational complexity.

In this vein of thought, it becomes evident that, while the development of software tailored to highly specialized functionalities may, in theory, present an attractive prospect for achieving the full control, one must not overlook the steep learning curve inherent to such software and the consequential need for a specific skill set. Hence, it is advisable to look for a pragmatic equilibrium between the facets of temporal efficiency and analytical precision.

1.4.4. Explicanda

After completing the software application step, the fifth stage involves presenting the results obtained from the software. This step is represented as a red section, encompassing all the hypothetical outcomes resulting from CLC (Figure 1.7). As seen in the previous section, the extent of software utilization varies significantly in terms of nuances, scope, and human implications. Nevertheless, regardless of these variations, the eventual results still need to be represented in a specific manner to facilitate interpretation.

This characteristic distinguishes CLC from other literary approaches since it involves presenting the data retrieved and/or stored by the software. The process includes a form of remediation and transformation, which morphs the investigated literary objects into something else. These transformed objects serve as the actual territory for ultimate interpretation. In addition, within our space-time representation, we have identified specific points labeled as 'exceptions.' We will explain the rationale behind this choice later in this paragraph. To preface the discussion, we have opted to visualize these points because, among all the reshaped or charted results, identifying exceptions requires the utmost attention and can significantly influence the analysis, potentially leading us towards alternative paths.
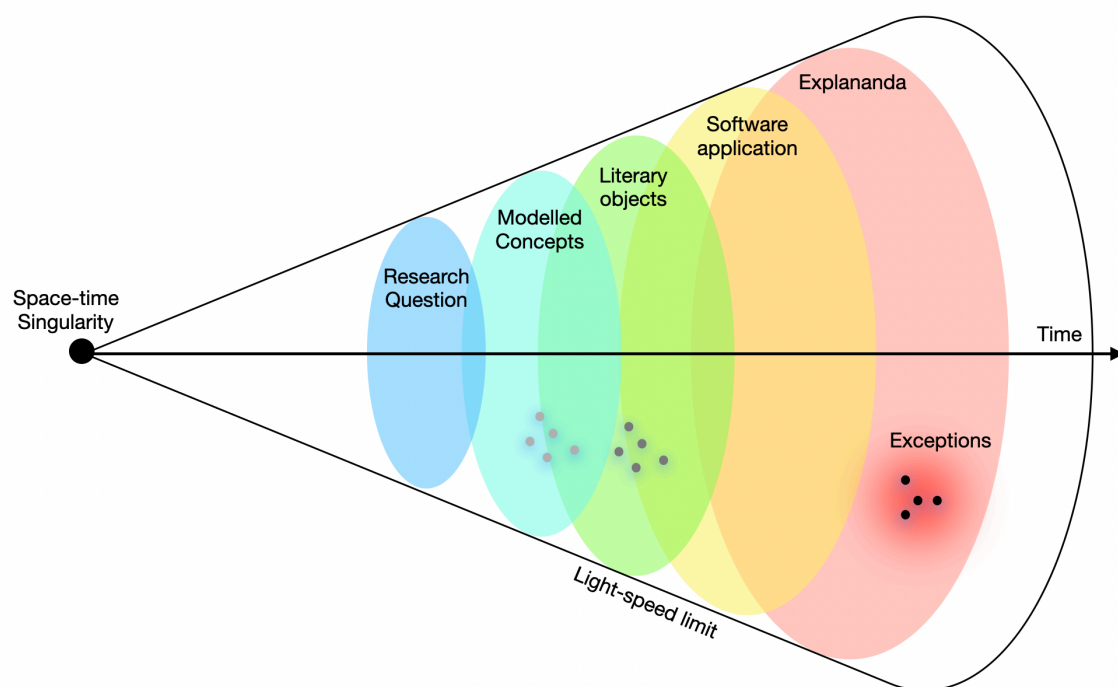
Figure 1.7. Visualization of the Explananda Formalization Step in Computational Literary Criticism.

Also, to emphasize the fact that some kind of visual, tabular, or textual representation of the results is not self-explanatory but requires another level of interpretative process, we refer to the object produced by this step as '*explicanda*,' or 'things that have to be explained'.[14]

These *explicanda* can take several forms, as demonstrated by our reference studies. Through a comprehensive analysis of these studies, we can discern that the visual representations showcased in critical works can be classified into the following types:

---

[14] Explicanda derives from the gerundive Latin form of the verb 'to explain' ('explico'). In Latin grammar, the gerundive is a verbal adjective that expresses necessity, obligation, or desirability. It is formed by adding the suffix '-ndus, -nda, -ndum' to the stem of a verb. When used substantively, the gerundive functions as a noun and often takes on a neuter gender. The word 'explicanda' is the neuter plural form of the gerundive of the verb 'explico,' which means 'to explain' in Latin. The resulting word 'explicanda' can be translated as 'the things that have to be explained' or 'the things to be explained.'

1) **Visual component entangled directly on the literary object**: Annotated literary text (Uenishi et al. 2022, Rizvi et al. 2022, Brill et al. 2020); web-based visualization platforms to visualize and compare texts (Howell et al. 2014, Menéndez de la Cuesta González 202116, Škrabal & Piorecký 2022, Asanov et al. 2021);

2) **Non-graphical schematization of data**: Tabular format (Smeets et al. 2021, Zehe et al. 2021, Uenishi et al. 2022, Murai 2020, Bode 2018, Janicki et al. 2023, Can & Cangır 2022, Kim et al. 2021, 24, Geybels 2023)

3) **Traditional data charting**: Network visualization or entity relation map (Kapitan & Wills 2023, Fan 2018, Rebora et al 2021, Uenishi et al. 2022, Murai 2020, Ciccoricco 2021, Varela 2021, Masías et al. 2017, Van Zaanen et al. 2020, Hazlon et al. 2021, Rakymzhan et al. 2022, Fischer et al. 2018), statistical types of graphs (word clouds, line charts, bar charts, box plots, scatter plots, dendrograms, tree-charts, etc.) (Queiroz et al. 2015, O'Sullivan et al. 2018, McCarty 2003, Zehe et al. 2021, Rebora et al 2021, Uenishi et al. 2022, Skorinkin & Orekhov 2023, Tian & Liu 2023, Brill et al. 2020, Kusi 2022, Bode 2018, Varela 2021, Rokib et al. 2018, Underwood 2019, Méndez 2020, Can & Cangir 2022, Kim et al. 2021, Li et. al 2021, Heuser et al. 2016, Berglund et al. 2022, Koidaki et al. 2021, Hosny 2018,

4) **GIS and types of digital cartography** (heatmaps, topological maps, chorographic maps, etc.) (Queiroz et al. 2015, Skorinkin & Orekhov 2023, Wrisley 2017, Varela 2021, Méndez 2020, Škrabal & Piorecký 2022, Borin et al. 2014, Heuser et al. 2016, Hazlon et al. 2021, Hosny 2018)

5) **Non-traditional data charting:** Specifically devised types of charts (Roberts-Smith et al. 2013, Heuser et al. 2016, Gervas 2009); heavily

information design-oriented interactive visualizations (Falcetto 2019, Dealberto et al. 2022, Grillo Gálvez 2019)

The first observation we can make pertains to the balance between creativity and complexity. While some critical works rely on hard sciences to generate graphs and avoid creating new types of graphs, others completely reshape the literary object in a creative manner. An aesthetically pleasing reshaping of the extracted data is a characteristic that has been proved helpful for visualisation usability (Cawthon and Vande Moere 2007), however it is not always useful for conveying the message, when it exceeds in an overly complex and intricate representation. In the realm of CLC, the issue is arguably exacerbated, particularly since literary interpretations could be considered inherently subjective, as highlighted by Bleich (2019) and numerous other scholars, with a distinctive emphasis.[15] Providing an even more subjectively created graph can add an additional layer of complexity. On the other hand, when research draws from other sciences and provide classical results, the reader needs to have competence in the science that studying is relying on (statistics, geography, etc.). This also can potentially create barriers to understanding the findings. To strike the right

---

[15] In the context of Bleich's theory, when we place the individual at the center of our critical concern and consider the act of literary reading, it becomes clear that this process involves subjective response, resymbolization, and negotiation, ultimately leading to validated knowledge. 'Subjective response' refers to the initial perceptual reactions towards a symbolic object (p. 96). 'Resymbolization' is the conceptualization of symbolized objects (p. 88), which emerges as a result of the need for explanation following the subjective response. 'Negotiation' is the subsequent response statement, encompassing a symbolic self-presentation, a contribution to a pedagogical community, and an articulation of the part of our reading experience that we believe can be transformed into knowledge (p. 167). According to this theory, literary interpretation both commences and concludes with the interpreting subject. Consequently, in the case of CLC, visualization may pose a challenge due to its potential idiosyncrasy. In our view, even if we were to base our CLC research on literary theories emphasizing the subjectiveness of interpretation, we should be cautious when explaining our visualization choice. Alternatively, we could opt for traditional charts to bridge the gap between the idiosyncrasies of literary interpretation and the necessity to convey our results to a broader community that needs to comprehend our findings, regardless of our perspective.

balance, critical works should aim to incorporate creativity in presenting data without sacrificing clarity and accessibility. By finding the sweet spot between creativity and complexity, researchers and critics can create impactful analyses that resonate with a broader audience while maintaining the integrity of the data and the message they wish to convey.

Secondly, the categories of time and space are extensively visualised, given their status as two pivotal aspects that shape the very essence of human experience, as numerous scholars have emphasized (Heidegger 2010, Lefebvre 1991).[16] Certainly, in traditional graphical representations such as scatter plots and line graphs, it is common to incorporate time or space as one of the dimensions. Certainly, fields such as geography, architecture, urban studies, and other sciences dealing extensively with spatial data have already dedicated significant thought to the visualization of spatial information over time. This is accomplished by connecting the technical and theoretical aspects of various Geographic Information Systems (GIS) and maps, as explored by Wade and Sommer (2006). These disciplines have delved into the societal implications of organizing space in specific ways, acknowledging that even the term 'information' within the context of Geographic Information Systems is subject to debate, as it is embedded in a distinct cultural framework. How space is categorized not only reflects society but also influences it, as underscored by Sheppard (1995). This also applies to Computational Literary Criticism, as fictional spaces display varying degrees of

---

[16] Among many others, Heidegger (2010) has certainly played a major role in steering Western culture towards a subjectively temporal and spatial perception of existence. In his well-known work *Being and Time*, he seeks to understand Dasein, i.e. the nature of human existence, and Dasein's relationship to Being. Dasein's existence is indeed characterized by both temporal and spatial dimensions, and its understanding of Being is shaped by its engagement with the world. Concerning the spatial dimension specifically, Lefebvre (2011) has certainly been one of the most prominent figures in what has since been called the 'spatial turn,' exploring the social construction of space as a dynamic and multifaceted process. Lefebre reflects theoretically on the categorization of space, theorizing the three interconnected spatial dimensions of 'perceived space', 'conceived space', and 'lived space'. Additionally, he underlined the pragmatic functions of social activity in the creation of space with the concept of 'spatial practice,' which emphasizes how people produce and shape space through their everyday actions.

referentiality that can be digitally mapped, thereby influencing our perception of the depicted locations.

On the other hand, visualizations can exhibit a remarkable level of creativity. They often integrate spatial and temporal concepts, either by blending them with other categories or by inventing entirely novel categories for visualization. Furthermore, these visualizations boast interactive elements that allow for dynamic data representation, effectively harnessing the potential of digital technologies and web-based visualization. In fact, numerous researchers have delved into both the strengths and weaknesses of visualization in the humanities. Their work spans from categorizing various visualization types (Sinclair et al. 2013; Benito-Santos & Sánchez 2020) to exploring the theoretical aspects of visualization when coupled with literary objectives (Moretti 2005; Drucker 2011; Barness 2015; Jänicke 2016; Graham 2017; Windhager et al. 2019). They specifically address the challenge of dealing with ambiguity and uncertainty, which is a unique feature of the literary realm and literary interpretation. This challenge is particularly daunting for metadata, which traditionally adheres to a binary opposition model.

Thirdly, among the types of visualizations present in critical works definable as CLC, we can draw another distinction. As this type of research involves a computational aspect, there are two main types of visual data involved. On one hand, there is the data regarding the literary objects and their features. On the other hand, data visualization can also be used to illustrate the effectiveness of the software itself, providing charts and graphs that explain the precision of the algorithms used – a practice that has become more and more frequent with machine learning and deep learning algorithms, which require a training process to develop a model (Chatzimparmpas et al. 2020). While both types of visualization are interesting, the first type holds predominant importance for CLC. This is because the primary goal of such an academic subject is not merely the computational efficiency in retrieving a type of data, but rather the interpretive results that arise from a reflection on literary-related objects. In fact, the most creative visualizations are not related to the training process of an algorithm but instead involve reshaping the literary objects themselves.

But above all, when it comes to visualization in CLC, arguably the most crucial observation is how this process abstracts and simplifies literary objects, transforming them into something entirely different that does not display but *create* knowledge: «Most information visualizations are acts of interpretation masquerading as presentation. In other words, they are images that act as if they are just showing us what is, but in actuality, they are arguments made in graphical form» (Drucker 2014, p. 10).

The process of literary interpretation no longer revolves solely around the original literary object from which the analysis began. Instead, it focuses on its reshaped aspects. Consequently, the critical evaluation of the abstractive process holds utmost importance. Only a positive assessment of what to visualize, ultimately becoming the central object of interpretation, can lead to an effective final analysis. If the focus of visualization strays towards something of little relevance, concealing more vital aspects, the very foundation of the interpretation risks being deemed incorrect.

This is why, in our space-time visualization designed to represent the acts of Computational Literary Criticism, we decided to emphasize certain points labeled as 'exceptions.' Visualizations have the power to make outliers and unforeseeable results more evident. For instance, the occurrence of a verb might be significantly higher than expected, or a literary work's style might differ greatly from others, causing it to be represented as a distinct and distant node in a network visualization, and so on.

These exceptions, when highlighted through visualization, can lead to new interpretive discoveries if the visualization is constructed through a correct retrieval process. Alternatively, they can raise awareness that the extraction process was flawed, necessitating a refinement of the algorithmic process. In other words, while a foreseeable trending line can certainly help confirm a traditional hypothesis of a literary era, it is the exceptions that deserve special attention within visualization, as they can lead to the unexpected and open doors to new discoveries.

1.4.5. Final interpretation

The final step involves interpretation, which is eventually represented as a pink section in our space-time visualization (figure 1.8). Through this step, researchers can

propose a new understanding of an author's literary work or tradition if they believe they have made a novel discovery. Alternatively, they can confirm an initial hypothesis, thereby supporting an existing critical tradition. Both options are valid, but they come with the risk of data manipulation. Researchers may downplay discovered exceptions to support a strong critical tradition that considers those exceptions impossible. On the other hand, they might actively search for something new, potentially overemphasizing unimportant outliers.
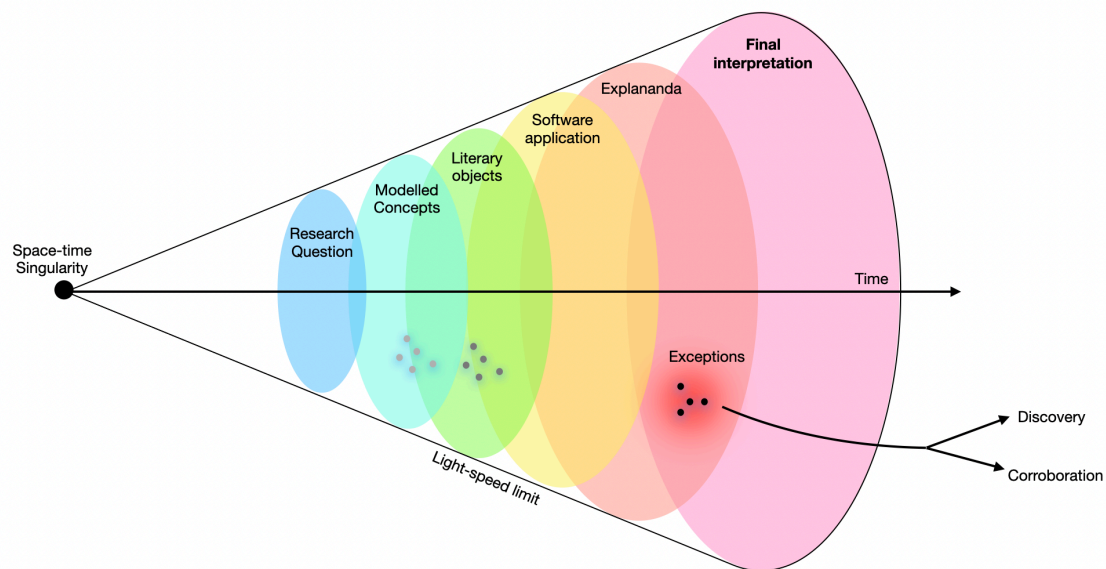
Figure 1.8. Visualization of the Interpretation Step in Computational Literary Criticism.

Drawing upon the recognition of the critical importance of exceptions as mentioned in 1.4.4, we have chosen to highlight how the identification of outliers and unexpected events guides the final interpretation. As a result, we have decided to illustrate with an arrow how the interpretation could stem from points representing exceptions, ultimately leading to the recognition of a discovery or the confirmation of the initial, possibly classical, literary interpretive hypothesis.

To summarize our findings on interpretation in our fifty reference texts, we observe that CLC research can be categorized as either content-focused or methodology-focused. The former (content-focused CLC research) encompasses: A) studies that delve

into the dynamics of interpretive outcomes, often reflecting extensively on fictional structures or attempting to establish connections between literature and various societal aspects; B) research that engages in debates about existing critical traditions related to the literary objects under investigation, with the goal of either affirming established viewpoints or uncovering new perspectives. The latter (methodology-focused CLC research) typically introduces new tools or engages in discussions about the utilization methods of existing ones.

These finding could be visualised in the following diagram (Table 1.9):

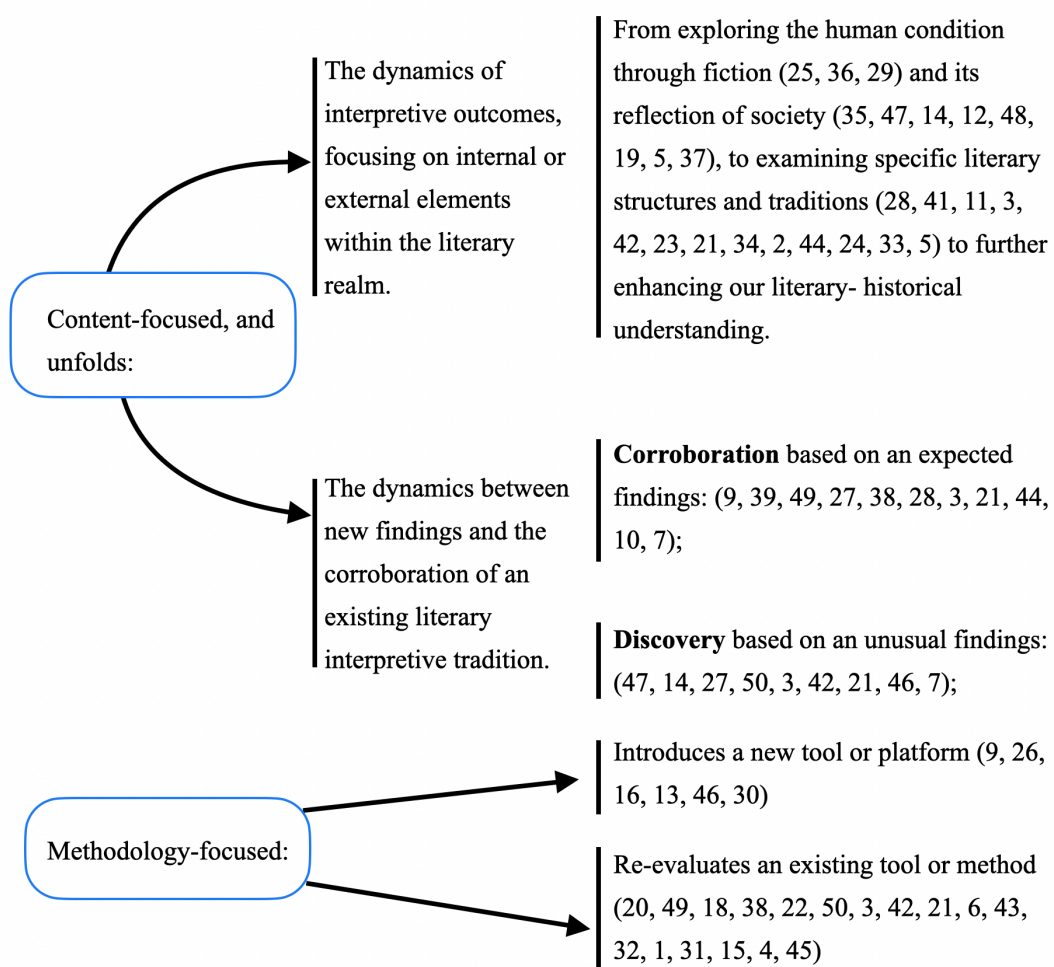| Content-focused, and unfolds:	The dynamics of interpretive outcomes, focusing on internal or external elements within the literary realm.	From exploring the human condition through fiction (25, 36, 29) and its reflection of society (35, 47, 14, 12, 48, 19, 5, 37), to examining specific literary structures and traditions (28, 41, 11, 3, 42, 23, 21, 34, 2, 44, 24, 33, 5) to further enhancing our literary- historical understanding.
	The dynamics between new findings and the corroboration of an existing literary interpretive tradition.	**Corroboration** based on an expected findings: (9, 39, 49, 27, 38, 28, 3, 21, 44, 10, 7);
		Discovery based on an unusual findings: (47, 14, 27, 50, 3, 42, 21, 46, 7);
Methodology-focused:		Introduces a new tool or platform (9, 26, 16, 13, 46, 30)
		Re-evaluates an existing tool or method (20, 49, 18, 38, 22, 50, 3, 42, 21, 6, 43, 32, 1, 31, 15, 4, 45)

Table 1.9. A schematization of the types of categories possible within the step of interpretation of CLC.

We can draw several insights regarding the interpretation step from these reference papers. Specifically, there are distinct dichotomous dynamics that are more specific to

CLC than to other approaches. The first conceptual dynamic involves the interaction between the pole of *content-focused research* and the pole of *methodology-focused approach*. On one hand, CLC approaches seem to prioritize methodological reflection. This entails a comprehensive reconsideration of the digital aspect concerning humanities or an evaluation of specific methods when applied to literary objects. It is no surprise that the field has been described as methodological since its inception, as McCarty recalls when he began teaching what he defined as «humanities computing:»

> In the mid to late 1980s […] I started teaching the subject. My students, from across the humanities and social sciences, taught me in turn that there was in fact a subject, and that it had to be about *method – the only scholarly concern all of us shared* [emphasis mine] (McCarty 2005, pp. 4).

Digital Humanities, along with CLC, have emerged from the application of digital approaches, and these shared methods form the very core that provides a common ground for various subjects within the humanities. On the other hand, other scholars primarily employ software applications in an instrumental manner, focusing on literary insights without delving deeply into the underlying methodology. From here stems the distinction between *content-focused* and *methodology-focused* approach that we indicated in our findings and consequently in our table helps categorize the interpretive results of the discipline more rigorously.

Furthermore, within content-focused research, papers and monographs present two sub-dynamics. Firstly, research oscillates between literary self-reflection and external reflection: researchers can employ digital methodologies to analyze literary structures, authors, or traditions, thus shaping a fully internal reflection that remains within the literary domain. Conversely, they may treat literary objects as symbolic means for exploring other concepts, placing more importance on the symbolic aspects than the literary object itself. This approach allows them to extend beyond the confines of the literary realm and explore societal contexts, examining their relevance at individual and collective levels. This aspect is not unique to CLC; even Frye (1957) in his second essay of *Anatomy of Criticism* accounted for five symbolic phases (literal, descriptive,

formal, mythical, and anagogic), corresponding to five types of symbols (motif, sign, image, archetype, and monad), and five types of literary criticism (textual, historical, interpretive, conventional/generic, and spiritual/religious). Though, this dynamic plays an important role in CLC, demonstrating that the field is not solely focused on (digital) methodology.

Secondly, content-focused computational literary critical research can oscillate between *corroboration* and *discovery*. As we suggested at the beginning of this section, CLC often confirms initial hypotheses when the data aligns with established critical traditions and interpretations. However, it can also venture into affirming new, even disruptive, literary interpretations based on unexpected data findings, leveraging the apparent authority of 'data' and hard sciences. This dynamic is arguably derived from computer science and data science, where - certainly much more than in literary studies - hypothesis-driven approaches are common (Smalheiser 2002). Researchers first formulate a hypothesis and then test it, eventually either corroborating the hypothesis or encountering data evidence that opposes it, leading to a discovery or something new in that case.

Finally, when conducting research, the main focus being on the interpretive aspects of methodology, digital literary scholars may introduce specially designed tools or platforms. Indeed, there has been a longstanding debate within the field about whether one of its specific goals is the capability to create something, be it a platform, a tool, or a code snippet (Ramsay & Rockwell 2012). Alternatively, researchers may engage in technical or theoretical discussions on existing methods without delving deeply (or at all) into case studies or literary examples to gain new insights or reevaluate existing ones. It is not coincidence that Svensson (2013, p. 66) pointed out that: «Humanities Computing as a whole maintains a highly instrumental approach to technology in the Humanities». The rebranding of 'Humanities Computing' as 'Digital Humanities', undertaken over the last two decades, especially following the establishment of the Alliance of Digital Humanities Organizations in 2005, aimed to expand perspectives and champion the subject's autonomy, rather than relegating it to a mere tool.

The division between methodology and content is not a random occurrence; it was a deliberate pursuit driven by specific motives. Initially, theorists – Moretti (2000) in the first place – engaged in distant reading because they aimed to read systematically and move beyond the traditional literary canon. Among them, Jockers (2016) asserted that «in the twenty-first century, literary scholars» could no longer «rely on anecdotal evidence or random data extracted from a few, even supposedly representative, texts» (Jockers 2016, p. 8). There was a mounting imperative to explore subjects beyond the conventional canon, thereby transcending the «theological»[17] approach criticized by Moretti in his work, *Conjectures on World Literature* (Moretti 2000, p. 57). He argued that literary scholars had grown excessively content with their habit of repeatedly delving into the same texts, often neglecting everything beyond them, even though culture as a whole demanded comprehensive investigation.

In other words, digital scholars had a twofold objective. Firstly, they aimed to broaden the scope of their research, which required the use of computational methods. Secondly, they sought to reduce the risk of overly subjective biases that had permeated literary studies in the post-structuralism era, where interpretation was largely idiosyncratic. In this context, some scholars were determined to untangle both the tools and methodologies from the actual content of computational-aided analysis. The only way to guarantee that the tools were impartial instruments, devoid of bias, thereby establishing a more robust foundation for interpretation, would have been to separate methodology from content, treating the methodological principles much like mathematical laws in applied statistics. In fact, Moretti underscored this very point by stating:

---

[17] «At bottom, it's a *theological exercise* [emphasis mine] – very solemn treatment of very few texts taken very seriously – whereas what we really need is a little pact with the devil: we know how to read texts, now let's learn how not to read them.» (Moretti, 2000, p. 57). Moretti argues that close reading treats specific works of literature as something of such significance that they could, ironically, be equated to the divine word. In his opinion, such a posture should be surpassed.

Quantitative research provides a type of data which is *ideally independent of interpretations*, I said earlier, and that is of course also its limit: it provides data, not interpretation. … Quantitative data can tell us when Britain produced one new novel per month, or week, or day, or hour for that matter, but where the significant turning points line lie along the continuum —and why—is something that must be decided on a different basis. (Moretti 2005, p. 9)

Far from being a relic of the early days of Digital Humanities, Rockwell and Sinclair (2022, p. 151) recently engaged in a similar debate on interpretive matters in their monograph *Hermeneutica: Computer-assisted Interpretation in the Humanities*. They made an attempt to distinguish between tools designed for analysis and the content of the analysis by posing the question: «What if we could consider things (like toys, websites, and tools such as Voyant) and their uses as forms of discursive practices?».

Following Moretti's methodology, their objective was to reduce bias and promote reproducible critical analysis. They aimed to incorporate these tools into the framework that underpins scientific discourse, much like a microscope serves as an instrument in biology without needing constant questioning. They even introduced a term, *hermeneutica*, to characterize the tools developed by Digital Humanities for analyzing cultural artifacts. These *hermeneutica* encompass both a software component (the code) and a hardware dimension (the servers hosting the code), giving them materiality and substantial presence. Due to their autonomy, they can be evaluated independently and then used as something granted, which does not need to be discussed repeatedly.

A similar perspective is articulated by William G. Thomas III in *A companion to digital humanities* (Schreibman et al. 2008, p. 66):

The goal for historians working in the new digital medium needs to be to make the computer technology transparent and to allow the reader to focus his or her whole attention on the 'world' that the historian has opened up for investigation, interpretation, inquiry and analysis.

The pursuit of transparency in digital tools mirrors the effort to mitigate bias. Since it is nearly impossible to entirely eliminate bias in interpretation, literary scholars who have turned to computational methods can only strive to reduce bias in the instruments they employ. This results in a notable distinction between methodology and content within the context of CLC, a distinction we previously asserted did not arise by chance.

Ironically, this very effort to reduce and ultimately eliminate bias, especially in methodology and the tools employed in digital methodologies, has drawn significant criticism because it has led opponents to criticize Computational Literary Criticism as «anti-interpretive.» They argue that it lacks depth, as exemplified by a highly cited article in the field of Digital Humanities, published in the *Los Angeles Review of Books*:

> Digital humanities has often tended to be anti-interpretive, especially when interpretation is understood as a political activity. Digital Humanities instead aims to archive materials, produce data, and develop software, while bracketing off the work of interpretation to a later moment or leaving it to other scholars—or abandoning it altogether for those who argue that we ought to become 'postcritical' (Allington et al. 2016).

Mitigating bias is crucial, but it could also mean refraining from adopting a political stance. And for those who see cultural critique as inherently intertwined with an intellectual perspective aimed at steering society towards progress, refraining from taking a political position can dilute the depth of cultural discourse and pull it towards its least constructive manifestation.

This debate is particularly intriguing because, with its circular movement, it harks back to the era of Structuralism. Just as discursive isotopies do, similar critical dynamics surface, vanish, and resurface in new forms decades later. Much like the challenges faced by Structuralism, CLC, in its quest to derive 'objective' formulas, has encountered criticisms similar to those another school of thought once grappled with, attempting to avoid a potential reduction in critical depth.

Now that we have raised all these possible critical positions toward the interpretive step of CLC, instead of prescribing what it should be or what the correct positions

should be based on, we would like to leave the discussion open, as critical discourses are contrastive, multifaceted, and inherently intertwined. No discourse or text lacks a – sometimes problematic – connection with another: «Any text is constructed as a mosaic of quotations; any text is the absorption and transformation of another» (Kristeva 1980, p. 66).

Chapter II

2.1. Introduction: (how) should algorithms cope with the Irish Troubles?

After proposing a hypothesis for a comprehensive theory of CLC, we are ready to present an illustrative example. Our argument revolves around the notion that CLC has the capacity to offer fresh insights and contribute to the rejuvenation of literary studies. It has the potential to shed light on how enduring questions can continue to be subjects of discussion, thus reopening the literary discourse on matters like: «What defines literature?» and «How should interpretation be understood?».

Our study is built upon the premise that harnessing the full potential of CLC necessitates staying up-to-date of cutting-edge methods in computer science. As we will see, our research incorporates state-of-the-art algorithms, rendering it contemporary in its approach, but provocative due to its high level of interdisciplinarity. To amplify the provocative nature of our work, we have chosen to concentrate on literary works associated with a contemporarily debated case study: the Northern Ireland Troubles. This topic has recently regained significance in post-Brexit Europe, where Ireland remains in the European Community while the United Kingdom does not. This situation underscores the challenges of border crossings. As Hall (1997) notably emphasized, culture—and fiction as a manifestation of culture—should serve as a focal point for societal reflection:[18] culture embodies the intricate dynamics of shared meanings, and exploring the cultural landscape can thus provide novel perspectives, offering a deeper understanding of societal occurrences. This is why opting for a focus on novels accounting for The Troubles could prove beneficial in the current study.

Additionally, given its sensitive nature, the political case of The Troubles also underscores a notable distinction between machines and human involvement. When delving into both fictional and non-fictional accounts addressing pain and terrorism

---

[18] «We should perhaps learn to think of meaning less in terms of 'accuracy' and 'truth' and more in terms of effect exchange - a process of translation, which facilitates cultural communication while always recognising the persistence of difference and power between different 'speakers' within the same cultural circuit» (Hall 1997, p. 10).

related to The Troubles, the choice needs consideration of the role played by algorithms and machines in the interpretation process. To put it simply, despite efforts to create artificial intelligence machines capable of displaying and recognizing human emotions (Picard, 2008), machines inherently lack empathy; they have never experienced pain. This raises ethical questions, including whether machines should handle sensitive documents in the first place and the associated ethical implications. These considerations extend to understanding what constitutes pain, how to respect actual testimonies, and the importance of avoiding the imposition of misinterpreted frames on data that could distort experiences having caused profound pain.

In a technical sense, we have the capability to regard sensitive testimonies and emotionally charged expressions of pain as mere soulless data, processing literature as we would any other form of information or text. However, the question arises: should we? What is the rationale behind employing computers to analyze literature, possibly erecting a barrier between the text and the reader? This alteration in approach may significantly impact the reader's reception by creating a distance from the immersive experience of the unique literary work.

Certainly, fiction transcends the realm of mere historical accounts. Nevertheless, when delving into the analysis of literary texts as imaginative narratives rooted in societal contexts, we can employ critical discourse analysis techniques, as suggested by Fairclough (2013). By incorporating narratological spatial tools, we gain the ability to explore the discursive pluralism that arises from the intricate interplay of what Fairclough identifies as 'ideological-discursive formations.'

> I propose to use for talking about institutional pluralism Pêcheux's term 'discursive formation' as well as Althusser's 'ideological formation' Pêcheux defines a discursive formation as 'that which in a given ideological forma- tion, i.e., from a particular position in a given conjuncture determined by the state of the class struggle, determines 'what can and should be said'' (Pêcheux 1982: 111). I shall refer to 'ideologicaldiscursive formations', in accordance with what I have said above about the inseparability of 'ways of talking' and 'ways of seeing'. (Fairclough, p. 43)

Language, including literary discourse as a specific form of discursive expression, is inherently intertwined with power dynamics. Contrary to a monolithic nature, social institutions present diverse sets of discursive and ideological norms. Fiction, among various textual forms, serves as a mirror to these dynamics, reshaping ideological norms through the unique lens of an author. In politically charged literary discourses, such as narratives recounting events like The Troubles, these power dynamics become more pronounced.

At the same time, like any political conflict, understanding the underlying reasons behind the dispute while respecting both political sides is inherently complex. It is essential to approach the matter with caution and strive to minimize bias. To achieve this, rather than taking sides, we have chosen to adopt the perspective of fictional victims of violence, irrespective of their political affiliations. In narratives depicting turmoil, literary characters undergo pain and articulate their experiences in various ways. This is where the intersection of critical digital humanities and spatial literary studies,[19] which we aim to explore, becomes evident: the correlation between space and violence in fiction can be analyzed using spatial theoretical categories to discuss how violence is portrayed. In this way, instead of aligning with a specific political stance, we can contemplate the repercussions of violence and harness the educational potential that fiction offers, as articulated by Tzvetan Todorov:

> Literature can do a lot. It can reach out to us when we are deeply depressed, guide us towards the human beings around us, help us better understand the world, and assist us in living. It doesn't intend to be a way to heal the spirit; nonetheless, as a revelation of the world, it can also, as we go along, profoundly transform us. (Todorov 2007, p. 52)

We argue that reflecting on violence and spatiality could help us better understand how culture elaborated on the pain of The Troubles, and we aim to deepen this understanding by investigating the dynamics of various ideological-linguistic

---

[19] An account of spatial literary studies and its intersection with Digital Humanities will be provided in paragraph 2.6.1.

formations. In other words, instead of hubristically asking algorithms to take a political stance on our behalf regarding The Troubles, or to empathize with painful experiences they cannot comprehend, we acknowledge that the educational value lies in the operationalization process (Bridgman 1927; Moretti 2013) that we are about to undertake.

When delving into the examination of a diverse range of literary works associated with politically sensitive case studies, and utilizing algorithms to handle confidential data and testimonies, it becomes imperative to consider how these machines should be trained. As discussed in the initial chapter, the choices we make guide the functioning of algorithms, and therefore, our judgment plays a pivotal role in determining how such matters should be handled: we need to introspect and engage in a preliminary debate about what is relevant to extract from these literary objects.

In light of this, according to the initial step of our theory, we have selected a precise abstract concept for the process of algorithmic retrieval and literary investigation: the relationship between pain and spatiality. As we delve into operationalizing the connection between spatiality and pain, our primary objective is to demonstrate the effectiveness and sensitivity of extracting these elements from literary texts. The outcome of this process, involving abstraction, text retrieval, and interpretation, should result in a deeper understanding of pain, specifically the anguish arising from the events of The Troubles.

The subject matter is far from simple, but by proceeding step by step, we aim to establish its cultural, conceptual, and ethical significance. Now, let us embark on this challenge.

2.2. Roadmap: Bridging from Testimonies to Fictional works

Due to the conceptual complexity of this chapter, it is crucial to provide a clear roadmap for the reader, outlining the progression of paragraphs. This will facilitate the reader's understanding of our current analytical position. If the ultimate goal of this chapter is to explore literary works that connect to The Troubles and investigate how the

perspective of victims is sublimated, and then eventually examining the intersection of pain and spatiality in fictional works, we must approach this task with precision, following the designated steps:

1) Introduction to the political context (Paragraph 2.3): Begin by acquainting the reader with the political landscape surrounding the topic at hand, encompassing root causes and offering a comprehensive description of the contemporary situation.

2) Creation of two distinct corpora (Paragraph 2.4): One comprises testimonies related to The Troubles, while the other encompasses novels depicting the same period.

3) Analysis of the perspective of victims regarding violence and pain through actual testimonies (Paragraph 2.5): Delve into the examination of firsthand accounts from the first corpus, shedding light on the experiences of victims and their perspectives on violence and pain.

4) Examination of the intersection between pain and spatiality in fictional works (Paragraph 2.6): This entails comparing the portrayal of pain in novels (by focusing on the second corpus) with the analyzed testimonies from the first corpus. The objective is to assess whether the novels under scrutiny depict pain similarly to or differently from real-world testimonies. Subsequently, the exploration delves into the interconnection of pain and spatiality in the realm of fiction, dissecting the structural components of their relationship.

As some may observe, we declare our intention to explore spatiality in literature while sidestepping an in-depth examination of the spatial aspects of real-life testimonies. While spatiality undoubtedly holds historical significance in actual testimonies by reflecting the locations where violence occurred, we are not interested in spatiality within testimonies. Our focus is on understanding the perspective of victims regarding pain, leading us to turn to testimonies, then transitioning to the investigation of the intersection of pain and spatiality in fictional works, as we assume that spatiality plays a significant role in fictional works due to its structuring and semiotic value.

2.3. Historical Account: From the Roots to the Contemporary Situation

The origins of The Troubles lie in historical tensions between predominantly Protestant unionists, advocating for Northern Ireland's retention within the United Kingdom, and predominantly Catholic nationalists, aiming to reunite the region with the Republic of Ireland (Darby 1995). While the antecedents trace back to the 17th century (Bardon 2011), we choose to focus on historical events of the last century concerning the causes of The Troubles. This focus is necessary as our study specifically centers on novels published within the conventional timeframe of the conflict, spanning from 1968 to 1998.

In the past century, our foundational timeline commences in 1920 when the British implemented the Government of Ireland Act, establishing two distinct parliaments: one in Belfast, overseeing six of the nine Ulster counties, and another in Dublin, governing the remaining Irish territory (Murphy 1986). By the following year, 1921, the Anglo-Irish peace treaty granted Ireland 'dominion' status, providing substantial autonomy in self-governance while remaining part of the British Commonwealth. In contrast, Northern Ireland remained part of the United Kingdom (Fair 1972). These tensions escalated into a civil war across the 26 Irish counties, pitting supporters and opponents of the Anglo-Irish Peace Treaty against each other, concluding with victory for the supporters of the Treaty (Hopkinson 2004). Subsequently, on Easter 1949, Ireland relinquished its dominion status, officially becoming a fully independent republic through The Republic of Ireland Act. This legislation not only designated Ireland as the Republic of Ireland but also vested the president of Ireland with the authority to exercise the state's executive power in external relations, acting on the advice of the Government of Ireland.

The Republic of Ireland Act remains in effect in Northern Ireland until 1969, during which the Parliament in Belfast manages domestic affairs. However, during this period, the Catholic population faces discrimination in political representation, as well as limited access to social services and public employment. The year 1969 marks a turning point when civil rights movements in Northern Ireland start highlighting the disparities

between the two communities. Tensions escalate with a Catholic uprising in Derry, triggered by opposition to a Unionist march through the Bogside, the historic Catholic neighborhood in the city (Hepworth 2021).

Among some of the most critical moments in the conflict, at least two are relevant. In 1972, following a civil rights demonstration in Derry, British Army paratroopers shot into the crowd, resulting in the deaths of 13 unarmed demonstrators. This infamous Sunday would later be known as 'Bloody Sunday.' The incident triggered an escalation of the conflict and increased activity from militia groups, including the IRA, as well as groups aligned with opposing political affiliations, such as the Ulster Defence Association (UDA), among others (Walsh 2000).

The conflict persisted in alternating phases, with Ira leading a resilient resistance through guerrilla and symbolic actions. Notably, they advocated for Irish political prisoners, exemplified by the 1981 hunger strike resulting in the death of Bobby Sands, a Provisional Irish Republican Army member. Sands, a key figure in the 1976 Balmoral Furniture Company bombing, succumbed to hunger while imprisoned at HM Prison Maze. The hunger strike, led by Sands, aimed to protest the removal of Special Category Status. This status granted certain prisoners exemptions, such as not wearing uniforms or engaging in prison work, residing within their paramilitary factions, and receiving additional visits and food parcels. Sands' demise, along with nine other hunger strikers, sparked a renewed wave of IRA recruitment and activity. Global media coverage drew attention to the hunger strikers and the republican movement, eliciting both praise and criticism.

After an extended and renewed phase of unrest, the Good Friday Agreement was ultimately signed on April 10, marking the establishment of a formal peace and bringing an end to the conflict. The decision to pursue such an agreement was grounded in a moral commitment to rebuild the institutional framework of Northern Ireland's society. This aimed to create an environment where individuals of Catholic and Protestant, nationalist and Unionist backgrounds would no longer feel diminished or threatened in terms of their distinct identities. The signing of the agreement vividly illustrated the human-made nature of collective identities – emphasizing that we define ourselves by

our choices, despite historical constraints that may portray them as unalterable outcomes (McSweeney 1998).

As of December 2023, concerns regarding the repercussions of The Troubles continue to be highly pertinent and have attracted increased attention. Despite potential formal efforts by Northern Ireland and the United Kingdom towards fostering peace, there have been numerous instances where achieving or maintaining such peace appeared challenging. The Good Friday Agreement, viewed by many as sustaining 25 years of peace (Tannam 2001), marked the formal conclusion of The Troubles. However, in reality, it served as the initial phase in a sequence of subsequent peace negotiations and agreements, leading to intricate political developments not necessarily anticipated by the Agreement's creators (Nagle 2017; Doyle 2018).

After the peace agreement initially implemented power-sharing in Northern Ireland's local assembly, London regained control following periods of unrest. It was only through subsequent negotiations and the 2006 St. Andrews Agreement that the local government was reinstated, but political deadlock has endured (Patterson 2012). Moreover, the recent departure of the United Kingdom from the European Union has rekindled challenges for the agreement, prompting worries about Northern Ireland's border with the Republic of Ireland and its connection with the rest of Britain (Gormley-Heenan & Aughey 2017).

In facts, the aspiration for peace to overcome societal divisions has not fully realized itself up to the present. Economic prosperity remains elusive in numerous areas, and Northern Ireland continues to experience profound segregation. Official data reveals that only 7 percent of schools are officially integrated, with the majority still identified as either Catholic or Protestant (Borooah & Knox 2017). Peace walls persist in dividing neighborhoods along identitarian lines, serving both as territorial markers and as tools for preserving and reinforcing identity (McAtackney 2011). Furthermore, paramilitary groups retain significant influence in various regions and persist in using violence, such as targeting young men accused of crimes, to assert control over neighborhoods (Bakke & Rickard 2023). In many of these areas, there seems to be strong public support for such extrajudicial enforcement, viewing these groups as more reliable in action

compared to the official authorities. This phenomenon, observed in post-conflict countries, poses a substantial barrier to achieving lasting peace, as highlighted by Dyrstad & Hillesund (2020).

The persistent difficulties confronting Northern Ireland, notwithstanding its comparative benefits in wealth, geography, and institutional infrastructure, underscore the significant barriers to establishing enduring peace. In other words, when people refer to the peace process in Northern Ireland, they are not referring to events of the 1980s: they are talking about an ongoing process that continues to this day. This is why exploring novels centred on The Troubles and the associated hardships could hold ethical significance.

2.4.1. Corpus #1: The perspective of victims

The next step, as outlined in paragraph 2.2, is creating the datasets, which is, in fact, the second stage of CLC. As mentioned previously, we built two distinct corpora: one comprising testimonies from relatives and loved ones who lost someone during The Troubles, and the other consisting of novels depicting the same historical period.

Regarding the first corpus, three key points serve as the foundation for delineating a corpus from which we can extract the victim's perspective on pain and violence:

1) Firstly, a key reference in understanding The Troubles is McKittrick's (2001) extensive journalistic report, spanning 1600 pages. This work meticulously documents every casualty of The Troubles, amounting to more than 3600. Each victim's account is accompanied by a concise biography, and each chapter includes testimonies from their relatives and loved ones.

2) Secondly, we can turn our attention to the 'Wave Trauma Center', also known as WAVE,[20] an organization founded in 1991 during Northern Ireland's Troubles. Initially established as a support group for widows, it has since evolved into the largest cross-community victim's organization in the region. Through the efforts

---

[20] Refer to: https://wavetraumacentre.org.uk/ [Last checked 2nd of December 2023]

of WAVE, we now have access to audio and video recordings of victim's interviews, providing valuable insights and perspectives.

3) Thirdly, we consulted the CAIN Archive project[21] – A thorough compilation of information and source materials on The Troubles and politics in Northern Ireland, covering the period from 1968 to the present. This collection, created by Professor Bill Rolston, stems from an ongoing research initiative initiated in 1997. The project has received funding from numerous research programs, notably the UK's nationally relevant Arts and Humanities Research Council (AHRC), the Economic and Social Research Council (ESRC), and the European Union's PEACE III program. Among its extensive resources, the collection prominently features diverse materials such as interviews and reports from victims.

Using these three resources, we meticulously gathered and examined data from a comprehensive pool of 388 victims, that can be listed as following:

- 98 testimonies in an audio format[22] from the 'Stories from Silence' section of the Wave Trauma Center platform: These testimonies last from one to five minutes, and they all come from direct relatives (sisters and brothers, partners, daughters and sons, etc.) of the victims.

- 14 testimonies from Smyth & Fay (2000): among the fourteen testimonies included in the book, six of the interviewees are Protestant, seven are Catholic and one is from an ethnic minority, thus provision a reasonable diversity. The

---

[21] Please see: https://cain.ulster.ac.uk/victims/index.html [Last checked 2nd of December 2023]

[22] The audio were converted from audio to textual form by using a Language Model-based python script, that leveraged AssemblyAI (https://www.assemblyai.com/) and its speech-to-text technology. For a in-depth exploration of how these models work, please see Wu et al. (2023). AssemblyAI's models utilize machine learning to first associate sounds and phonemes and then identify grammatical errors by assessing the likelihood of a word following another. Consequently, the podcasts were automatically transcribed and subsequently corrected as necessary based on English grammar rules. The software yields results that typically boast an accuracy rate of over 95% for the English language. All the testimonies can be heard first-hand on the site: https://wavetraumacentre.org.uk/who-we-are/wave-stories/ [Last checked on December 5th, 2023]

interviewers interviewed 77 victims and then decided for the most representative ones, with respect to gender balance, geographical diversity, and urban/rural provenance.

- 58 testimonies from the project 'Accounts from the conflict', a digital archive of personal accounts of the conflict, based in Ulster University and designed to provide for the long-term storage of stories related to The Troubles in, and about, Northern Ireland. It contains testimony in textual form and video format.[23]

- 155 short testimonies from McKittrick (2001): Each chapter of the reportage begins with a series of short testimony citations that does not primarily focus on describing actual violence. Instead, they document emotional pain, which the editors argue is more impactful.[24]

- 1 testimony from White (2000), which recounts the self-narrated story of journalist Shane White, detailing his childhood in 1960s South Armagh, where his father served as a Royal Ulster Constabulary (RUC) sergeant. The book vividly depicts the locales of South Armagh through anecdotes involving drunken postmen, RUC pursuits, and childhood friends who later became casualties of The Troubles.

- 1 testimony from Grimaldi (1998): The book features a textual and photographical testimony from the reporter Fulvio Grimaldi, who was conducting a reportage in Derry/Londonderry when the events of Bloody Sunday occurred.

---

[23] Please see: https://accounts.ulster.ac.uk/repo24/index.php [Last checked on December 5th, 2023]

[24] Testimonies include examples such as: «I was grazed by a bullet and Patrick seemed to fall along the wall. I thought he fainted from seeing me bleed but then 1 saw that the back of his head was covered with blood and 1 knew the flashes had been bullets and that Patrick was shot.' -The father of schoolboy Patrick Rooney» (McKittrick 2001, p. 30). Although brief, these testimonies have proven to be valuable resources for capturing the emotional aspects of victims recalling pain.

- 1 testimony from Myers (2006), a firsthand account of The Troubles by a Irish Catholic journalist. The book vividly depicts the escalating conflict, detailing the actions of loyalist gangs, IRA members, and the impact on society, by combining personal and political narratives.

- 1 testimony from Spencer (2005), that features interviews with 15 close individuals connected to the victims of the Omagh bombing occurred on 15th August, 1998.

- 4 testimonies from Rolston & Gilmartin (2000): The text delves into the role of the state in a significant portion of the fatalities during the Northern Ireland conflict. It uses firsthand accounts to underscore the challenges faced by regular individuals.

- 1 testimony from Restorick (2000), in which Rita Restorick recounts the story of her twenty-three-year-old son, Stephen Restorick, who was killed by a sniper's bullet on February 12, 1997, while manning a checkpoint in South Armagh.

- 1 testimony from chapter 10 from testimony from Pringle & Jacobson (2000): The book covers the events of Bloody Sunday, detailing the human stories behind the tragic event and its profound impact on Irish history.

- 15 testimonies from Chapter three of O'Connor (1993): The book includes a series of interviews with Catholics from every walk of life. It explores attitudes toward Britain, the Republic of Ireland, the Church, Protestants, and the IRA's campaign during the period of the Conflict.

- 8 testimonies from Chapter 4 of MacKay (2000): The publication discusses the findings of research carried out by the renowned journalist Susan McKay. It entails the presentation and analysis of more than sixty in-depth interviews conducted with a diverse array of northern Protestants.

- 1 testimony from Chapter 7 and 8 from McClean (1997), with a first-hand authorial account of the violence perpetrated in the infamous event.

- 1 testimony from McCann (1974): In this work, McCann details his personal experience growing up Catholic in a Northern Irish ghetto during key events in

Derry. He observed the civil rights movement's shift from a peaceful pursuit of British Democracy to an assault on the British state.

- 3 testimonies from Lindsay (1998), which present personal narratives of British soldiers who served in Northern Ireland. These testimonials offer insights into a life on the precipice, the evolution of the conflict, and the profound impact of fear and trauma on the human experience.

- 4 testimonies from Kerr (1996): This book chronicles the experiences, beliefs, perspectives, and aspirations of a diverse group of individuals from Derry who played integral roles in the challenges of Northern Ireland. The content is derived from a set of interviews conducted between September 1995 and April 1996.

- 18 testimonies from Kennally & Preston (1971): The books describes the resistance faced since the Civil Rights campaign began in 1968, leading to government crises and unfulfilled promises of reforms. The author points out the misuse of the British army by the Unionist Government for repression, and features relevant testimonies from first-hand witnesses.

- 1 testimony included in chapter 8 from Harris & Healy (2001), which challenges the passive victim portrayal of Northern Irish women in the conflict by documenting nationalist women's varied experiences, highlighting rural perspectives and everyday human rights violations.

- 1 testimony from Chapter twelve of Devlin (1969): the book recounts the protest movement of 1968-1969 and emphasizes the personal cost of preserving integrity, and the willingness not to compromise.

- 1 testimony from Daly (2000): The book chronicles Edward Daly's journey from his childhood in Belleek, through his priesthood, focusing on his impactful years as a curate in Derry during the Northern Troubles. His experience as a bishop of Derry makes his testimony particularly valuable.

The total number of testimonies is 388. This effort resulted in a corpus of 409,999 words. Each testimony not originally in textual form underwent transcription (in the case of audio or video format) or digitization (if not already in digital format).

2.4.2. Corpus #2: Novels relating The Troubles

Regarding the second corpus, we compiled a collection of literary texts centered around The Troubles. This undertaking aimed to enable a comprehensive comparison of the portrayal of violence and spatial dynamics derived from both authentic testimonies and fictional narratives.

To compile the corpus of novels related to The Troubles, we drew upon the extensive groundwork already accomplished and documented by the CAIN Archive project. This project has diligently cataloged a comprehensive list of novels addressing The Troubles. Utilizing this curated list, we crafted a corpus consisting of 41 novels published between 1968 (the conventional starting point for The Troubles) and the present day. This process resulted in a corpus of 3587702 tokens, eventually serving as an exemplary research field, offering insights into the literary discourse surrounding the historical period of The Troubles. The list of titles included in our corpus is the following:

Author	Title	Year of publication
Bateman, C.	*Divorcing Jack*	1995
Bateman, C.	*Of wee sweetie mice and men*	1996
Bateman, C.	*Maid of the Mist*	1998
Berry, F.	*Northern Spy*	2021
Bradby, T.	*Shadow Dancer*	1998
Breslin, J.	*World Without End, Amen: A Novel*	2012
Burns, A.	*Milkman*	2018
Clancy, T.	*Patriot Games*	1988
Clarke, S.	*Soldier E: Sniper Fire in Belfast*	2013
Davies, M.	*The Drumbeat of Jimmy Sands*	1999
Dickinson, P.	*The Green Gene: A Crime Novel*	2015
Doyle, R.	*A star called Henry*	1999
Easterman, D.	*Day of Wrath*	2014
Edwards, R. D.	*The Anglo-Irish Murders*	2001

Healy, D.	*A goat's song*	1994
Herron, S.	*The Whore-Mother*	2014
Higgins, J.	*A Prayer for the Dying*	1987
Higgins, J.	*On dangerous ground*	1994
Higgins, J.	*Angel of Death*	1996
Higgins, J.	*Drink with the Devil*	1996
Higgins, J.	*Day of Reckoning*	2000
Higgins, J.	*Without Mercy*	2005
Higgins, J.	*The Violent Enemy*	2008
Hutson, S.	*Knife Edge*	1997
Lingard, J.	*Across the Barricades: A Kevin and Sadie Story*	2003
Lingard, J.	*The Twelfth Day of July: A Kevin and Sadie Story*	2003
Lingard, J.	*Into Exile*	2017
Madden, D.	*One by one in the darkness*	1996
McCabe, P.	*Breakfast on Pluto*	1998
McDonald, I.	*Sacrifice of Fools*	1996
McLaim R.W.	*Eureka Street*	1996
McNamee, E.	*Resurrection Man: A Novel*	1995
O'Brien, E.	*House of splendid isolation*	1994
Park, D.	*The rye man*	1994
Park, D.	*Stone kingdoms*	1996
Park, D.	*Stone Kingdoms*	2015
Patterson, G.	*The International*	1999
Petit, C.	*The Psalm Killer*	1996
Silva, D.	*The Marching Season: A Novel*	1999
Strong, T.	*The Tick Tock Man*	1994
Weber, K.	*The Music Lesson*	1999

We will now proceed delving into the next step of CLC – the software application step. Simultaneously, we will revisit the process of abstraction related to pain and spatiality. In fact, to retrieve these concepts from our corpora, it is essential to computationally precisely describe the linguistic constituents being extracted and elucidate the methods employed by the machine to extract them: without this detailed description, our thesis would be deemed indefensible.

2.5.1. Pain #1: Macro-analysis of the depiction of pain in testimonies

To compare the perspective of victims regarding pain and the depiction of pain in novels, we need to extract and analyze passages depicting pain from real-world testimonies. The primary assumption here is that pain finds linguistic expression through lemmas associated not only with the direct experience of pain ('to suffer', 'agony', 'distress', among others), but also with lemmas related to violence ('cruelty', 'coercion', 'to kill', among others).

This connection is particularly relevant as we delve into the exploration of pain induced by acts of (terror) violence committed by groups like the Ulster Defence Association (UDA) or the Irish Republican Army (IRA), among others. In our specific context, the intertwining of pain and violence is inevitable. Hence, we require an initial list of lemmas encompassing both violence and pain. We therefore started by using Miller's Wordnet Project (Miller 1995), and compiled an initial list of lemmas relating to the word 'pain' and to the words indicating the weapons used by IRA according to Oppenheimer (2008, pp. 264-268) with their lexically related words. Such a list of lemmas includes the following: 'Attack', 'axe', 'bleed', 'blood', 'bomb', 'bombing', 'bruise', 'brutality', 'brutalize', 'bullet', 'choke', 'conflict', 'coercion', 'cruelty', 'crush', 'cry', 'damage', 'death', 'decapitate', 'explosive', 'fear', 'fire', 'firearm', 'fracture', 'grenade', 'gun', 'harass', 'harm', 'hit', 'hostility', 'incendiary', 'injure', 'intimidate', 'kill', 'knife', 'lash', 'machete', 'maim', 'massacre', 'mutilate', 'oppress', 'paramilitary', 'pistol', 'projectile', 'revolver', 'rifle', 'scream', 'shotgun', 'shout', 'smite', 'stab', 'strangle', 'tears', 'terrorize', 'torment', 'violate', 'violence', 'wound'.

However, if we were to extract lemmas from a precompiled list, there might be concerns about bias. This is because our initial compilation is rooted in a language register closely associated with the contemporary variant of English, while our corpus is chronologically situated a few decades in the past.

To address this potential bias and attempt to disentangle our choices from linguistic predispositions, we turned to word vectors.[25] Along this line of thought, after establishing an initial set of lemmas associated with pain and violence, we process the

---

[25] Word vectors, also known as word embeddings, constitute numerical representations of words within a continuous vector space, initially introduced by Mikolov (2013). These mathematical representations capture the semantic meaning of words based on their context in a given dataset, grounded in the hypothesis of distributional semantics (Lenci 2008), a linguistic and computational approach that focuses on understanding the meaning of words based on their distributional patterns in a given context. Within a word vector space, words with similar meanings are positioned closer to each other, reflecting their semantic similarities and facilitating the identification of analogies based on the proximity of word representations. Techniques such as Word2Vec (Mikolov et al. 2013), GloVe (Pennington et al. 2014), or FastText (Bojanowski et al. 2017) are commonly employed to generate word vectors, which have become integral to various natural language processing (NLP) tasks. These vectors empower algorithms to comprehend and process language more effectively by representing words in a manner that preserves their semantic relationships.

witness corpus with a pre-trained[26] GloVe model, fine-tuned[27] the model on our corpus of testimonies, and eventually find the closest lemmas in terms of semantic distribution. This refinement ensured that the mathematical linguistic representation aligned precisely with the variant of English found in the corpora we compiled.

Lastly, we computed the lemmas predominantly employed in the context of those initially listed. This process entailed evaluating similarity using cosine similarity between the embeddings of lemmas and words within the corpus. Consequently, we expanded the list of lemmas associated with pain and violence, aligning it more closely with the corpus characteristics. We obtained a second list that encompassed the initial words related to 'pain' and 'violence' and all words with similar linguistic context in the witness corpus. The results are detailed in the following table:

Target Lemma	Similar lemmas
assault	attack, arm, raid, robbery, charges, fire, rape, kidnap
attack	suicide, bombing, raid, kill, ambush, retaliation, blast
axe	ax, knife, hatchet, scissors, spear, machete, batter, blade

---

[26] In the context of humanities studies, a pretrained model refers to a machine learning model that has been trained on textual data before being fine-tuned or adapted for specific tasks within the humanities. These models are typically based on advanced natural language processing techniques and have learned patterns, language structures, and contextual understanding from a wide range of sources. Once trained and loaded into the specific code being utilized, these models serve as powerful tools for humanities researchers, as they can be applied to tasks such as text analysis, sentiment analysis, language translation, and other linguistic inquiries.

[27] Fine-tuning a model in the context of humanities studies involves adapting a pre-existing machine learning model to perform specific tasks or analyses related to the humanities. Instead of building a model from the ground up, fine-tuning utilizes a pre-trained model and adjusts its parameters using a smaller, domain-specific dataset. In the current scenario, the pre-trained Glove model, initially trained on a 6 billion-token corpus and thus equipped with a general knowledge of the English language, underwent fine-tuning on our dataset. This dataset consists of novels and testimonies that document or recount The Troubles. The fine-tuning process enables the model to provide mathematical representations of words tailored to the variant of English found in those texts.

bleed	suck, breathe, inhale, vanish, weep, tremble, wither, swell
blood	heart, bone, tissue, body
bomb	blast, explode, explosive, explosion, detonate, ignite
bombing	suicide, raid, terrorist, assassination, incendiary, improvised
brutalize	oppress, manhandle, victimize, objectify, debilitate, traumatize, terrorize, harrass
bullet	shrapnel, gunshot, wound, grenade, chest, stab, riddle, shotgun, shot
choke	strangle, throttle, hook, shackle, rope
crush	quell, crackdown, rebellion, crack, battle, fight, uprising, rebel, separatist
cry	shout, scream, loud, cut
damage	injuries, impact, harm, severe
death	die, execute, execution, murder, father, brother, victim, child, son
decapitate	behead, mutilate, disfigure, abduct, strangle, enslave, asphyxiate, decimate
explosive	bombs, bomb, detonated, detonate, grenades, weapon, grenade, explosion
fear	worry, anger, danger, worried, threat, afraid, doubt, concern
fire	attack, explosion, blast, blaze, ground, destroyed, artillery, gunfire
firearm	handgun, handguns, weapon, semiautomatic, felon, felony, gun
fracture	tendon, tibia, ankle, bone, ligament, fibula, cheekbone, dislocated, sprain, elbow
grenade	mortar, assailants, propelled, detonated
gun	handgun, rifle, weapon, pistol, assault, firearms, fire, cannon, handguns
harass	intimidate, terrorize, frighten, humiliate, imprison, coerce, persecute
hit	struck, shoot
incendiary	explosive, bombs, flammable

injure	maim, humiliate, terrorize, abduct, demoralize, endanger, kill, mutilate
intimidate	harass, terrorize, frighten, coerce, humiliate, discredit
kill	destroy, shoot, attack, poison, suicide, abduct, kidnap
knife	blade, scissors, stab, sticks, ax, stick, machete, neck, axe
lash	pummel, rip, goad, wreak, shove, rattle, yank, whip
machete	stab, knife, wield, truncheon, ax, hack, shrapnel
maim	injure, mutilate, abduct, terrorize, imprison, sicken, victimize, brutalize
massacre	killings, atrocities, genocide, murders, slay, assassination, murder, bomb, kill
mutilate	maim, decapitate, disfigure, enslave, asphyxiate, mistreat, castrate, injure, brutalize, deface
oppress	brutalize, persecute, subjugate, enslave, demean, terrorize, humiliate, repress, denigrate, victimize
paramilitary	arm, militia, militiamen, army, rebel, soldiers, policemen, police
pistol	rifle, caliber, revolver, semiautomatic, handgun, gun, shotgun, submachine
projectile	launcher, pellet, propellant, mortar, bullets
revolver	pistol, magnum, caliber, handgun, shotgun, rifle
rifle	shoot, caliber, gun, semiautomatic, artillery, carbine, shotgun
scream	shout, cry, yell, shriek, holler, cries, shouts, yells
shotgun	gauge, grenade, pistol, revolver, rifle, handgun, semiautomatic, bullet
shout	yell, cry, scream
smite	asphyxiate, flamethrower, launcher, rifle, revolver, pistol
stab	gunshot, wound, bullet, shrapnel, knife, machete, chest
strangle	suffocate, smother, hobble, humiliate, starve, stifle, choke, cripple

tears	cry, grief, sadness, weep, whip, terror, frighten
terrorize	terrorise, intimidate, harass, demoralize, frighten, terrify, humiliate, oppress, enslave, subjugate
torment	anguish, loneliness, agony, despair, humiliation, misery, endure, nightmares
violate	violation, contravene, infringe, enforce, abide, comply, prohibit, breach, provisions
violence	violent, bloodshed, clashes, conflict, sectarian, fighting, attacks, killings, incidents
wound	chest, shoulder, neck, gunshot, bullet, leg, broken, injuries, wrist
harm	endanger, consequences, threaten, cause, damage, affect, jeopardize, harmful
bruise	contusion, thigh, sprain, ankle, knee, elbow, laceration, scar, fracture, ligament, thrash

You can also use Principal Component Analysis to visualize these results in two subsequent figures (Figure 2.1, Figure 2.2). In Figure 1, we can see the mathematical representation of the lemmas from the original list, along with additional lemmas identified by the GloVe-based model within similar linguistic contexts. On the x-axis, zero represents identical relationships, reflecting the mathematical similarity of words, while the y-axis represents another dimension. It is important to note that the Glove model considers 300 mathematical dimensions to measure similarity, which are eventually reduced to two dimensions through Principal Component Analysis[28] for

---

[28] Principal Component Analysis (PCA) is a statistical method frequently used to analyze and interpret complex data sets. In humanities research, PCA is employed to reduce the dimensionality of large datasets, capturing the most significant patterns and variations within the data. By transforming and identifying key components, PCA helps researchers uncover hidden relationships, similarities, or differences among variables, enabling a more concise and interpretable representation of the data.

visualisation clarity. In brief, the graph demonstrates the similarity in terms of linguistic occurrence between words: the closer the words are, the more frequently they occur.
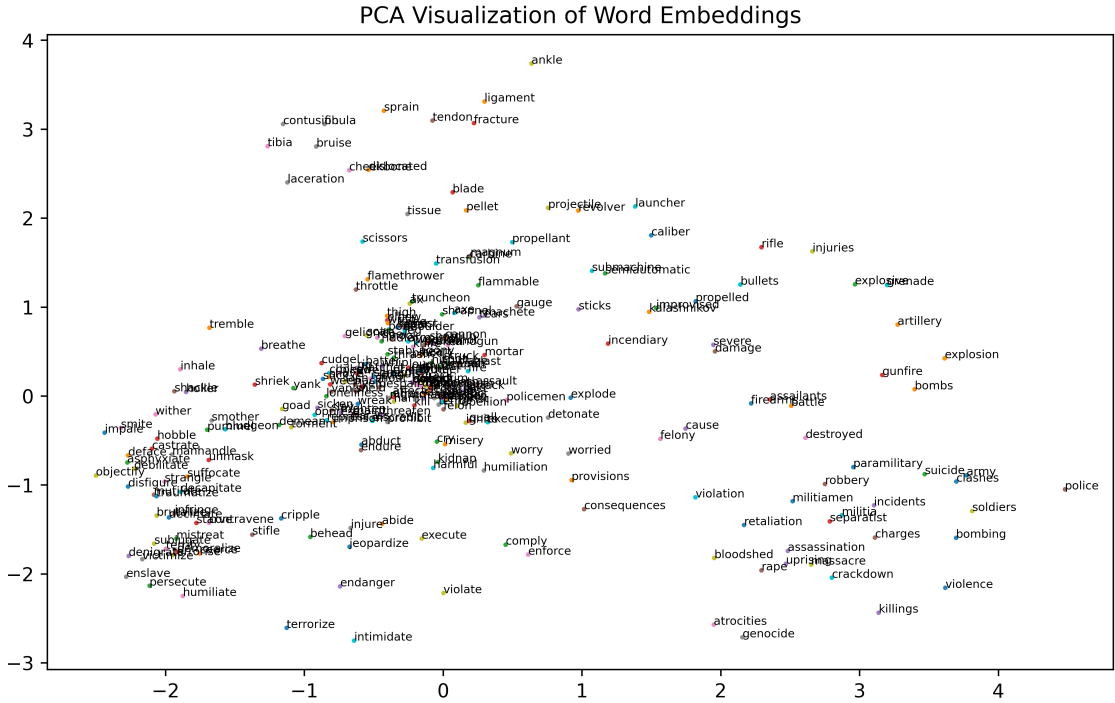
Figure 2.1. Visualization of principal component analysis on glove embeddings fine-tuned with novels and real-world testimonies relating The Troubles.

Upon examining the chart in Figure 1, we can already make some preliminary observations about the expression of pain and violence:

• Long-range weapons, such as 'grenades,' and 'bombings' (located on the right side of the graph), are more likely to be associated with contexts describing general damage ('damages,' 'injuries') and are mathematically closer to systemic events ('bloodshed,' 'violence,' 'killings,' 'genocide'). In addition, long-range weapons are associated with lemmas that involve a systemic act of objectification and humiliation through violence, as expressed by terms such as 'persecute', 'humiliate', 'enslave', (located in the lower-left part of the graph).

• Short-range weapons (such as 'blade,' 'knife,' 'shotgun') appear to be close to the actual descriptions of injured body parts (expressed with words such as 'fibula,' 'ligature,' 'ankle,' situated in the middle-upper part of the graph).

Also, the graph in Figure 1 exhibits a distinct cluster emerging, situated around the centre. Since this cluster of words is challenging to comprehend, we have decided to zoom in on that portion in Figure 2. In this enlarged view, it becomes possible to read the words one by one:
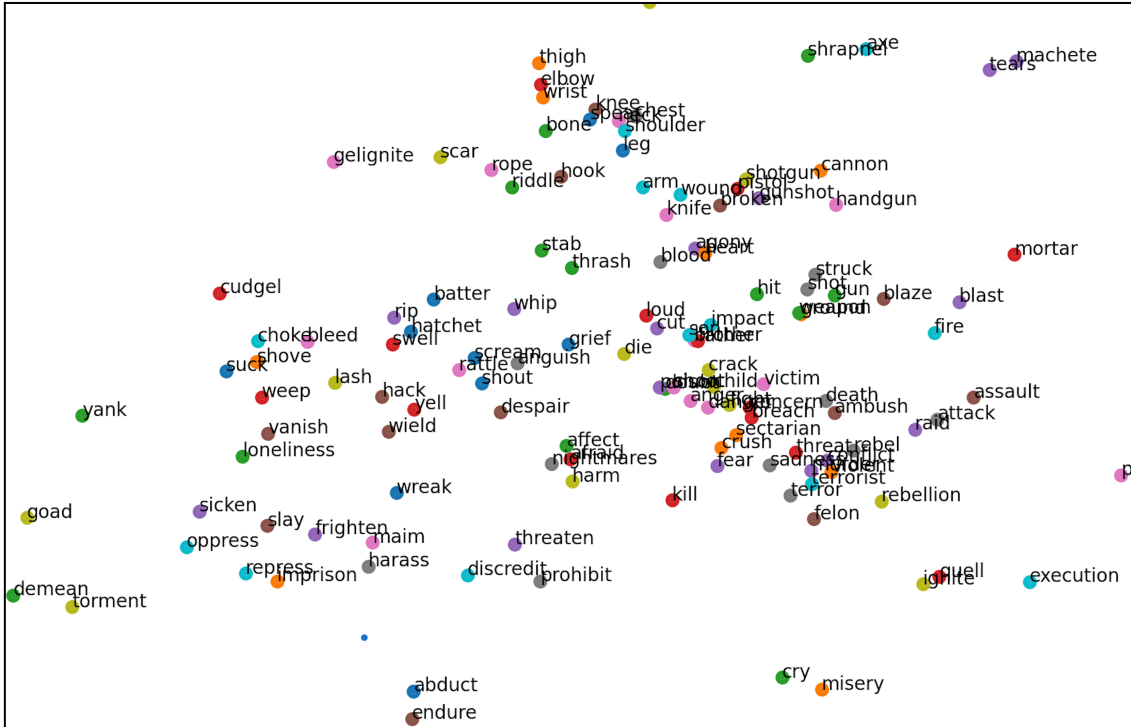
Figure 2.2. Expanding upon the most prominent word cluster in the PCA representation of Figure 1.

Interestingly, in the upper part of Figure 2, we can observe a distinction in the use of verbs describing violence in terms of actionality (Vendor 1957):

• Non-durative/punctual violent verbs such as 'shoot,' 'impact,' 'die,' and 'stab' are semantically closer to the moment of death ('kill,' 'loud,' 'victim,' 'death').

• Durative verbs expressing violence ('batter,' 'hit,' 'choke,' 'shove') are closely associated with psychological states such as 'loneliness,' 'fright,' and verbs expressing the reaction of the victim ('weep,' 'yell,' 'scream,' 'shout').

By intersecting the previously made observations, we can categorize them into three patterns:

1) Punctual violent verbs (such as 'cut', 'hit', 'impact', 'shoot', 'break', and 'stab' are closely associated with describing the moment of death and are specifically linked to particular parts of the body.

2) Durative violent verbs ('batter', 'harm', etc.) play a role in describing the process of experiencing pain and are closely linked to lemmas expressing the emotional responses of witnesses.

3) Long-range weapons are used to describe the chaotic effects of an attack rather than specific outcomes, and they involve systemic acts of violence.

On the other hand, short-range weapons exhibit linguistic closeness with parts of the human body and provide a general description of the event of death.

Having made those observations by examining the data from a mathematical perspective, we subsequently conducted a in-depth manual analysis in context to confirm the findings, as we will explore in the following paragraphs.

2.5.2. Pain #2: Microanalysis of the depiction of pain in testimonies

We will now delve into the examination of patterns within actual passages from the second corpus that we created. This corpus contains testimonies from victims[29] who have been affected by The Troubles. Our goal is to scrutinize the usage of the three identified patterns outlined in section 2.5.1 within their respective contexts.

1) Punctual verbs (highlighted in blue) are more closely connected to the description of the instant of death or to violence seeking death (highlighted in red) and are closely related to specific parts of the body (highlighted in green). Look at the following two examples:

---

[29] The word 'victim' is used here to refer to both individuals murdered by terrorist groups and the relatives of those people. These relatives fall under the category of victims, as they have been affected by the violent events.

They brought him from one club to another club tortured him, took his shirts off and treasures off him… His boxer, shorts off him, hung him up on the beam. Hung… hung… hung up with a rope. They broke his back with pickaxe handle, stabbed in the thighs and the bollocks with a knife. He had severe burn marks in his forehead, they broke his fingers […]. They took him out of a club, brought him to a place called Mercy State, shot three times in the head. My brother suffered terrible death ('Testimony of Richard McCartan', Wave Trauma center).

I was bandaged from my ankles right up. My elbow was broke, my wrist was broke and I'd fractured kneecaps, swollen kneecaps, my ankles were broke and I'd slices all over my arms and legs and my head. I think my eyes were sitting out, and my nose got broke. They just didn't care where they hit me. They just went at me, hoping that they would maim me for life or half kill me – leave me for dead (Marie Smyth & Marie-Therese Fay 1999, p.. 126)

In these passages, the moments of violence aimed at causing death – regardless of the fact that the second victim survived – are described in great detail. Punctual verbs, whose direct object is frequently the injured part of the body, are explicit most of the time. It seems that when the relative of a victim recounts the actual moment of the death of a loved one or an incident of violence intended to be deadly, they feel the urge to be specific, mentioning both the action of injuring and the injured parts. This serves as a coping mechanism to make sense of the trauma.

Indeed, as emphasized by Tuval-Mashiach et al. (2004), following a traumatic event, individuals – specifically those who have been indirectly affected by violence through the loss of a loved one – form a narrative recounting the most significant events. This occurs even when they have not witnessed the events firsthand, as it is an integral aspect of the recovery process. Interestingly, the results of our analysis suggest that we could connect this coping mechanism of narration with actionality (Vendler 1957) by demonstrating that punctual verbs seem to be used more frequently when describing the precise moments of death.

If we agree with Busch and McNamara (2020, p. 328), who point out that «the interconnection of the experience of trauma and the role of language [...] is not causal and mono-directional but complex and multilayered, [...] but literary or biographical texts often provide insights into connections between traumatic and linguistic experience,» we could agree that some traces of traumatic events are indeed reflected in linguistic patterns. In this light, following van der Kolk's (2014, p. 176) view, who asserts that the imprints of traumatic experiences are organized not as a coherent logical narrative but as fragmented sensory and emotional traces, such an actional characterization, i.e. the fact that punctual verbs are more likely to be used when describing the very moments of death, could be linked to the idea that witnessing or imagining those moments could be so impactful and disruptive for the perception of the subject that they end up being more prone to imagine fragments or segments of what happened, rather than the full and iterated process of physical violence.

2) Durative verbs (highlighted in blue), such as 'batter' and 'harm,' play a crucial role in describing the process of experiencing pain. In other words, they are closely linked to lemmas expressing the emotional responses of witnesses (highlighted in red), such as 'be frightened,' 'be harassed,' and 'be oppressed.' Consider the following examples:

> I wasn't physically or mentally forced out. It was more a foreboding that the family could be harmed. Out of thousands, four or five bigots is all that is needed to cause immeasurable suffering (Marie Smyth and Marie-Therese Fay 1999, p. 65)

> He was very angry. He was obviously totally out of control and he could not stop shouting and screaming and moving about. It was very, very frightening to see him so full of anger and pointing his rifle at innocent civilians (*Those Are Real Bullets, Aren't They?*, Peter Pringle and Philip Jacobson, 2000)

While the first passage recounts the climate threats occurring during the period of the conflict, linking the verb 'to harm' with the general «immeasurable suffering» caused by the continuous menace of violence, the second leverages an *-ing* form to highlight the durative semantic value of the verb 'to point'. From a cognitive perspective, the linguistic connection between durative verbs and lemmas expressing emotional responses could highlight that the subject is reimagining the scene in a slower, more detailed manner. This linguistic pattern not only mirrors their exploration of the fragmented nature of moments preceding death but also delves deeper into their examination of the traumatic experience, providing a comprehensive analysis of emotional responses.

As Kearney (2007) suggests, drawing on Aristotle's *Poetics*, «the recounting of an experience through the formal medium of plot, fiction, or spectacle enables us to relive the past in a forward direction» (p. 51). Narrating pain is equivalent to reliving, and hopefully processing, pain to the extent that a witness who reexamines their emotional responses—fear, sadness, disgust, numbness—might even undergo a moment of emotional catharsis. More intriguingly, Kearney posits that «the seemingly unspeakable traumas of death, terror and pain […] are a true test for narrative powers and limit of catharsis» (Kearney, p. 56). This parallels the ability to articulate – meaning to linguistically formalize – the emotional responses to traumatic experiences with the psyche's ability to cope with those experiences: if trauma survivors are unable to express the emotions they have yet to process, then by delving deeper into their emotional reactions, using durative verbs and words associated with emotional distress, they may have progressed to a different phase in their process of healing from trauma.

3) Long-range weapons seem to be more closely linked to systemic acts of violence and are often used to describe chaotic and less specific acts of violence through hypernyms. To illustrate this point further, consider the following examples. The first two exemplify the use of long-range weapons (highlighted in blue) to describe the chaotic effects of an attack (highlighted in red) rather than specific outcomes. The next two examples illustrate the correlation between long-range weapons and systemic acts

of violence (highlighted in green), such as the intent to 'blow people apart' or the deliberate planning of a massacre:

> We actually witnessed the bomb go off. We were on a hill part of the golf course and we could see right across the town and just saw a huge plume of smoke and dust ('Artistic responses to the Omagh Bomb Conversations with Carole Kane and Malachi O'Doherty', Boris Pinto)

> We're looking at all the people outside coming and going, and the next thing, the explosion, the bomb went off […]. Luckily, I only suffered minor injuries and my daughter was all right. But the screaming, the whole effects of it, I'll never forget. And the panic, people trying to get out, looking for ones we did see dead. But it was so confusing, so many people lying about and just the confusion and seeing everything, and the smell, the water running down the street, and then meeting people we knew and that say: «Are you all right?» ('Helen Carr's testimony', Wave trauma center)

Lastly, we present three counter-examples involving short-range weapons (highlighted in blue), underscoring their linguistic contextual connection with nouns denoting parts of the body (highlighted in red) and not involving general or systemic acts of violence:

> And the police told me that I had found a body and they thought it was Chucky and could I make a formal identification. He was shot twice in the head with a son of shotgun and it was a mess. We had to identify him with the tattoos he had on us on his arm. ('Michael Bennet's testimony', Wave trauma center)

> His attackers couldn't find a pistol so they used a rifle instead. He blew off one of his legs completely and left the other one barely hanging on. The men panicked and ran off. An ambulance wasn't called for an hour and the victim bled to death on the street.(Lost lives, p 891)

> The police were coming in and the next thing was… Plastic bullets were fired, and Seamus was shot at point-blank range in the chest. (Kathleen Duffy's testimony, Wave Trauma center)

The correlation between these factors initially seems logical due to the nature of these weapons. It's not surprising that weapons designed for close-quarters combat allow both victims and witnesses to identify and later describe the specific body parts that are injured. In contrast, grenades and bombs do not allow enough time for an immediate recognition of the full extent of the damage, owing to their tremendous force.

However, a more in-depth exploration of this linguistic phenomenon could unveil a crucial aspect. Different weapons involve distinct dynamics between the perpetrator of violence and the victim: when a weapon is employed from a greater distance, neither the victims are in direct contact with the aggressor, nor does the aggressor witness the actual pain they are causing.

This, in turn, allows for a certain degree of disconnection, reducing the personal accountability of the perpetrator – a phenomenon which can be ascribed to a form of a moral disangagement (Bandura 2011), consisting in «psychosocial maneuvers by which people selectively disengage moral self- sanctions from inhumane conduct». By steering clear of the victim's gaze, the perpetrator seeks to distance themselves from violence by avoiding the elicitation of empathy. Empathy, indeed, the affective component of understanding another's emotional state, can be conceptualized in its most basic form as the capacity to perceive another person's immediate emotional state. As Decety and Jackson (2004) argue:

> Developmental research indicates that we are, from birth, not only acting and thinking selves, but we also express an intuitive need to relate ourselves to other people. It has been shown that very young infants express what Trevarthen (1979) terms intersubjective sympathy—that is, they are innately predisposed to be sensitive and responsive to the subjective states of other people. This can be demonstrated through several means, including spontaneous face-to-face interaction between infants and their mothers and

through more specialized 'still-face procedures' […], which can lead to withdrawal by the infant (Decety and Jackson, p. 77).

Face-to-face violent interactions (carried out with short-range weapons) leave the door open for the elicitation of empathy, therefore making it increasingly difficult for the perpetrator to partake in systemic acts of violence, as they are directly confronted with the suffering of their victims. On the contrary, the more impersonal the engagement (as it is with long-range weapons), the more emotionally detached the violence become, heightening the probability of recurring systemic violence, ultimately culminating in massacres, slaughters, and mass murders.

2.6.1. Space #1: Computational Literary Studies and Spatiality

As mentioned in paragraph 2.1, having outlined patterns regarding the depiction of pain in real-world testimonies, our objective is to investigate, within the context of novels related to The Troubles, whether pain is narrated differently and how it intersects with the spatial dimension. We have also posited that spatiality in fiction holds structural significance and is often layered with semiotic values. To elucidate our theoretical framework and provide the reader with a research context in terms of spatiality in narratology and CLC, we now present a concise literature review.

Within the realm of digital humanities and CLC, spatiality has been playing a pivotal role, given the relatively low learning curve of mapping technologies such as QGis[30] and arcGIS[31], and their frequent usage in many digital projects (Foley & Murphy 2015; Jankowska 2021; Duever & McGinn 2020). Spatiality has often been investigated in its

---

[30] QGIS is an open-source geographic information system (GIS) software that provides a user-friendly platform for creating maps, handling spatial data layers, and conducting geospatial analysis. Further information at: https://qgis.org/en/site/ [Last checked 1st of December 2023].

[31] ArcGIS is a geographic information system (GIS) software that enables users to create, analyze, and visualize spatial data. It is widely used in fields such as urban planning, natural resource management, and geospatial analysis. Please see: https://www.arcgis.com/index.html [Last checked 1st of December 2023]

direct connection with non-fictional spaces, with projects that map existing locations mentioned in literary texts. However, the attempts of reflecting on spatial literary theory and benefit from the categories that non-digital literary theory has used on fiction has not been a path widely investigated. There are some notable attempts, such as the Lancaster university project 'Chronotopic Cartographies' (Bushell et. al, 2021), that harnesses the topologic rather than topographic aspects of (literary) geography, or Barbara Piatti's project 'A literary Atlas of Europe' (Reuschel et al., 2009; Piatti et al., 2009), that exploited other literary spatial dimensions (such as character density in relation with the geographic category of population density) and above all reflected of the matters of representation stemming from the ambiguity of literary spaces. Nonetheless, there is still much to be done, despite researchers attempting to computationally investigate aspects of literary spatiality. Some have even endeavored to offer academics a comprehensive spatial ontology, utilizing machine learning to extract and analyze literary spaces. An example is Schumacher (2023), whose aim was to provide digital humanists with new computational-theoretical tools applicable to literary texts.

In this context, it is also notable that critical digital humanities (Grimshaw 2018; Berry 2022) has not played much with the critical value of spatiality conceived in a narratological sense or in its structural value for narrative. As we anticipated, most times space is investigated in its connection with the real world, rather than coined as a category that is inserted in a structural dynamic with other components of fiction, such as time, characters, levels of narration and so on. However, following Westphal's (2011) intuition, that conceives spatiality as the preferable start for a societal reflection towards literature and culture, the category of space should be regarded as a function of literature and investigated in relation to other categories rather than merely represented on a (digital) map.

We posit that digital spatial literary studies should harness the narratological frameworks provided by literary theorists. Among the theorists who have investigated the topic, Genette (1980) emphasized the necessity of a fictional space for every narrative, thereby underscoring the pivotal role of space in storytelling. According to

him, «it is easier to conceive of a description devoid of any narrative element than the reverse, since even the most sober designation of the elements and circumstances of a process can pass for a beginning of description» (Genette 1980, p. 29).

Hamon also provided narratological tools for delving into the study of literary descriptions and proposed a formula for identifying and analyzing descriptions as follows: 'P + F + T-I (N + PR$_q$ + PR$_f$)', where 'P' represents one or more characters, 'F' is a predicate of the type 'acting on,' 'talking about,' or 'looking at,' and T-I is an introductory theme that includes the spatial setting or the object being described. This introductory theme is further elaborated into a nomenclature (N) consisting of elements that stand in a metonymic relation to the introductory theme. In simpler terms, for instance, the description of a landscape is broken down into sub-descriptions of rivers, mountains, flowers, etc. These sub-themes can be expanded through qualifying predicates (PR$_q$) or functional predicates (PR$_f$), which connote the sub-theme in terms of its function or qualitative aspect (Hamon 2004, pp. 340-341).

Another significant contributor to the spatiality debate is Herman, who, among other narratological categories, introduced the distinction between 'figure' and 'ground' (alternatively referred to as 'located objects' and 'reference objects'). The semantic framework of spatial relations can be conceptualized as a dependency between two or more entities, where the former serves as a backdrop, and the latter assumes a position relative to the former.The correlation between these entities, whether in terms of laterality, externality, proximity, etc., is linguistically conveyed through adverbs of place (such as 'here', 'there', 'around') and prepositions ('in', 'between', 'over'). These linguistic elements not only convey information about the entities involved but also encompass general characteristics, enabling us to discern whether an object possesses volume, surface area, among others (Herman 2002, pp. 274-5).

Special mention should be given to Bakhtin (1975), who introduced the now well-known concept of the chronotope, emphasising the inseparability of space and time in narrative. Bakhtin (1975) introduced the concept of chronotope, by saying that: «We shall call chronotope […] the substantial interconnection of temporal and spatial relations of which literature has taken possession artistically. [...] In this term is

expressed the inseparability of space and time (time as the fourth dimension of space)» (Bakhtin 1975, p. 231). A chronotope, in Bakhtinian terms, refers to the intrinsic connectedness of temporal and spatial relationships that are artistically expressed in literature. It is the fusion of time and space, and Bakhtin used this concept to analyze how the representation of time and space in a literary work shapes its meaning.

In the realm of cognitivism, Ferguson & Hegarty (1974) delved into the cognitive aspects of mentally reconstructed maps, identifying three potential symbolic values attributed to the spaces included in those representations: 'anchors', 'paths', and 'landmarks'. 'Anchors' represent key reference points on the overall map, often places of significant emotional importance, such as home, the city center, or a church where pivotal story events unfold. 'Paths' hold value in the context of characters' journeys from one location to another, potentially carrying allegorical significance related to story events. 'Landmarks' encompass the starting or ending points of a character's path.

Likewise, emphasizing the similarity between fictional and natural narratives, Linde & Labov (1975) outline two strategies for constructing the narrative's settings: firstly, the 'map' strategy, involves describing the environment from a static vantage point, typically situated above the depicted space. Here, orientation is established using spatial vectors like 'right,' 'left,' 'above,' and 'below.' The second strategy, aptly termed the 'tour,' dynamically portrays spaces from a mobile deictic center, following the journey of a character navigating these surroundings. In this approach, relative cues like 'further ahead' and 'on the left' are employed, with spatial descriptions being strongly linked to a character or an object.

In addition to commenting on the strategies by which fictional spaces are constructed, other theorists have explored the symbolic values that authors can layer onto these spaces. According to Lotman (1975), narrative space is not just a backdrop for the actions of characters; instead, it constitutes a dimension within the text that reflects a *weltanschauung*. In other words, it recreates a conception of the world, incorporating specific social and cultural beliefs. Lotman classifies all the arts as secondary modeling systems and specifically contends that literary space should be

perceived as a signifying continuum that also organizes other levels of the text. Spaces, therefore, express non-spatial relations and are linked to concepts from different semantic spheres. Environments characterize characters, giving rise to what Lotman terms 'two-level', 'ethical-spatial metaphors' where places are imbued with moral values. Particularly, Lotman discusses the concept of the 'frontier' as a boundary space that divides environments into distinct portions, leading to semantically connoted oppositions such as 'inside'/'outside', 'us'/'others', 'high'/'low', 'good'/'bad', 'city'/ 'country', and 'nature'/'culture'. Frontier spaces play a pivotal role in narrative patterns and textual dynamics; the plot of a novel is often set in motion by the violation of a frontier, and characters can be analyzed based on their relationship to such a boundary space.

In terms of the overall structuring of the narrative space, Torop (2002) outlines three levels of chronotopic analysis: 'topographical chronotopy', which pertains to the time and place within which the plot unfolds; 'psychological chronotopy', concerning the mode of space-time experience as perceived by singular fictional characters; and 'metaphysical chronotopy', which delves into the author's space-time (Torop 2002, pp. 18-19). Ryan (2003, 2012), on the other hand, identified five levels of narrative spatiality: 1) Frame: The immediate environment surrounding an action as it occurs; 2) Setting: The overarching socio-economic and cultural context in which events unfold; 3) Story Space: The collective set of spatial frames, encompassing all environments mentioned, whether explicitly or implicitly referenced by the characters; 4) Storyworld: The narrative space constructed through a blend of cultural understanding and real-life experiences. 5) Expanded Storyworld: Encompasses the original storyworld and incorporates all counterfactual worlds generated through the characters' thoughts, assumptions, dreams, and beliefs.

By drawing on these diverse perspectives, digital spatial literary studies can attain a richer and more nuanced understanding of the interplay between space and narrative in the digital realm.

2.6.2. Space #2: Operationalising the intersection of pain and spatiality

Now, to draw a parallel between linguistic features related to the depiction of pain (extracted as explained in paragraph 2.5.1, where we detailed both the operationalization process and the extraction methods) and spatiality, we first need to identify the specific spatial information we would like to extract. Secondly, we will elaborate on how we retrieve such an information from the texts.

In this context, we argue that spatiality can be effectively framed within the three categories of 'Geo-political entities' ('GPE': countries, cities, states, etc.), 'Facilities' (FAC: mountains, rivers, lakes, forests, etc.), and 'Locations' ('LOC': buildings, airports, airways, bridges, etc.). GPE, FAC and LOC account for the majority of spatial indications from real-world locations, are the ones that are most efficiently retrievable with state-of-the-art machine learning algorithms through a process called Named Entity Recognition (NER), a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories.[32]

Our idea is to utilize machine learning algorithms and NER to trace these spatial indications back from texts. Subsequently, we will employ the narratological tools presented in the previous paragraph to interpret the relationship between spatiality and the depiction of pain.

Framing spatiality within a framework that highlights the presence of real world space is defensible in our case study by arguing that, since we are investigating novels related to a historical event, the locations falling under our interest are those from the real world. Our assumption is that, as our novels depict the settings of The Troubles, and those settings are places that have been subjected to terrorist attacks, they will inherently display a specific relation with pain. This connection will guide us towards

---

[32] Technically, we used the Spacy library for Python, through which we loaded, tokenized, and pre-processed the second corpus (the one including the novels). We then iterated over the sentences of the corpus to recognize spaces and places, eventually storing the spatial mentions in a database along with the sentences in which those spatial indications occurred.

the objective of the chapter, which is to study the intersection between the fictionalised spaces of The Troubles and the depiction of pain.

As stated in 2.5.1, we identified three patterns related to the depiction of pain in testimonies. To explore how these patterns were represented in fiction and examine their correlation with spatiality, we developed three distinct Python scripts. These scripts were designed to identify each pattern and analyze the surrounding spatial indications categorized as GPE, FAC, and LOC.

To guide the Python scripts, we initially categorized the lemmas discovered in 2.5.1 into relevant groups, including punctual/durative verbs, lemmas expressing emotional responses, parts of the body, systemic acts of violence, and short-range/long-range weapons. The categorisations is presented in the following table (Table 2.3):

Body part	'ankle', 'arm', 'blood', 'body', 'bone', 'bruise', 'cheekbone', 'chest', 'contusion', 'doubt', 'elbow', 'fibula', 'fracture', 'heart', 'injury', 'knee', 'laceration', 'leg', 'ligament', 'neck', 'scar', 'shoulder', 'thigh', 'tibia', 'tissue', 'wrist'
Durative violent verbs	'To ambush', 'To asphyxiate', 'To attack', 'To batter', 'To bleed', 'To choke', 'To coerce', 'To disfigure', 'To fight', 'To hack', 'To jeopardize', 'To kidnap', 'To lash', 'To manhandle', 'To mistreat', 'To persecute', 'To rape', 'To repress', 'To smother', 'To starve', 'To stifle', 'To strangle', 'To subjugate', 'To suffocate', 'To threaten', 'To throttle', 'To torment', 'To yank'

Emotional response	'agony', 'anger', 'concern', 'fear', 'grief', 'loneliness', 'misery', 'sadness', 'terror', 'to affect', 'to anguish', 'to be afraid', 'to cry', 'to debilitate', 'to demoralize', 'to endanger', 'to endure', 'to frighten', 'to harass', 'to holler', 'to humiliate', 'to intimidate', 'to shout', 'to sicken', 'to shriek', 'to terrorize', 'to traumatize', 'to tremble', 'to weep', 'to yell', 'scream', 'tears', 'to harm', 'to harass', 'to demean'
Long-range weapon	'artillery', 'blast', 'bomb', 'cannon', 'explosive', 'flamethrower', 'grenade', 'incendiary', 'launcher', 'mortar', 'propellant', 'semiautomatic', 'submachine', 'sniper rifle', 'high-powered rifle'
Punctual violent verbs	'abduct', 'assault', 'behead', 'breach', 'break', 'cripple', 'crush', 'cut', 'damage', 'decapitate', 'destroy', 'detonate', 'die', 'dislocate', 'execute', 'explode', 'fire', 'grind', 'hit', 'ignite', 'impact', 'injure', 'kill', 'maim', 'mutilate', 'rip', 'shoot', 'shove', 'slay', 'smite', 'sprain', 'stab', 'wreak'
Shorts-range weapons	'axe', 'bullet', 'baton', 'brass knuckles', 'club', 'dagger', 'firearm', 'gun', 'gunfire', 'handgun', 'hatchet', 'knife', 'machete', 'magnum', 'pistol', 'projectile', 'revolver', 'rifle', 'scissors', 'shiv', 'slingshot', 'stun gun', 'truncheon', 'whip'
Systemic act of violence	'atrocities', 'bloodshed', 'clash', 'conflict', 'crackdown', 'execution', 'explosion', 'genocide', 'incident', 'raid', 'rebellion', 'to brutalize', 'to decimate', 'to denigrate', 'to discredit', 'to massacre', 'to objectify', 'to oppress', 'to quell', 'to victimize', 'uprising'

After incorporating these lemmas into the three separate Python scripts, and the following paragraphs will provide a detailed explanation of the algorithmic process used to implement the lists into the scripts. Each of three following paragraphs corresponds to one of the three patterns of pain depiction identified in 2.5.1.

2.6.2.1. Punctual verbs and general reference to death

The first characteristic we identified in real-world testimonies of pain accounts was the specific use of punctual violent verbs. These verbs were closely associated with the descriptions of the moment of death or killing, and with parts of the body undergoing violence. In order to uncover how the cognitive patterns found in testimonies were used in fiction, we developed a script that iterated through the corpus, identified punctual verbs like 'cut,' 'hit,' 'impact,' 'shoot,' and 'stab,' and then examined whether these verbs were linguistically linked to bodily semantics and used to describe the exact moment of death rather than the process of dying.

In the case of actual testimonies, our interpretation was that these cognitive structures (the usage of punctual verbs to describe only the instant of death) were employed to acknowledge the loss of a loved one while avoiding excessive reflection on the traumatic event. This avoidance seemed to be linked to an individual's inability to psychologically process the trauma.

Now, we will analyze some examples that our algorithm found in literary texts to determine if similar linguistic correlations were employed in a similar manner. We will highlight punctual verbs that describe pain or violence in blue. The first excerpt is from *Eureka Street* by Robert McLiam Wilson:

> This man read, haltingly but confidently, a poem entitled *Poem to a British Soldier About to die* […]. The poem told the young British soldier why he was about to *die*, why it was his fault, how it had been his fault for eight hundred years and would probably be his fault for another eight hundred, why the man who was going to *shoot* him was a fine Irishman who loved his children and never beat his wife and believed firmly in democracy and freedom for all, regardless of race or creed, and why such beliefs gave him no option but to *murder* the young British soldier (about to *die*).

In this passage, the narrator, Jake, a young Catholic man, serves as one of the two main protagonists in the novel alongside his Protestant friend Chuckie. Jake comments on the oversimplification of attributing all blame for The Troubles to the British government. This political stance is implied in a poem being read aloud by another character, in a scene narrated with a comic style. Punctual verbs are used to describe the moment of death without delving into the details of the death process.

However, in this narrative, the linguistic pattern that links punctual, violent verbs to an inability to confront the reality of death is mediated by a three-tiered narration. Initially, Jake (the first narrator) relinquishes the narrative spotlight to an impromptu translator (the second narrator) who happens to be translating the poetry of a nationalist poet (third narrator). Through this process of adaptation, Jake employs irony, a linguistic device that is not typically found in actual testimonies where humor or any form of comic discourse is usually avoided. In other words, the discerned pattern within testimonies exhibits is intricately imbued with an ironic sense, a characteristic made feasible through the medium of fiction, which possesses the capacity to observe pain in a manner distinct from that of real-life experiences, mitigating the extent of personal involvement on the part of the witness or victim.

The second excerpt comes from Jake Higgins' *Angel of Death*:

> McGurk looked stunned: «It can't be.»
> «Oh yes it can, old son,» Dillon told him, bringing up the Browning and *firing* through the open door, knocking McGurk on his back, then swinging and *shooting* the man beside him through the head. The man at the wheel of the Rover pulled forward, drew a pistol and *fired* through the open passenger window, then put his head down and took off. Dillon *fired* twice at him, shattering the rear window, but the Rover turned the corner and was gone.

In *The Angel of Death*, the literary narrative unfolds the tale of Grace Browning, a mysterious former actress turned assassin, associated with the 'January 30' terrorist group. This woman plays a crucial role in preventing the assassination plot against Protestant terrorist Quinn, a mission led by the former IRA operative Sean Dillon (a recurring character in Higgins' novels) and his ally Hannah Bernstein. The reported excerpt commences with the impending victim, McGurk, who anticipates Dillon's arrival before witnessing the dramatic scene of Dillon using a Browning to eliminate two individuals. McGurk is convinced that Dillon has tracked him down, and the ensuing dialogue is vividly cinematic.

The punctual actionality of the verb 'to fire' reinforces the connection between this semantic actionality feature, the mention of precise body parts, and the general depiction of the moment of death, as found in 2.5.1. Fiction, though, not only inherits the linguistic function of this actionality feature but also, in the case at hand, introduces a pictorial aspect that, while non-essential for conveying information, becomes indispensable for the graphic details that captivate the reader. Similarly to the preceding example, the patterns observed in testimonies have been inherited but transformed to serve a linguistically layered purpose.

The third example comes from Davies' *The drumbeat of Jimmy Sands*:

«You and the bastard who injured Liam. I am going to find him and *kill* him for what he did to Liam and then I am going to *kill* you for being a cold-hearted bastard who can't understand a father's love for his son.»

(*The drumbeat of Jimmy Sands*, Murray Davies, p. 202)

The narrative unfolds through the perspectives of two soldiers, Fife and O'Keefe, both hailing from challenging backgrounds – one born in Glasgow and the other in Belfast. In the section of the novel from which the excerpt is extracted, O'Keefe is portrayed confronting a British soldier who has abducted his son, Liam. The specific excerpt reports O'Keefe's thoughts as he engages in conversation with the British soldier.

The verb 'to kill' in this context signifies the imagined moment where O'Keefe, serving as the narrator, contemplates seeking revenge and ending the life of the soldier suspected of kidnapping his son. Interestingly, the use of counterfactual violence and hypothetical narrative plans diverges from actual testimonies: testimonies, typically recounted after a violent event has transpired, seldom involve victims publicly expressing a desire for revenge or indulging in the imagination of hypothetical violent retaliation. This departure of fiction from reality underscores a unique aspect of storytelling: in the realm of fiction, the narrative can articulate emotions and scenarios that would be impossible or implausible in real-life testimonies.

The fourth example comes from Park's *Stone Kingdoms*:

> There were a dozen other bodies visible, and probably the same number in the surrounding bush, those who had been pursued and *executed* or who had crawled away to *die*. We saw a mother and her baby, the child still enfolded in her arms, probably *killed* by the same bullets; an old man frozen on his knees like a toppled statue; two women, *raped* and *mutilated*, their clothes strewn around like rags.

> (*Stone Kingdoms*, David Park, p. 134)

In the novel, the protagonist Naomi is haunted by The Troubles in Northern Ireland and finds herself confined in a land between mountains and the sea. Seeking an escape, she decides to leave Ireland and venture to a refugee camp in the heat of Africa, changing the familiar city streets of Belfast for the arid desert.

In this passage extracted by the algorithm, Naomi witnesses a massacre in a central African camp. This horrifying event triggers memories of the violence from The Troubles, a past she had hoped to leave behind. The narrative navigates through Naomi's experiences, allowing us to witness her walk though the dead bodies step by step.

In a manner reminiscent of actual testimonies, Park seems to be portraying a character who avoids confronting the psychological despair of death: the use of punctual verbs ('execute', 'die', 'kill', 'rape', 'mutilate') to capture the moment of

death without delving into the details of the killing process, depicting the character's struggle in facing the harsh reality of death. This time, it seems that fiction is almost directly drawing from what we find in testimonies, as per linguistic usage. However, beyond aligning with our interpretation that the use of punctual verbs is intertwined with an inability to process trauma, this cognitive-linguistic pattern is employed to make the moment almost a-temporal. The rest of the description utilizes semantics (referencing the verb 'to enfold' or the similitude «like a toppled statue») slightly higher in terms of diaphasia than the register usually found in testimony. The pattern of interest intersects with an aesthetic aim.

The last example comes from Petit's *The psalm killer*:

> In the general confusion there were occasions when the army thought it was fighting the UDA only to discover afterwards that it was their own undercover soldiers they were *shooting* at. In one such battle, which Candlestick heard about from Lena, over a hundred rounds were *fired*, two soldiers were *shot* and a wounded plainclothes man was seen being *executed* with a bullet to the head.

(*The psalm killer*, Chris Petit, p. 100)

The narrative in set in mid-1980s Belfast, centering around the Detective Inspector Cross. He becomes entangled in a complex web of death, each murder executed by the Psalm Killer, a murderer who taunts Cross with biblical messages. In the specific excerpt, Candlestick, the murderer, is being made aware by the prostitute Lena of a clash where the UDA accidentally killed two of their undercover soldiers. The depiction of violence through punctual verbs, once again, does not dwell excessively on the specifics of the deadly events. The literary intent is to introduce another narrative layer, broadening the fictional universe by recounting other violent episodes known to the inhabitants of Belfast, episodes that have become normalized in their awareness.

From the analyses of these excerpts, it appears that fiction is incorporating the interplay between punctual actionality and the depiction of death coupled with the absence of specific details. Yet, as to the psychological aspect discussed in section 2.5.1,

this pattern does not always align with an inability to confront trauma. This is because fiction, as it explores possible worlds (Pavel, 1975), does not always portray characters as firsthand witnesses to violent events. Characters may instead be informed of gruesome details about others' deaths, imagine hypothetical violence, and articulate a desire for physical revenge – thus using the same verbs used when describing deadly moments, but not to account for an occurred experience. As a linguistic consequence of this, the pattern identified in real-life testimonies slips into fiction, serving a variety of linguistic purposes, ranging from comedic and ironic expressions to cinematic nuances in the representation of violence.

Now, let's explore the interaction between this pattern, i.e. the pattern linking punctual verbs and descriptions of death that do not dwell in gruesome details, with spatiality. We'll start with an example from Silva's *The Marching Season*, where punctual verbs (marked in blue) coexist in the same context as geopolitical entities (marked in red):

> There was also evidence to suggest that its leaders [the leaders of Ulster Freedom Brigade, ed.] would go to extraordinary lengths to safeguard internal security. Charlie Bates, a Protestant suspected in the murder of Eamonn Dillon, had been discovered *shot to death* in a barn outside *Hillsborough* in *County Armagh*, and the bombers in *Dublin* and *London* had both *died* in the explosions—a fact that had not been made public.
>
> (*The marching season*, Daniel Silva, p.113)

The narrative is set in the initial phase of the peace process in Northern Ireland. In a singular night, a faction of dissenting Protestant extremists tries to rewind time through three ruthless acts of terrorism—executed in Belfast, London, and Dublin, each designed to disrupt the peace and solidify Ulster's connection to the United Kingdom. At the story's core is Michael Osbourne, a retired CIA officer, compelled to rejoin the Agency due to the nomination of his father-in-law, former U.S. Senator Douglas Cannon, as the new American ambassador to London, who Michael discovers as targeted by the conspiracy aiming for his execution. In the selected passage, Michael

receives information from a colleague indicating that the Ulster Freedom Brigade is suspected of murdering internal members to safeguard its own security. The use of geo-political entities (such as the town of Hillburgh, and the cities of Dublin and London) suggests a location of violence in broad terms and from a broader geographical perspective. Specific details, such as street names or location-specific spots, are absent, allowing the character receiving the information – along with the readers – to generally grasp the events. This further emphasizes the use of punctual verbs to convey death in a general sense, rather than delving into the intricate details of the violent process. It involves a general depiction or mention of the moment of death, coupled with wide-ranging spatial indications.

Yet another illustration, drawn from Clarke's *Soldier E - Sniper Fire in Belfast*, further substantiates our argument:

> During that meeting, every word of which was picked up by the probe, Quinn confirmed that Captain Dubois and Lieutenant Cranfield together had made numerous illegal trips across the border in *Eire* to snatch wanted IRA men and bring them back to *Northern Ireland* to be *'captured'* by the RUC and *imprisoned*. Because of this, it was widely believed that the two men, and certainly Lieutenant Cranfield, *had killed* O'Halloran as an act of vengeance for the deaths of ten invaluable 14 Intelligence Company 'Freds', or turncoats, and the subsequent suicide of Corporal Phillips.

(*Soldier E - Sniper Fire in Belfast*, Shaun Clarke, p. 146)

Set in the 1970s, a tumultuous era marked by frequent sectarian violence, the narrative unfolds against a backdrop of terrorist groups financing their operations through acts of robbery, fraud, and extortion. The focus of the book lies on the Special Air Service (SAS), a crucial component of the British Army portrayed as an indispensable force with highly skilled soldiers capable of operating covertly to counteract terrorist activities. In the given passage, punctual verbs ('capture', 'imprison', 'kill') are once again employed to portray violence in broad terms, steering clear of delving into specific details. Here, the violent event – abduction, rather than

murder, representing a distinct form of violence – is recounted from one character to another. The communicative goal in this fictional scenario is to convey information. Once again, the emphasis on locations is geographically broader, reflecting a higher hierarchical perspective. Counties and countries are highlighted, as opposed to neighbourhoods or specific buildings. The goal is to apprise the listener of the overall context of these actions, rather than eliciting empathy for the specific act of violence in its specifics.

In conclusion, upon a comprehensive analysis of the connection between the primary feature identified in actual testimonies – specifically, the linguistic relationship between punctual actionality, the general depiction of death avoiding specific details (with its occasional link to an incapacity to process death trauma) and spatiality, we have discerned an intriguing pattern. Geopolitical entities such as cities, counties, and centuries, especially when frequently mentioned within the same linguistic context, tend to be associated with broad descriptions of violence, deliberately steering clear of graphic particulars.

Leveraging the spatial categories outlined by Ryan (2003), we can clarify this tendency in narratological terms by asserting that when portraying the moment of death using punctual verbs, the emphasis is placed on depicting the storyworld. Meanwhile, the spatial frames tend to remain blurred or not excessively characterized, aligning with a cognitive interpretation of this first pattern within the context of the challenges associated with processing trauma – in other words, the traumatic nature of death tends to blur the specificity of spatiality.

This spatial inclination could also be rephrased by highlighting how fiction employs the aforementioned spatial strategies discussed by Linde & Labov (1975), as indicated in the relevant footnote (please see paragraph 2.6.1). When employing punctual verbs to depict the general moment of death, narratives tend to favor the 'map strategy' in portraying the story world. In fact, when utilizing the 'tour strategy', as seen in the excerpt from Parks's *Stone Kingdom*, precise spatial indications become blurred.

This pattern, which combines the portrayal of pain and spatial elements, is more characteristic of fiction than testimonial texts. This can be elucidated by considering the

structural requirements of spatiality in fiction. In the realm of fiction, writers aim to construct a more organized world, intending to lead readers through the narrative. Even in cases where the scene is depicted with internal focalization[33] and the protagonist, who serves as the focal point, is too traumatized to articulate the specific spatial details they observe, the narrator must find a way to anchor the events. In other words, if the specificity of spatial frames cannot be provided due to the character being portrayed as overwhelmed by trauma within the narrative, the narrator resorts to offering general spatial indications to illustrate the story world by referencing geopolitical entities, such as cities and Counties.

This hypothesis will gain additional support in the following paragraph, which delves into the connection between durative verbs and precise spatial indication, as the paragraph will argue that durative verbs, contrasting with punctual verbs in terms of actionality, are linked to acts of violence that involve a more detailed description of the dying process and greater specificity in depicting violence.

2.6.2.2. Durative verbs and emotional responses to pain

Moving on to the second characteristic, we initially algorithmically traced back the lemmas of durative verbs, selecting them from the list compiled in paragraph 4.1 (marked with blue). We then refined our chosen results by considering linguistic contexts that included lemmas expressing emotional responses (marked in purple) or describing different types of violence in detail (marked in green). Only in the final stage, as we will see in the concluding part of this paragraph, did we assess the spatial aspects. Let's examine the results, beginning with examples where durative verbs were linked to emotional responses. Here are the findings:

---

[33] As for the narratological category of 'focalisation', also definable as 'focal point', refer to Bal, (2009, p. 146). Focalization refers to the narrative technique or perspective through which a story is presented to the reader. It involves the selection of a specific point of view from which the events of the narrative are conveyed.

She had a broken and lacerated upper left arm and a nasty cut on her head. When the helicopter paramedic saw the head wound—heads *bleed* a lot—he brought her here as a precaution. We ran a complete head protocol on her, and she's fine.

(*Patriot Games*, Tom Clancy, p. 446)

Patriot Games, one of Clancy's bestselling novels in the series, revolves around Jack Ryan preventing a terrorist attack by the Ulster Liberation Army (ULA) on the Prince of Wales and his family. This heroic act earns him the admiration of the entire nation but also incites the enmity and hatred of the ULA, making him a target for terrorism. In the selected excerpt, Jack Ryan's wife has been involved in a car accident resulting from an attempted act of revenge again Ryan himself.

In the provided passage, the durative verb 'to bleed' is utilized to depict the sustained consequences of a violent impact, coupled with a detailed description of the specific injuries sustained by Ryan's wife. This discovery supports our interpretation that the witness is more adept at handling trauma and violence when recounting events with durative verbs. Such actionality expresses the duration of the (violent or painful) actions. Indeed, in the excerpt, it is the paramedic character who observes the wound, and their professional background makes it easier for them to assist and cope with such violence. The next example come from Eastermann's *Day of Wrath*:

The man, half choking, nodded again. Slowly, Abu Hida withdrew his hand, pressing the Peskett in harder as he did so.
«You fucking Irish bastard,» the man grunted.

(*Day of wrath*, Daniel Easterman, p. 198)

The story revolves around the events occurred in a prestigious Muslim Leaders' Conference held in Belfast. While security measures are stringent, they prove insufficient: Irish terrorists claim the lives of four British agents, all transpiring under the gaze of a British army unit. Declan Carberry, an Irish agent grappling with a

drinking problem and a failing marriage, shoulders the responsibility of unraveling the mystery of the identity of the perpetrators behind the violence. However, when the terrorists not only murder his daughter but also frame him for conspiracy, Carberry discovers that the web of corruption extends further up the chain of command than he initially believed.

In the selected passage, Abu Hida, a Muslim soldier seeking vengeance for his fallen companions, employs a Peskett[34] to compel a man to disclose the positions of other sentinels in a war post he is attempting to infiltrate. The chapter from which this excerpt is drawn revolves around heightened tension as Abu Hida endeavors to avoid being discovered. Contextually, the use of a durative verb serves the purpose of eliciting identification from the reader regarding the tension of the situation and describing Hida's menacing violence within such a heated context. Here, the durative violence serves a narrative function: it is essential for making the reader perceive the tension, and the description of violence, rather than being purely informative, aims to depict the narrative situation within a specific heated context. A cold, punctual verb describing violence in general would not convey the tension that the passage aims to create.

The next passage, from Anna Burn's *Milkman*:

> Nothing stops them,' she said, 'till death stops them. You'll regret it, daughter, finding yourself ensnared in the underbelly of all that alluring, mind-altering, unruly paramilitary nightlife. It's not all it seems. It's on the run. It's war. It's killing people. It's being killed. It's being put in charge. It's being beaten. It's being tortured. It's being on hunger strike.

> (*Milkman*, Anna Burns, p. 63)

Milkman is set in 1970s Northern Ireland during The Troubles, focusing on an unnamed 18-year-old girl, only mentioned as 'middle sister.' She becomes the target of Milkman, a paramilitary officer, leading to a series of unsettling events. Despite her efforts to avoid him, she eventually accepts a ride, only for Milkman to be killed the

---

[34] A Peskett is a close-combat weapon, a British special forces device from the Second World War.

next morning by British forces. The next day, another stalker threatens her later, but she's rescued by women in a club bathroom, marking the end of her ordeal and a return to normalcy.

The specific passage in question involves the narrator, the middle sister, recounting her mother's advice against marrying a 'rebellious man' affiliated with a paramilitary organization. In this narrative, durative verbs such as 'to beat' and 'to torture' are juxtaposed with punctual verbs (or periphrastic expressions) declined in the *-ing* form, such as 'being killed' and 'being put in charge,' emphasizing the prolonged nature of the actions. This choice aligns with the mother's intent to vividly convey to her daughter the immense difficulties she would encounter if she chose to marry someone associated with a paramilitary group. Once again, the passage confirms our hypothesis: the use of durative actions is intricately connected to a detailed depiction of violence, serving various communicative purposes and contrasting with the fragmented or general mention of deadly moments linked with the use of punctual verbs.

The next passage, from Bradby's *Shadow Dancer*:

> I didn't shoot unarmed protesters demanding their legitimate rights. I didn't bash down people's doors and get them out of their beds in the middle of the night. I didn't *torture* people and *beat* them and *harass* them…'

> (*Shadow Dancer*, Tom Bradby, p. 272)

The books follow Colette McVeigh, a widow and former IRA member turned MI5[35] informer. After being arrested in London, she faces a choice: reveal information and see her children again or remain silent and go to prison. Guided by her MI5 handler, David Ryan, their journey back to Ireland challenges trust and loyalty, especially as Colette puts herself in increasing danger to fulfill her part of the deal.

The extracted passage hold particular relevance for our arguments. It depicts a poignant moment where Ryan and Colette grapple with the harsh realities of war. Ryan

---

[35] The Security Service, commonly referred to as MI5 (Military Intelligence, Section 5), serves as the UK's domestic counter-intelligence and security agency.

has just learned that Colette played a role in the bombing that led to the death of his uncle. He points accusatory fingers at her, eventually labeling the conflict as «her war». Colette, in turn, reacts with anger, forcefully pulling his shoulder and, catching him off balance, causing him to tumble to the ground. She vehemently asserts that she never desired any of the tragic events that unfolded.

The passage serves as a prime example of the intricate relationship between the portrayal of violence, its emotional impact, and durative actionality: Colette employs verbs such as 'to torture,' 'to harass,' and 'to beat' to articulate the pain she witnessed due to sectarian violence, underscoring its weight and lasting impact. The heightened tension is accompanied by a plethora of emotional cues: while not explicitly featured in the algorithmically selected passage mentioned above, these emotional indicators are prevalent throughout the entire paragraph, shedding light on the profound suffering experienced by the relatives of the victims.[36]

Now, let us direct our focus to the intricate relationship between this linguistic phenomenon – specifically, the connection between a detailed depiction of violence, encompassing its emotional repercussions, and durative actionality – and spatiality. Durative verbs are highligthed in blue; spatial indications are highlighted in red, and emotional repercussion of violence are marked in green.

To illustrate our points, we will begin with an example drawn from Shaun's The *Whore Mother*:

> He gave his mind to the street and the traffic. His hands were wet and slipping on the wheel, his face was pouring sweat, his neck was *bleeding* behind from a smarting cut. He swung on two wheels into *Elmwood* and down the *Lisburn Road*; left into *Dublin Avenue*.

---

[36] The emotional cues can be perceived from direct speeches preceding and following the passage algorithmically found. First, the words spoken by Ryan: «You're damned right you didn't know who he was! Well, I'll tell you, he wasn't a soldier to me and my mother»; *«*Two tiny children. […] And do you want to know what they do now? They still cry for their father in their bloody sleep,» *Shadow Dancer*, p. 272; second, Colette's response: «He's not the only one. I didn't notice any of your people crying when they shot down Davey in a stinking churchyard,» *Shadow Dancer*, p. 272.

The Whore-Mother narrates the journey of John McManus, driven by a moral compass to join the Provisional IRA. However, as he becomes deeply entangled in a world of senseless violence, he desperately want to break free from its clutches. Fleeing from IRA pursuers, he begins a desperate journey across Ireland, attempting to escape the repercussions of his tumultuous past.

The computationally identified passage recounts a scene in which McManus endeavors to escape by car from IRA assailants seeking to harm him. The text portrays the severity of the situation, detailing the bleeding cut on his neck, mentioned within a paragraph that, rather than emphasizing the physical aspects of pain, recounts the heightened emotions of fleeing by car from his assailants. The description underscores the McManus's feelings by depicting particulars such as anxiety-induced sweat or the wheel slipping from his hand due to haste.

In terms of spatiality, the passage accurately pinpoints the sequence of events by specifying the names of Dublin streets, such as Lisburn Road, Dublin Avenue, and Elmwood. As anticipated at the end of the preceding paragraph, it appears that a connection arises between durative actionality, the detailed depiction of violence, and the deliberate incorporation of urban facilities or specific spatial indications into the unfolding drama.

Let us move to another example:

> In the end it did not come easy to him as he was by no means a person of natural cruelty. They *lifted* this Taig on the Newtonards Road who *struggled* against them which meant the use of blunt force. The fact was that this put his nerves on edge.

Resurrection man narrative unfolds in Belfast, where a group of ruthless Protestant criminals led by Victor Kelly commits acts of brutality. The main character, raised in the

middle of a rigid, oppressive mother and a father characterized by mysterious silences, immerses himself in American gangster films: his sole reality, coupled with a deep-seated animosity towards Catholics, revolves around a cinematic imagery that he obsessively strives to emulate. Authorities refrain from delving into the matter, leaving journalists Ryan and Coppinger as the lone investigators seeking understanding.

In the provided passage, three gang members – Victor, Big Ivan, and Willie – are in the process of carrying out a murder on a «Taig.» The use of two durative verbs ('to lift' and 'to struggle') conveys the act of seizing the victim by force, and then the victim's violent attempt to break free. The specific location ('Newtonards Road') is employed to vividly illustrate the setting, further emphasizing the connection between a particular depiction of violence, surroundings, and the mentions of precise spatial indications. Following this line of reflection, let us now turn to a counter-example to support our thesis:

> A quick leap through the war of independence followed, and then the focus went on to accounts of discrimination in *Northern Ireland*, sectarian oppression and what were euphemistically described as spontaneous outbursts of anger by the nationalist people. There were shots from a play showing brave women *being battered* by evil gloating policemen and the whole thing finished with one of them holding up a clenched fist and declaiming: 'We may be the most oppressed people ever, but we'll proudly gain our freedom in the end.'

(*The Anglo-irish Murders*, Ruth Dudley Edwards, p. 96)

The narrative of *The Anglo-irish Murders* revolves around the occurrences at a conference where, somewhat unwisely, the British and Irish governments appoint Baroness Troutbeck, a tactless and impatient figure, to preside over discussions on Anglo-Irish cultural sensitivities. When a delegate falls from the battlements, the circumstances surrounding the incident remain uncertain, with neither the authorities nor anyone else able to determine whether it was accidental or intentional. The subsequent death presents a similar dilemma, leading opposing factions to accuse each

other of murder while politicians attempt to downplay the situation in the pursuit of maintaining peace.

The selected passage centers around a moment in the fictional conference, where a film is being screened. The documentary focuses on the Victorian period and the victims of the Famine, briefly revisiting earlier hardships such as the penal laws, Cromwell's sack of Drogheda, and mass hangings after the 1798 rebellion. In other words, it explores how the past has led to the current situation of The Troubles. In this depiction, the durative verb 'to batter' takes on an *-ing* form to emphasize the progressive nature of the violence.

However, the central focus of the passage is not on narrating the violence of The Troubles. Instead, it delves into more distant form of violence that emotionally reconnects with this tumultuous period. The author does not intend to evoke reader empathy, and consequently, locations are presented in a broad sense, with spatial indications limited to 'Northern Ireland.' In other words, the crucial characteristic here is what Ryan (2003) defines as the spatial setting, distinct both from spatial frames and the story world. Spatial indication is mentioned merely because of the ironic similarity between the space of The Troubles and the space of the Famine period. The irony lies in the fact that the documentary within the novel was intended to reconcile the two sides but eventually portrayed the violence perpetrated by the British during the rebellion as a singular outburst rather than something systemic, with this misrepresentation of facts resulting in escalated tension at the conference.

In conclusion, our interpretation for the intersection between the second patter retrieved from actual testimonies – which is the linguistic pattern blending durative verbs, particularised depcriptions of violence and emotional responses to pain – and literary spatiality, is that durative verbs and detailed description of violence are accompanied by precise spatial indications clearly describing the spatial frames only when the literary aim is to arouse empathy from the reader.

This could hint that fiction, could actively engage in the usage of the two patterns identified from real-world testimonies to present trauma (when, with punctual verbs, death is recalled but not precisely detailed) or help cope with it (when, with durative

verbs, death and violence are deepened to the point of delving into the goriest details and into the emotional reponses to those acts).

If we align ourselves with Martha Nussbaum's arguments (1986), which aim to dismantle the well-established sharp division between literature and philosophy, we acknowledge that: A) Literary texts uniquely appeal to «emotion and imagination»; B) that our interaction with literature in general, and tragic poetry in particular, indisputably involves an emotional response as a «fact;» C) that this emotional response either constitutes or contributes to cognitive activity and knowledge (pp. 15–16). By attributing significant cognitive value to the emotional responses associated with transforming historical pain into fiction, Nussbaum could be reinterpreted to suggest a connection between evoking empathy and utilizing spatiality to more vividly portray spatial frames.

In this interplay, there is an evident intention to acquaint the reader with pain, enabling them to relive and ultimately transcend it, especially when spatial indications are fine-grained and strictly connected with real-spaces where the violence of The Troubles have been unfold – from Derry/Londonderry to London, Belfast, Omagh and many other infamously known places.

In such a theoretical context, the interplay between linguistic patterns depicting pain and violence and spatiality assumes new semiotic value. As Keen (2006) cleverly points out while debating a theory for narrative empathy, «character identification often invites empathy, even when the fictional character and reader differ from one another in all sorts of practical and obvious ways, but empathy for fictional characters appears to require only minimal elements of identity.» Having argued that the various focusing possibilities offered by spatiality and the degree of detail with which spatial frames could be depicted are linked to the possibility of identifying with a character (the greater the spatial details, the more easily imaginable is the situation in which the action is unfolding), we can now offer an interpretation for the first two patterns analyzed.

By leveraging the possibility to actively focus on smaller or larger spaces through focalization, fiction could be one of the factors helping deepen the arousal of empathy

from the depiction of pain, thus activating a specific cathartic function. Ultimately, this could serve to move beyond the pain of The Troubles.

2.6.2.3. Long-range weapons and systemic acts of violence

Regarding the third and last characteristic pertaining to pain identified from real-world testimonies, which involves the intersection between the type of weapons and the type of violence described, we initiated the process by identifying long-range weapons and their connection to systemic violence.

Our script initially processed the novels to identify long-range weapons. Subsequently, it filtered the results once more, focusing specifically on excerpts where this type of weapon (highlighted in *blue*) was associated with indications of systematic violence (indicated in *red*). As is customary, the spatial aspects will be analyzed at the end of the paragraph. As we noted in the macro-analytic part of our analysis, long-range weapons were more easily associated with systemic violence. To support our findings and discuss them, the results could be presented through the following exemplary passages:

> Sergeant Harris was the driver, with Cranfield sitting in the front beside him. […] It had been a rather bad year: the humiliating fall of the Tory government; the creation of a non-elected, supposedly neutral power-sharing executive to replace direct rule of Ulster from London; the collapse of that executive under the *intimidation* of the Ulster Workers' strike and IRA *violence*, including the horrendous Birmingham pub *massacre*; the Dublin *bombing*; an IRA truce through Christmas and New Year of 1974-5, and finally the collapse of that truce.
>
> (*Soldier E - Sniper Fire in Belfast*, Shaun Clarke, p. 20)

The first passage is extracted from Clarke's *Soldier E*, whose storyline has been previously introduced in paragraph 2.6.4. In this particular excerpt, two sergeants from MI5 are engaged in a discussion about recent events related to The Troubles, recounting how, despite a ceasefire between the IRA and the British Army, the year has proven to be quite challenging. The Dublin bombing is mentioned alongside other massacres, being portrayed as an instance of the overall violent behavior exhibited by the IRA.

Notably, the reference to the bombing lacks a specific description of the events; instead, it is integrated into a broader framework that outlines the violent situation in which these events are unfolding. Let's move to the second example:

> His head was spinning with new ideas and he saw vividly in his mind the flashes of *explosive* and *the chaos and destruction he would take to London*. Only then might he play a part in achieving a just peace, the cause to which he had devoted and sacrificed so much of his life. He poured himself a straight Black Bush and thought how much he loved his daughter, how much he owed her.
>
> (*The tick tock man*, Terence Strong, p. 51)

The narrative follows Tom Harrison, also known as the 'Tick Tock Man,' as he undertakes the risky task of disarming terrorist bombs. During his journey from Belfast to London to deactivate explosive devices, his abilities face a severe challenge due to a fresh IRA bombing initiative. This campaign distinguishes itself from previous ones, with each device proving increasingly lethal and sophisticated, posing greater difficulties for him in rendering them inert.

The chosen passage centers on Dougan, an IRA member, delighting in the idea of placing a bomb in London. The section just before this excerpt details Dougan honing his bombing skills and acknowledging the inevitable occurrence of casualties. Here, fiction presents something impossible in testimonies: a (positive) emotional response from a potential perpetrator of violence. Once more, long-range weapons are embedded within a theoretical context of widespread and systemic violence. Bombing, in this scenario, is employed to further the overarching goal of instilling a sense of terror across the entire population, thereby coercing the British government onto a specific course of action. Another passage from the same novel, exhibit an analoguos pattern:

> There had suddenly been a lot of gunfire and he knew something had gone wrong, sensed it. He wondered whether there had been a shortfall in *CS*

gas. They'd discussed it earlier, how the mullioned steel windows would make a *heli-assault* difficult, even if the sills weren't rigged with *explosive*.

<div align="right">(<i>The tick tock man</i>, Terence Strong, p. 468)</div>

In this second passage algorithmically found, Harrison is trying to get into a building where IRA people is holding as an hostage his son Archie. The type of violence being recounted is an assault carried out from helicopter to the ground, by launching gas grenade and then let the SAS entering the building as a team. Here, in the chaotic events being depicted, there is no space for singular wounds to be described; the focus is more on the operational aspect of violence.

Let's now redirect our focus to the integration of spatiality (whose indications will be marked in *green*) within the context of this linguistic phenomenon. To illustrate, let's begin with an initial example:

«The *execution* of the IRA terrorist Eamonn Dillon, the *bombing* of the *National Library* in *Dublin*, and the *bombing* of the *Underground* at *Heathrow Airport* were carried out under the orders of the military council of the Ulster Freedom Brigade».

<div align="right">(<i>The marching season</i>, Daniel Silva, p. 26)</div>

This passage captures a pivotal moment in the novel, wherein a representative of the Ulster Freedom Brigade claims responsibility for a series of terrorist attacks spanning three different locations: Belfast, Dublin, and London. The inclusion of the word 'execution' not only frames murder as a punitive act but also places it within a systemic political framework. Also, this entire excerpt is contextually woven into a larger discussion surrounding the recurrent acts of violence perpetrated by the UFB. Their objective is clear: to compel the British government to reconsider its stance on the Good Friday Agreement.

Spatially, the narrative details three spatial frames, encompassing both specific facilities such as the National Library and the Underground at Heathrow Airport, as well

as geopolitical entities like Dublin. The threefold portrayal of spatial frames adds a layer of depth to the narrative, highlighting the strategic nature of the assaults. However, despite this attention to detail, the passage refrains from delving into the specifics of the individual locations and the violence they involve. They are mentioned to anchor the reader in the story world, but the particular one is not described in depth. Another example:

> He had been one of the first on the scene when UDA *bombs* killed twenty-five people in *Dublin* in 1974. For many years after that he had lived, like so many of his countrymen, with the dread that the conflict *in the North* would escalate and drag *the South* into a *cycle of terror and retaliation*, and, finally, *civil war*. His grandfather had been a fighter in the twenties and had seen at first hand how *intercine strife tore families and towns apart*.
>
> (Daniel Eastermann, *Day of Wrath*, p. 70)

This excerpt from Eastermann *Day of Wrath*, revisits the early years of Declan, the protagonist, who, according to the narrator, witnessed terrorist attacks and bombings from a very young age. This exposure led him to live in constant fear of the conflict escalating to a larger scale. Notably, the connection with space is portrayed through focalization: the broader the spatial focus, the more systematic the recounting of pain. Consequently, the narrative often centers on long-range weapons due to their capacity to inflict greater suffering. Bombings, grenades, and various weaponry described in the narrative cause immense pain, resulting in numerous casualties. The depiction of such violence is situated on a city level ('Dublin') or even on a regional scale ('The North', 'the South') giving rise to themes of 'terror,' 'retaliation,' 'civil war,' and 'interline strife.'

One final example remains to bolster the interpretation we are sketching, which we will expound upon in the concluding part of this paragraph:

He turned on the radio and they listened to the news about *Ulster*. A soldier had been shot dead by a *sniper*, another was seriously ill. The incident had happened in a *Catholic estate* in *Belfast*.

«It's time they locked your lot up,» said Sadie.

«They're not my lot,' said Kevin angrily»

(*Into Exile*, Lingard Joan, p. 111)

The narrative of Lingard's novel revolves around two adolescents hailing from Belfast – Sadie, a Protestant, and Kevin, a Catholic – trying to sustain their relationship amidst religious differences. After marriage, they relocate to London in an attempt to escape The Troubles. However, they encounter challenges as they grapple with the status of outsiders and the hardships of living with limited finances. The storyline unfolds with accounts of job losses, struggles in forging friendships, and conflicts arising as Sadie feels neglected while Kevin immerses himself in his new job. Their bond faces additional strain when tracing events befalls Belfast, forcing Kevin to navigate the delicate balance of supporting his family back home while striving to construct a new life with Sadie in London.

The selected passage discusses a news report revealing a murder depicted as an incident involving a long-range sniper rifle in Ulster, specifically in a «Catholic estate in Belfast». This revelation triggers a confrontation between Sadie and Kevin, with Sadie accusing the latter of taking implicitly part of an organization responsible for a specific type of terrorist attack. The emphasis here is not solely on the isolated act of violence, but rather on the recurring nature of such incidents that pose challenges for the entire region of Ulster – meaning the systematicity of violence.

Spatial references become more generalized, such as 'Ulster' and 'Belfast,' when recalling systemic pain. Consistently, in alignment with other findings in this chapter, these references are more commonly associated with long-range weapons. The more precise the account of violence, the narrower the spatial indication becomes, ranging from geopolitical locations like 'Ulster' and 'Belfast' to more specific sites such as a 'Catholic estate.'

In essence, beyond inheriting linguistic patterns observed in real-world testimonies, as seen in this case, the link between long-range weapons and systemic violence intersects with spatiality as a structural function. Rather than primarily associating spatial indications with a specific type of weapon, they vary in the extent of the area they represent based on the specificity of the violence being recounted. Systematic acts of violence, often discussed in a context involving long-range weapons, require placement in the story world by emphasizing their diverse positions within a broad area, hence being indicated with geopolitical entities. On the other hand, specific acts of violence, more frequently linked to short-range weapons, are easily associated with facilities, depicting a singular area of violence.

2.7. The Intersection between Pain and Spatiality: A Lotmanian Perspective

To summarize the findings of this second chapter, we can argue that the interplay of pain and spatial elements in fiction is significant in eliciting varying degrees and types of empathy and identification. Firstly, following Genette's (1980) argument that spatiality is always present as a fundamental backdrop for every narration, even when trauma hinders characters' ability to provide specific spatial details – a characteristic we linked to durative violent verbs – authors use general indications or geopolitical references to construct an organized narrative structure.

Secondly, in analyzing the patterns found in testimonies, we noted that fiction utilizes linguistic elements like durative verbs and detailed violence descriptions with precise spatial indications to evoke empathy from readers, thus suggesting a connection between empathy, spatiality, and the vivid portrayal of historical pain. This connection aims to immerse readers in the trauma of The Troubles and facilitate a cathartic experience, allowing them to transcend the pain.

Thirdly, the link between long-range weapons and systemic violence is structurally tied to spatiality. Spatial indications vary based on the specificity of violence being recounted, with systematic violence often associated with long-range weapons requiring

broader geopolitical references. In contrast, specific acts of violence, often involving short-range weapons, are depicted with a focus on individual facilities or singular areas.

These observations could be further substantiated by adopting a Lotmanian perspective (1975), wherein narrative spatiality is more profoundly comprehended as a dichotomic interplay between poles. In this context, the interplay between broader and narrower spatial indications, intersecting with poles such as short-range and long-range weapons, durative and punctual violent verbs, and type of violence/pain depicted, can be delineated in the following schema (fig. 2.3):
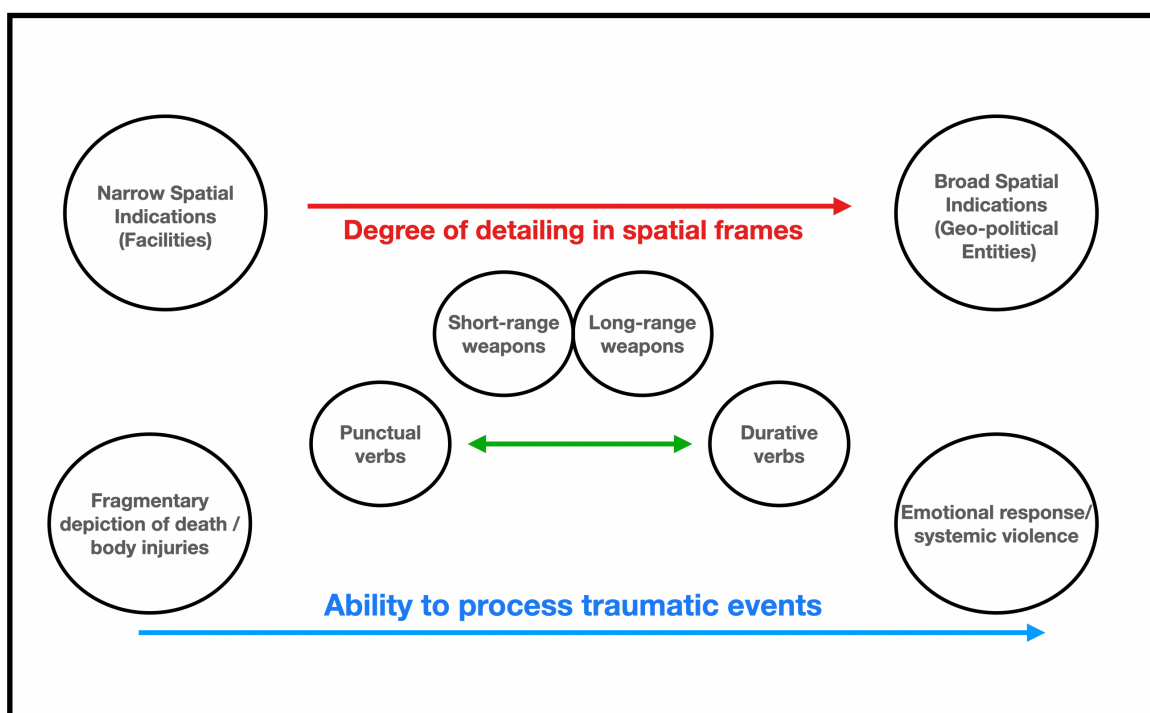
Figure 2.3. A Lotmanian schematic representation of the various polarities involved in the interplay between pain and spatiality in the novels depicting The Troubles.

Short-range weapons are often closely associated with punctual verbs. And punctual verbs are preferred by authors when they aim for a fragmentary yet precise description of pain and violence, utilizing narrow spatial indications such as specific facilities, street names, and neighborhoods. We interpret this fragmentary style as a reflection of one's ability (or inability) to cope with and process traumatic events. The lower the ability to cope with traumatic violence, the more fragmented the depiction of death becomes, characterized by a high presence of punctual verbs.

On the flip side of the spectrum, contrasting polarities emerge: long-range weapons are often associated with durative verbs, more commonly employed to depict systemic acts of violence (such as simultaneous terrorist attacks carried out across different locations). These actions, owing to their systematic nature, are frequently situated at a higher hierarchical spatial level and involve broad spatial references, such as counties, cities, and states. Systematic traumatic events are often connected to a conceptualized familiarity with distressing occurrences. In fact, to grasp the systematic nature of painful events, characters must have experienced or become acquainted with more than one terrorist event, they must have comprehended that they have occurred, and eventually place them within their psychological chronotope[37] within the context of a comprehensive narrative world.

This schematic representation is not intended to be prescriptive but rather descriptive. It recounts tendencies among polarities, as they are observed in novels depicting The Troubles. Its purpose is to help locate a novel in terms of the intersection between spatiality and pain. And, as mentioned at the beginning of this chapter, the analysis is aimed both to provide critical insights and to be valuable in terms of procedural thinking. Without the need to operationalize pain and spatiality in such a mathematical manner, we might not have fully understood the linguistic closeness of certain lemmas and concepts. The linguistic proximities identified with artificial intelligence algorithms eventually prompted a further analysis that culminated in the description of polarities.

In this context, computational literary criticism proves its worth not only due to its scalability but also because it compels the (computational) literary scholar to uphold an unprecedented level of rigor. In the specific case under consideration, it guided our reflection towards the exploration of an intriguing relationship between the fragmentary depiction of death and a particular type of spatial indications. A reflection facilitated by artificially intelligent algorithms, yet harnessing a distinct form of intelligence to shed

---

[37] For the concept of psychological chronotope, please refer to the aforementioned definition by Torop (2002), mentioned in the footpage number 29, located in paragraph 2.6.1.

light on a human-oriented interpretation. In fact, the ultimate goal was to encourage ethical reflection on The Troubles.

Fiction transcends pain by intricately weaving it into a tapestry of interplaying and dynamic elements, where spatiality assumes a relevant role in relation to the other features that compose literary works. Through a careful examination of these interplays, we can discern the elements that fiction inherits and sublimates from real-life testimonies. This process involves layering these narratives with a veil of aesthetic beauty, enabling readers to identify with the stories and facilitating a cathartic experience that aids in coping with traumatic violence. In essence, through the lens of beauty, we strive to transcend terror:

Let everything happen to you
Beauty and terror
Just keep going
No feeling is final

(Rainer Maria Rilke, *Book of Hours*, I 59)

Chapter III

3.1. Introduction: Eoin McNamee's *Resurrection Man* as a Focal Point

After outlining the general framework for CLC in the first chapter and considering a corpus of 41 novels in the second chapter, we will now illustrate our overarching theory through a case study focused on a singular novel: *Resurrection Man* by Eoinn McNamee. This particular novel was published in 1994—four years prior to the Good Friday Agreement—and will serve as a noteworthy example for elucidating our conceptual framework.

While in the second chapter of the dissertation we analyzed a corpus of novels with the goal of assessing the differences between real-world testimonies and novels that depict the spaces of The Troubles, in this chapter, with the same objective of delving into a politically charged case study, we will commence by examining journalistic documentation, with the aim - yet again - to verify communalities and differences between literary narrations and non literary narrations depicting pain and violence.

Our focus will be on a representative novel, aiming to compare its representation of terror spaces with the reports found in newspaper depicting the same spaces portrayed in the novel. We will first analyse non-fictional texts to then delve into the literary discourse to ascertain whether any peculiarities emerge when subjected to analysis using algorithms and artificial intelligence. *Resurrection Man* is well-suited for this purpose, as it is loosely based on a recognizable historical event: the Shankill Butchers incidents.

3.2. The Shankill Butchers: an historical overview

The more reliable historical review of the Shankill Butchers is in Martin Dillon's *The Shankill Butchers: A Case Study of Mass Murder* (Dillon, 1999), where Dillon had extensive access to the case files of the Royal Ulster Constabulary (RUC). The Shankill Butchers were a notorious group of loyalist paramilitaries operating in Belfast, Northern

Ireland, during the 1970s. Led by Lenny Murphy, the group was responsible for a series of gruesome and sectarian murders. The Shankill Butchers primarily targeted Catholic civilians, and their brutal methods included abductions, torture, and killings, leaving a dark mark on the troubled history of the region during the period of sectarian violence.

Lenny Murphy was the youngest son of Joyce and William Murphy from the loyalist Shankill Road area in Belfast. Murphy, associated with the Ulster Volunteer Force (UVF), was known for attending paramilitary crime trials to familiarize himself with legal procedures.

In 1972, Murphy was arrested for the murder of William Edward Pavis at his home in West Belfast, but was acquitted due to disruptions during the trial. After various escape attempts, Murphy was re-arrested and imprisoned, and during the imprisonment Murphy got married and had his first daughter, with a woman named Margaret Gillespie. Released in 1975, Murphy established a paramilitary team, which included his brother, John Murphy. The gang, famously known as the 'Shankill Butchers', carried out gruesome murders that have become infamous in history due to their brutality. The perpetrators frequently utilized stolen knives and meat cleavers in these horrific acts.

The gang targeted Catholics, and their victims endured brutal and sadistic killings, with Murphy's involvement in the murders continuing indirectly even during his further imprisonment. In 1979, eleven members of the gang were convicted, with Murphy implicated as the driving force. Even though the members of the band received the longest combined prison sentences in United Kingdom legal history (Dillon 1999, p. 312), Murphy and two other prominent members, referred to in the news as 'Mr. A' and 'Mr. B,' faced no charges due to a lack of evidence.

After Murphy's release in 1982, he resumed his killing spree, targeting both Catholics and Protestants, with the formers singled out more for personal reasons than for sectarian disputes. After the period of imprisonment, he assembled a new gang and continued his violent activities. Ultimately, Murphy was assassinated by the IRA in November 1982 while in front of his girlfriend's house in Gleincairn, with his family denying his involvement in the Butchers' crimes.

From July 21, 1972, when he took part in the murder of Francis Adams, to November 16, 1982, when he was fatally shot, Murphy was actively engaged in terrorist actions or serving time in prison for them, with him and his unit being responsible for the deaths of possibly up to two dozen people. The majority of the victims were arbitrarily chosen men presumed to be Catholic based on the places they frequented or the neighborhoods they lived in. Some were Ulster Volunteer Force (UVF) members killed in retaliation for disciplinary infractions, such as selling weapons to Catholics or robbing elderly Protestant women. The crime scenes were recognisable with most of the victims having multiple stab wounds, and some being nearly decapitated, with butcher knives. Despite many working-class Protestants being aware of the identities of the 'Shankill Butchers,' who were involved in these brutal acts publicly, the activities persisted for a decade. This endurance was due to the thorough intimidation of the community and Murphy's cunning, which matched his brutality.

From this context, *Resurrection Man* emerged. And, as we will observe in the analysis, Belfast is not explicitly mentioned frequently; however, these events had such a profound impact on the general public that it was quite straightforward, especially around the time of publication, to recognize the connection between the novel and the events unfolding in Belfast in recent years.

3.3. The spaces of *Resurrection Man* in the academic discourse

In terms of critical reception, Eoin McNamee's *Resurrection Man* has garnered 90 and counting academic publications, with scholars debating its various aspects and significance. Diverse voices have either critiqued or praised its portrait of Belfast.

Generally speaking, the novel represents a fictionalized re-telling of a real and particularly gruesome series of sectarian knife-killings perpetrated on Catholics by a Loyalist gang (the 'Shankill Butchers') in 1970s West Belfast (Stainer 2006). Haslam (2000) asserts that the narrative of *Resurrection Man* appears complicit with the sadism of the killers, thereby perpetuating additional violence against the actual victims of the Shankill Butchers. On the other hand, as we delve into forthcoming spatial

characteristics, the relative blurriness of fictional spaces makes the novel be viewed as a postmodern and ambivalent work, devoid of any clear ethical stance. In fact, according to Johnson (1999), the novel challenges and disrupts social-scientific explanations for violence, both in terms of psychology and sociopolitical geography, with the city of Belfast defined by elements such as work-ethic, purpose, progress, modernism and industrial materialism.

Delving into the characterization of Belfast, Scanlan (2001, p. 42) asserts that, «although the text names innumerable Belfast streets, parks, prisons, hotels, and bars – the Shankill, Crumlin Road, Ormeau Park, the Europa – never is 'Belfast' once named.» However, this assertion is not supported by computational analysis, as we found five instances of the space-name 'Belfast'. Nevertheless, we consider that the relatively infrequent mentions of the word 'Belfast' could be interpreted in line with Scanlan's criterion: McNamee's political explanation for avoiding a clear mention of the city setting may reflect a desire to convey that, at least for working-class citizens residing in the poorest neighborhoods of the city, 'the Shankill' or 'Crumlin Road' hold genuine political significance, while 'Belfast' suggests an illusory unity.

Certainly, McNamee's focus on the actual names of streets and buildings serves to underscore fragmentation, nostalgia, and the remnants of a colonial era: «The city […] has withdrawn into its places-names. Palestine Street. Baklava Street. The names of captured ports, lost battles, forgotten outposts held against inner darkness» (McNamee 1997, p. 3). The names of places bear remnants of an era that the present-day residents of the city have overlooked— not a vision of the past shaped by Irish nationalism or even one by an Ulster native, but rather a past rooted in imperial Victorian times.

In such a context, there are also proponents of a historical and economic interpretation of place names, wherein street names acquire symbolic significance. Dermot McCarthy, for example, underscores that these names «evoke Belfast's involvement in both the ascent and decline of the British Empire» (McCarthy 2000, p. 138). Consequently, as Elmer Kennedy-Andrews argues (2003, p. 126), such framing of Belfast's urban spaces results in street names becoming unreliable, as they confine the citizens of Belfast «within a linguistic contour that no longer corresponds to the

landscape of fact» (Kennedy-Andrews, 2003, p. 126), and can no longer be trusted to delineate the contemporary narrative of Belfast: the remapped city no longer aligns with the deeply ingrained topography of Victor's racial memory.

While these fictional urban locales, deeply rooted in Belfast, may be distinctly identified from McNamee's novel, the spaces of *Resurrection Man* give the impression of a territory almost detached from historical context and characterized by a constantly changing cartography. Cartography itself appears to diminish, eroded by the city's inherent resistance to permanence, as argued by McGlynn (2016), with McNamee placing emphasis on street names rather than explicitly mentioning the city. This approach presents the city as something fragmented, thereby challenging the conventional dichotomy between 'Us' and 'Them,' as asserted by Herman (2015, p. 692-3), since every part of the city seems to be intricately molded into the other.

According to Herman (2015), this spatial ambiguity positions *Resurrection Man* within a «cultural tradition of sorts» encouraging the reader (or audience) to adopt a more expansive understanding of terrorism, not necessarily related to the Irish Question. Herman suggests, «One might include such disparate works as [...] Eoin McNamee, *Resurrection Man*. Even though these works are distinct in terms of time, location, subject, genre, and ideological stance, they intricately challenge both the notion of terrorism and the customary responses to it by breaking down any overly simplistic categorizations». McGlynn (2016, p. 58) even contends that, in such a context, the Catholic victims in *Resurrection Man* could seamlessly be replaced with Jews or blacks, the streets could represent those of any European city, and Victor Kelly could embody an English, French, or Italian figure: while nationality emphasises uniqueness, social class – which ends to be, by locating the action in the streets of the Shankill neighbourhood – identifies the universal and transcends temporal constraints.

Moreover, the spatial dimension is intricately linked to the portrayal of the characters' inner selves, as highlighted by Bennet (2014, pp. 230-1), who points out that, despite the limited attention paid by writers to the dark world of loyalist paramilitaries, and McNamee himself delves into loyalist terror in *Resurrection Man* without placing excessive emphasis on the ideologies of the murder gangs or the political goals they

pursue. Instead, McNamee's focus lies on the inner lives of the characters, presenting a portrait of the internal landscape of Ulster loyalism that is relentlessly bleak, shaped by psychopathologies characterized by neurosis, inadequacy, and sexual fear.

Aligned with the inner conflicts of the characters, the setting of Belfast is depicted as notably ominous. The second protagonist, Ryan, a journalist reporting on the brutal murders committed by the Loyalist paramilitaries, articulates the description of a emblematic scene within this Belfast:

> The city was always empty after five o'clock. Street lighting was sparse as if areas of darkness had been agreed. You got the feeling of single cars cruising the streets with sinister gleams from their windscreens. Drizzle falling from a vigilante sky

(McNamee 1997, p. 25)

Differing perspectives on this matter are presented by Kelly (2004, p. 168), who observes that while critics often condemn this novel for perpetuating the notion of Belfast as a stagnant, infernal backdrop abandoned by global history, he contends that Resurrection Man is «a text that disrupts and overcomes the simplistic categorization that has plagued the mainstream political and cultural portrayal of the conflict» (Kelly 2004, p. 168). He rather suggests that the novel indeed paints Belfast as a nightmarish setting without indulging in stereotyping but showing the harsh reality experienced by many residents during that period of terror – intensified significantly by the actions of the Shankill butchers.

In this context, Belfast takes center stage in the discourse on Irish fictional places, emerging as the protagonist in the elegiac depictions of the North, as observed by Fadem and Maureen (2015), while with a wild nihilism, McNamee's Belfast conveys to us that «partition was not merely a rupture; it was a burial» (Fadem & Maureen 2015, p. 6). In other words, McNamee delivers a politically charged and boldly realistic novel, offering candid critiques of The Troubles, partition, and the broader political history without apology.

3.4. Two corpora: fictional and non-fictional texts

As outlined in section 3.1, the primary objective of this chapter, akin to the preceding one, is to juxtapose the portrayal of pain and violence in fictional accounts of The Troubles with other forms of narration. The goal is to underscore the diverse qualities of fiction in conveying the experience of pain.

To accomplish this, two distinct corpora are required (which consist in the second step of the theoretical framework outlined in chapter one):

1) The initial corpus comprises a representative novel, and we have chosen *Resurrection Man* by Eoin McNamee, resulting in a corpus consisting of 73,064 words;

2) The second corpus includes news reports detailing the same events that served as the foundation for *Resurrection Man* (i.e., the Shaking Butchers case). This corpus comprises 134,759 words and includes:

Author/Source	Title	Date	Type of text
Dillon, M.	The Shankill Butchers: the real story of cold-blooded mass murder	1999	Journalistic essay
Dillon, M.	A Legacy of Hatred	2009	News article
Belfast Telegraph	Slaughter in Back Alley	1975	News article
Tinney, A.	Lenny Murphy's Brother was Real Boss of Shankill Butchers Gang, Says New Book	2017	News article
Jordan, H.	Lenny Murphy – Local Hero, Shankill Road Style	2011	Book Chapter
Browne, M.	On the First Anniversary of Martin O'Hagan's Murder	2003	News article
Walker, G.	Shankill Butcher is Free	1996	News article
The Independent	Exposure Sealed Fate of Notorious Activists	2000	News article
Carroll, J.	Tales from the Dark Side	2007	News article

Sengupta, K.	Spate of Savage Killings Blamed on Loyalist Feud	2002	News article
McAleese, D.	Shankill Butcher William Moore Found Dead in Flat Was Due to Face Questioning	2009	News article
BBC News	Murder Link to Shankill Butchers	2004	News article
Preston A.	Shankill Butcher Eddie McIlwaine Spotted Putting up UVF Flag	2023	News article
Reid, K.	Shankill Butcher Eddie McIlwaine Spotted Marching in Whiterock Parade	2023	News article
Gorman, R.	Sick New UVF Memorial Honouring Notorious Shankill Butchers Gang Causes Outrage	2023	News article
BBC News	Winston Rea 'Criticised Shankill Butchers' in Boston Tapes	2020	News article
BBC News	Shankill Butchers Held Belfast in Grip of Terror	2011	News article
BBC News	Loyalist Bar Killing 'Horrific'	2006	News article

3.5. Methodological approach

As for the research methodology, after reviewing the academic literature on the *Resurrection man*, we determined that two categories could be pertinent to the selection process of abstracted concepts (as declared the initial phase of our Computational Literary Criticism theory, as argued in Chapter one):

1) Space Mentions: According to our literature review, the absence of specific place mentions may be more noteworthy than their presence. As fictional works leverage this interplay and integrate it into a meaningful system of signification. To validate this hypothesis, our approach involves identifying place mentions in both the journalistic and fictional corpora. This will enable us to compare how news and fiction characterize the most significant places in relation to the events

involving the Shankill Butchers.

2) Itineraries of Characters: This category, closely related to the previous one, defines itineraries according to the journeys that characters undertake between distinct fictional locations. Narratological debates[38] have underscored the relevance of itineraries in structuring character systems. By discerning the types of itineraries present in both news reporting and fiction, then scrutinised in a comparative analysis, we aim to pinpoint where fiction diverse from journalistic texts, and provide commentary on them.

To address the first category and identify place mentions, we employed two distinct approaches:

A) Machine Learning Approach: We utilized Named Entity Recognition (NER) through the Spacy library, as elaborated in Section 2.6.2. Spacy employs a pre-trained Language Model that has undergone training to recognize locations through exposure to substantial examples of human-annotated data. This application of artificial intelligence allows for the substitution of humans in tasks that are repetitive and mechanical.

B) Gazetteer Approach: To enhance the efficiency of place-mention retrieval, we employed a gazetteer approach. This involved triangulating the analysis by utilizing a set of lists containing geographic information about places in Northern Ireland. A script iterated through every word in the corpus, and if a word was found to correspond to one of the places mentioned in the gazetteer, it was annotated. For

---

[38] Please refer, for instance, to Herman (2002), p. 275. Itineraries, as journeys that characters undertake between landmarks, could hint at a particular semiological value of the character system.

more effective referencing, the gazetteer lists comprised a database of place names[39] categorized as follows:

- Six list of small cities and towns, divided for each of the six Northern Ireland Counties
- A list of Belfast Street Names
- A list of Cities included in Northern Ireland
- A list of Northern Ireland Counties
- A listi of Northern Ireland Regions
- Six lists for the Northern Ireland townlands

As for the second category that we aimed to extract – namely the itineraries – we began by reviewing relevant literature on motion verbs in English (Collitz 1931; Slobin 1996; Stamenkovic 2013; Hickmann et al. 2017) and compiled a list of motion verbs. Subsequently, employing automatic Part of Speech Tagging[40] with the Spacy Library, we initially identified those motion verbs. We then examined whether any spatial indications were present among the child tokens of the motion verbs using Named Entity Recognition (NER). We filtered the results based on prepositions indicating motion to a place ('to,' 'towards,' 'at,' 'in,' 'in the direction of,' 'for,' 'onto') or indicating motion from a place ('from,' 'out,' 'out of,' 'off,' 'away from'). This filtering was based on Zhang (2002), who classified motion prepositions for the English language. In simpler terms, we successfully determined when a motion verb expressed movement and

---

[39] The names of locations in each category were extracted from Wikipedia and geographic.org, both of which provide a comprehensive list of geographical locations along with corresponding Google Maps references. In case of interest, please refer to: https://geographic.org/streetview/northern_ireland/belfast/belfast.html [Last checked on December 21, 2023]

[40] Part-of-speech tagging (POS tagging) is a natural language processing (NLP) task that involves assigning a specific part of speech (e.g., noun, verb, adjective) to each word in a given text. The goal of POS tagging is to analyze and understand the grammatical structure of a sentence, which is crucial for various NLP applications such as text parsing, sentiment analysis, and machine translation.

subsequently classified the movement as either a motion from or to a specific place. This pattern allowed us to identify itineraries for both individuals featured in the news and characters performing actions in fiction.

3.6.1. Data analysis #1: Spatial indications in the Journalistic Corpus

As stated before, each of the two corpora – the one comprising *Resurrection Man* and the one comprising news about the Shankill Butchers – underwent processing to extract first-place names and secondly itineraries. Let us start with the results on the journalistic corpus.

The initial findings of this phase of the analysis reveal that within the journalistic corpus, 49 different cities in Northern Ireland are mentioned. This accounts for a total of 652 references out of 4380 spatial indications within the corpus, indicating that only 14.88% of the spatial indications pertain to city-level locations. This result is not surprising, given that the Shankill Butchers primarily operated at an urban level in Belfast, which is expected to be almost the only mentioned city. Such a relative scarcity of city mentions can also be explained by the composition of the journalistic corpus, which includes news articles from local Belfast newspapers and the extended account from Dillon. The latter mentions the setting of the events a few times at the beginning of the book and then assumes it for the remainder of the work.

A second relevant observation based on the gathered data is that out of the 4380 place mentions, 687 are streets from Belfast. This represents 15.68% of the total, with 103 different street names being referenced. This mirrors the fact that, since the Shankill Butchers carried out their attacks in Belfast, particularly around the infamous Shankill Road, reporters needed to precisely locate the events within various parts of the city, somehow circumscribing the events and consequently creating a very focalized image of that Belfast – that is to say designing 'another' Belfast, of a new map of the city.

The top twenty most frequently mentioned roads are as follows, along with their respective counts:

Shankill Road	104	Donegall Street	11
Crumlin Road	44	Brown Street	10
Tennent Street	37	Cavehill Road	10
Cliftonville Road	34	Manor Street	10
Antrim Road	28	Cambrai Street	9
Millfield	27	Royal Avenue	9
Falls Road	25	Highfern Gardens	8
North Street	14	Gresham Street	8
Mount Street	14	New Lodge Road	8
Rumford Street	12	Union Street	7

If we visualize these roads on digital maps (please refer to *Figure 3.1* and *Figure 3.2*), we can pinpoint the majority of events associated with the Shankill Butchers is in West Belfast, close to the city center:
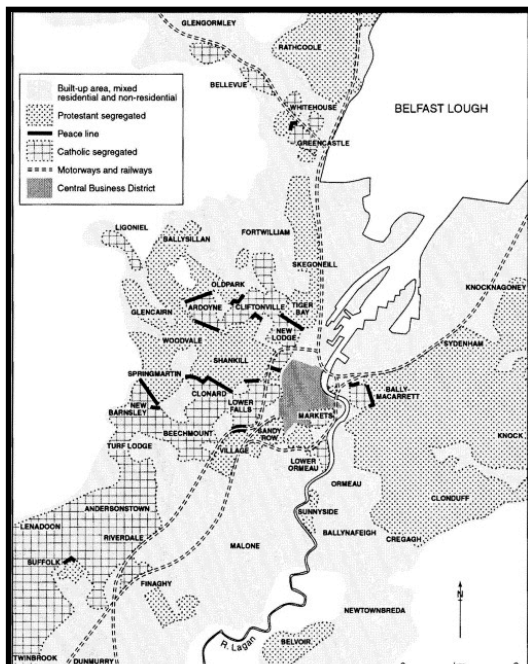
Figure 3.1. Peace lines and ethnonational segregation in Belfast city around 2000 (image from Boal, 2002).
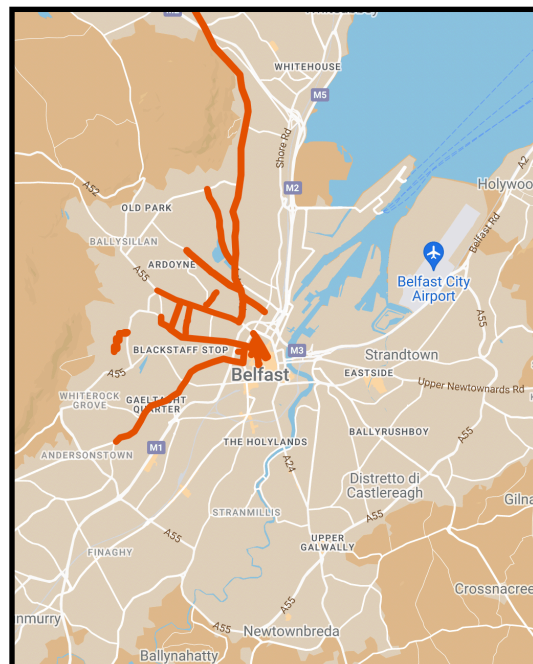
Figure 3.2. The twenty Belfast roads (marked in orange) most closely associated with the Shankill Butchers in the journalistic corpus (our image).

Violence is particularly pronounced along the peace walls and in the suburb of Shankill, which shares its name with both the suburb and Shankill Road. Notably, many victims of the gang met their tragic end in this area. According to the Cain Archive,[41] the sixth victim of the gang, 19-year-old Noel Shaw, was fatally shot in a taxi on Nixon Street, just off the Shankill Road; Edward McQuaid, aged 25, became a victim of a drive-by shooting while walking along the Cliftonville Road; and Cornelius Neeson, aged 49, was discovered beaten to death at the junction of Manor Street and the Cliftonville Road, among other victims.

Among the spatial indications that surpass the categorization of the gazetteers (such as 'West Belfast,' 'Northern Ireland,' bars and pubs, gardens, and cemeteries, among other general spatial indications), which, on the other hand, were identified through NER, an interesting pattern emerges. Let us examine the counts of places that are not contained in the gazetteer list but have been categorized as spatial indications by the NER process carried out with Artificial Intelligence:

Windsor Bar	46	Lowood Park	9
Brown Bear Bar	35	Crumlin Road Courthouse	8
Crumlin Road Prison	34	Lawnbrook Club	8
Maze Prison / Long Kesh	15	Belfast Telegraph	7
Long Bar	13	Royal Victoria Hospital	7
Chlorane Bar	10	Tennent Street Station	7
Unity Flats	10	Mater Hospital	5

Bars and pubs such as 'Windsor Bar,' 'Brown Bear Bar,' 'Long Bar,' and 'Chlorane Bar' are referenced quite infrequently in comparison to other spatial indicators,

---

[41] The CAIN project, a 20-year investigation on The Troubles from Ulster University, also encompasses a detailed account of conflict casualties per year, specifying the location of each death. If interested, please refer to: https://cain.ulster.ac.uk/sutton/chron/1976.html [Last checked on 21 December 2023].

constituting 214 out of 4380 mentions. This accounts for approximately 4.88% of the total references.

Even less frequently mentioned than recreational places are those related to legal proceedings, such as Crumlin Road Prison and Maze Prison, or Crumlin Road Courthouse. The combined occurrences of these legal sites amount to 57, which equals 1.3%. Interestingly, both Crumlin Road Prison and Maze Prison held significance for Lenny Murphy, the leader of the Shankill Butchers. During his initial imprisonment in the former in 1972 (Dillon 1999, p. 46), he met his future wife Margaret Gillespie, with whom he had a child while still incarcerated. Meanwhile, in the latter, known as 'The Kesh,' Murphy continued to give orders to a formalized team of terrorists (Dillon 1999, p. 168).

This tally of mentions, whether regarding bars and pubs or buildings associated with legal actions, may seem insignificant. However, it will prove to be a remarkable difference compared to McNamee's novel, as we will explore in the following paragraphs. While strategically important for Murphy's life events or strategic activitie, according to our analysis of the news corpus, these places were not considered particularly described. The news corpus seems less concerned with uncovering the strategic activities surrounding the Shankill Butchers' murders and instead focuses on reporting and locating the violent events they perpetrated. The same observation applies to mentions of pubs and bars. While the upper floor of the Brown Bear pub served as a meeting place for the gang, located at the intersection of Mountjoy Street and Shankill Road near Murphy's home (Dillon 1999, p. 55), bars are not frequently emphasized as spatial indicators in this corpus. Many of these bars have since closed and no longer exist, but their locations can be traced thanks to a few sources. For example, the Brown Bear Pub, though now closed, was situated on the corner of Upper Shankill and Mountjoy Street (Taylor 1999, p. 153), depicting a strategic geographical point.

Let's reconsider, keeping this in mind, a few notable references to these spatial indicators in context:

1) Murphy suspected three UVF men who frequented the nearby **Windsor Bar**. He had a word with their boss and arranged for them to be 'arrested' and taken to a room above the **Brown Bear pub** which Murphy used as his own HQ (Jordan 2011, p. 202)

2) The UVF leadership then quizzed Murphy about Robinson's death, but he denied any knowledge of it. However, within days the truth emerged and the leader of the **Windsor Bar** team ordered Waller's death in revenge. On 29 November 1975, Waller – a married man with two children – was shot dead as he sat in his car outside a loyalist club in Downing Street. Murphy was furious at this challenge to his authority and ordered his men to abduct Waller's killers. (Jordan 2011, p. 203)

3) During the first month of his freedom he spent much of his time in the **Brown Bear pub** on the Shankill Road, which he established as a headquarters and meeting place. In the same area he found himself faced with a competing unit which operated from the **Windsor Bar** and was run by a man of Murphy's generation: Anthony 'Chuck' Berry. (Dillon 1999, p. 55)

Based on the aforementioned excerpts, despite the sparse mentions in the overall news corpus, the Windsor Bar functions as the headquarters for another paramilitary group that rivals the Shankill Butchers, while the Brown Bear Pub serves as the gathering point for the gang itself. In terms of strategic significance, these locations should hold greater importance, but in the news corpus, they go almost unnoticed.

Similarly, the mention of two Belfast prisons indicates their strategic relevance when analyzed in context. When read in context, out of the 49 references to the two prisons, all mentions of 'The Maze' are linked to the strategic activities that Murphy was carrying out, regardless of his confinement. It even appears that being enclosed facilitated Murphy's activities, leaving him unsuspected of being responsible. In contrast, Crumlin Road Prison is consistently mentioned and described as a place where the police had stronger control over prisoners. Notable examples of this differing dynamic can be found in the following excerpts:

1) Out of reach of the police for six years, he could continue, with Mr A.'s assistance, to control the Brown Bear unit and give orders for further killings from his prison cell. Once Murphy was locked away in the **Maze Prison** (formerly Long Kesh) he felt able to put into action a plan which would not only satisfy his craving for violence but would fool Nesbitt into believing that he was not, in fact, the cut-throat murderer. (Dillon 1999, p. 166)

2) His meetings were designed to persuade Moore and Bates to put in writing their knowledge of the roles played by Murphy, Mr A. and Mr B. Each visit necessitated sealing off part of the prison building and placing prison staff known for their trustworthiness in charge of the interview area. A prison officer who was serving in **Crumlin Road Prison** at that time told me that only two or three officers were chosen for the task and no one else was allowed near the interview room. (Dillon 1999, p. 238)

3) At this time **Crumlin Road Prison** was a difficult place in which to provide the kind of protection sometimes necessary to keep a prisoner alive. (Dillon 1999, p. 40)

Prisons undeniably played a pivotal role in The Troubles and the acts of terrorism in Belfast, such as the Shanking Butchers events. They served as both witnesses to the incarceration of those responsible and catalysts for the emergence of new political parties. The concentration of politically motivated prisoners from the same paramilitary organizations fueled the politicization process, with Sinn Féin itself being the most renowned example established in the 'Long Kesh'/'The Maze', as noted by McAtackney, L. (2014).

Despite playing a crucial strategic role, the two Belfast prisons are infrequently mentioned in the corpus in relation to the Shankill Butchers events. To be specific, as we mentioned, only 1.3% of the total spatial indicators directly reference the two prisons, particularly in the context of juridical proceedings like trials or simple imprisonment. This scarcity of mentions can be attributed to the communicative intent

of the texts, which primarily seek to provide information about the assassinations, focusing on the terror narrations and the criminal issues involved, rather than on the strategical aspects of the paramilitary groups.

3.6.2. Data analysis #2: Spatial Absences in Fiction

As for the fictional corpus, comprehending *Resurrection man*, after having processed the text with same algorithms used in the previous paragraph, the total spatial indications found are 223. These spatial indication then underwent a categorisation with the gazetteer lists presented in 3.3.

We have discovered that out of 223 spatial indications, 71, equivalent to 31.83% of the total indications, pertain to streets in the city of Belfast. Paradoxically, the mere mention of Belfast occurs only five times throughout the entire novel. What is noteworthy is the scarcity of references to this specific location, which was highly prominent in the journalistic corpus. We posit that McNamee intentionally made this choice, and to delve into and interpret this decision, we will analyze the five instances where the name 'Belfast' appears:

> 1) Coppinger came in and ordered a pint of Bass. He'd been drinking in Tiger Bay. Listening to stories about the Blitz, Kingdom Brunel in **Belfast**, the construction of the Titanic in the shipyard. He said that a cousin of his father's had accidently been sealed in the Titanic's double hull and the body had never been found. (McNamee 1997, p. 21)

> 2) Dawn had broken. There was a fine drizzle falling and Heather laughed at the way rain beaded in Victor's hair like a hairnet you'd see an old woman wearing. […]At first Heather thought he was driving at random; then she saw the pattern. He was driving carefully along the edges of Catholic west **Belfast**. She had never been this close before although she had seen these places on television. Ballymurphy, Andersonstown. The Falls. Names resonant with exclusion. (McNamee 1997, p. 43)

3) 'Right now? I don't see why not. Come on over, like.' She spoke the last sentence in an imitation of McClure's accent. **East Belfast** chant. Rising to a falsetto on the last word of the sentence with an intimation of fear. (McNamee 1997, p. 176)

4) 'Did he have a name?'
'It was a Protestant-sounding name. Some wild **Belfast** name. You know like one of those names, when you say it, you have to screw up your face and squint like you'd a fag in your mouth. McClure, it was. Billy McClure. Called me missus all the time. (McNamee 1997, p. 176)

5) He began to spend nights in the back room of the Pot Luck again, his head lost in chemical dazzle until the speed began to wear off with the advent of another glassy, **Belfast** dawn. At these times he started to think about his mother. He was reluctant to approach the house in case the police or even former members of Darkie's unit had it watched but it brought tears to his eyes to think of the hardness of her life. (McNamee 1997, p. 216)

The first example is drawn from the third chapter and unfolds in a pub where Coppinger is scheduled to meet with Ryan. Belfast is mentioned here, associating it with Kingdom Brunel, a prominent British civil and mechanical engineer considered one of the most influential figures of the industrial revolution. However, Belfast doesn't serve as a geographical reference for the characters in the fictional world, rather it links the city's name to an almost iconic British figure, thereby distancing Belfast from the narrative world in which the characters find themselves.

The second passage mentioning Belfast is particularly significant as it establishes an isotopy that will reemerge towards the end of the novel. In this case, Victor Kelly and Heather are preparing to drive along the borders of Catholic west Belfast. Victor's careful driving, marked by a pattern initially unnoticed by Heather but later recognized, aligns with the possibility for Victor to perform a certain action. The paragraphs commence with a temporal indication – the dawn – which will be repeated in the fifth

and passage mentioning Belfast. These two instances, from excerpt number two and excerpt number five, are the only occasions directly referring to the geographical location of the fictional world within Belfast, making this pairing distinctive.

The third and fourth instances of the mention of 'Belfast' both pertain to the characterization of McClure, first to the character's voice and then to his urban origins. Again, as in the first excerpt, 'Belfast' is not used directly to pinpoint the narrative world in the capital of Northern Ireland, it hints that one of the characters involved with the resurrection men (i.d. McClure) hails from East Belfast, indirectly situating the narrative world around the city but without explicitly mentioning where the action is unfolding.

The fifth and final example completes the circle of the isotopy introduced in this paragraph, echoing the aforementioned second mention of Belfast. This passage comes from the penultimate chapter of the novel. Victor Kelly knows he is about to meet his demise, and so does his partner Heather, who refuses to abandon him and is prepared to witness his death. Once again, in the second instance where the word 'Belfast' is directly used to locate the fictional action in the city, the characters represented in the fictional scene are Victor and Heather, and the time of scene is the dawn. However, in this instance, Victor is no longer driving through the streets of Belfast but instead stays «nights in the back room of the Pot Kuck,» under the influence of alcohol and chemicals.

The general absence of Belfast as a spatial indication within the novel, paired with the fact that only two direct mentions are used to describe similar scenes, creates a circular and isotopic composition, that serves to underscore two contrasting states of mind of the main character. While for the majority of the novel, Victor is audacious, daring, violent, and holds power over the city through his deadly activities, in this final part of the novel, he has come to realize his failure.

The interplay between the absence and occurrence of spatial indicators, most prominently observed in the journalistic corpus, allows us to illuminate one of literature's arguably most idiosyncratic functions: the ability to utilize silence as a structural feature.

This leads us to a significant observation regarding the use of artificially intelligence-oriented algorithms for processing and interpreting literature: algorithms struggle to process and categorize elements that are not explicitly present in the text. As algorithms operate on explicit verbal data, any information not articulated in words cannot be effectively processed. However, as we have witnessed, it is precisely from silences and absences that, at times, a structural interpretation of a novel can emerge.

This underscores the importance of recognizing that, regardless of the complexity of the algorithms we employ, we must not only guide them carefully but also must be ready to fill the gaps and evaluate their outputs using the diverse intelligence that humans possess. While machines excel at mechanical, repetitive tasks and can swiftly execute identical steps millions of times, always counting on the presence of items, a literary scholar possesses the ability to categorize not only what is explicitly stated but also what can be inferred from the absence of information.

In essence, this highlights that, alongside utilizing advanced algorithms, human judgment remains indispensable in discerning nuanced aspects of literature that go beyond the capabilities of even the most sophisticated artificial intelligence systems.

3.6.3. Data analysis #3: Itineraries in the Journalistic Corpus

As for the itineraries, when applied over the Journalistic Corpus the algorithm presented in 3.5. initially found 499 positives instances. After a manual check of all of them, the results were then refined to 295 true positives, with many cases considered not relevant under the category of 'itinerary'. The itineraries were then classifiable as following:

1) 134 out of 295 (amounting to 45.42%) itineraries are described with clear and precise spatial indications regarding Belfast. A few illustrations are as follows: «At Murphy's command, Moore drove to the Lawnbrook Social Club», (Dillon 1999, p. 108); «Moore drove off into Upper Library Street and onto Peter's Hill and the Shankill Road», (Dillon 1999, p. 117); «They drove up the Shankill and

into Ainsworth Avenue, where he stopped the car and a van passed them», (Dillon 1999, p. 216);

2) 73 out of 295 (amounting to 24.74%) itineraries are described with general indications, avoiding precise street names and with a lot of deictics and indexicals involved. Several instances include the following: «It was discussed that we got out and get a Taig», (Dillon 1999, p. 163); «He then made his way to the hijacked car and Moore drove the car away at speed», (Dillon 1999, p. 114); «This came about because of the violence of the early 1970s when Catholics were burned out of their homes» (Dillon 1999, p. 100);

3) 46 out of 295 (amounting to 15.59%) are short itineraries, describing brief movements within the same spatial setting. Various samples comprise the following: «It is a strange fact that Waller did not move after McAllister hastily left the vehicle» (Dillon 1999, p. 92); «Bates rushed from the taxi and helped Murphy drag Rice, who by now was slightly dazed, into the vehicle» (Dillon 1999, p. 117);

4) 23 out of 295 total entries (equivalent to 7.79%) are itineraries in which the landmark (i.e., the reference point for the starting or ending point of an itinerary) is mentioned as granted, as previously stated in another part of the text. This is especially applicable to indications such as 'Murphy's place' or similar, where the spatial reference should be familiar to the reader. Examples of this include: «Big Sam and me and Artie got back into the car and we went back to [Mr A.'s] house» (Dillon 1999, p. 195); «Patrick walked to his brother-in-law's house and was driven home» (Dillon 1999, p. 268);

5) 15 out of 295 itineraries (equivalent to 5.08%) include stops at bars and pubs. The following examples are included: «Mr I. took me back to the Windsor Bar and to the rear of it» (Dillon 1999, p. 143); «Hagan and his earlier passenger

were taken to the Windsor Bar with a warning that they would be held hostage there for half an hour» (Dillon 1999, p. 136); «Nesbitt rushed over to find that in fact the Windsor Bar itself was the murder scene» (Dillon 1999, p. 81)

Based on this precise data, it appears that journalistic reports emphasize entanglements with specific locations in Belfast, explicitly naming them. General indications are also widespread, following, in terms of occurrence, by the category of brief itineraries detailing specific individuals involved in minor movements within the same location. While this analysis may seem overly detailed, it serves as a foundational point for comparing how itineraries are depicted in fiction, as we will see in the next paragraph.

3.6.4. Data analysis #4: Itineraries of Fictional Characters

We then applied the algorithm used in 3.6.5 to retrieve itineraries on *Resurrection Man*. The software application initially identified 246 instances of itineraries. The results were then manually checked, resulting in 141 valid itineraries, while the remaining instances were identified as false positives (including phraseological verbs, idiomatic expressions, among others). Upon reviewing all 141 valid cases of itineraries, we categorized them as follows, along with their respective percentages:

1) 38 out of 141 (which equal to 26.95 %) are itineraries in which the landmark (i.e., the reference point for the starting or ending point of an itinerary) is mentioned as granted, as previously stated in another part of the text, especially when is the place of living of another character. The following examples are included: «Dorcas was preparing to shave James when Victor arrived» (McNamee 1997, p. 227); «They took their duty as Victor's mates seriously, and visiting Heather while Victor was an act of duty» (McNamee 1997, p. 93);

2) 37 out of 141, (which equal to 26.24 %) are itineraries that are described with general indications, avoiding precise street names. Examples of this could be the following: «Afterwards Dorcas would admit without shame that having moved house so often was a disturbance to Victor's childhood» (McNamee 1997, p. 1); «Darkie would go drinking in local discos to bring back new girls» (McNamee 1997, p. 30);

3) 32 out of 141 (which equal to 22.69 %) are short itineraries, describing brief movements within the same spatial setting. Various samples comprise the following: «James came across the yard and pushed past him into the doorway», (McNamee 1997, p. 168); «McMinn was taken into the toilets where a shot was fired into the wall beside his head», (McNamee 1997, p. 13);

4) 26 out of 141 itineraries (which equal to 18.43 %) include stops at bars and pubs. The following examples are included: «Ryan went into the Horseshoe bar and ordered a drink» (McNamee 1997, p. 21); «When Big Ivan came down to the bar that day to say that Victor wanted to talk to him he admitted to a small feeling of dread» (McNamee 1997, p. 159); «Victor was in Maxies the night they got John McGinn» (McNamee 1997, p. 25);

5) 14 out 141 itineraries (which equal to 9.92 %) are itineraries are described with clear and precise spatial indications regarding Belfast. Examples of this could be the following: «Dorcas and James came from Sailortown in the dock area», (McNamee 1997, p. 1); «On his day off Victor would go down to Crumlin Road magistrates' court», (McNamee 1997, p. 9); «When McMinn rammed him off the road in Amelia Street the reaction was extreme», (McNamee 1997, p. 13);

The results for this part indicated that treating of the itineraries is profoundly different between the journalistic corpus on *Shankill Butchers* and *Resurrection Man*, as

we can see from *Figures 3.3.* and *3.4*:
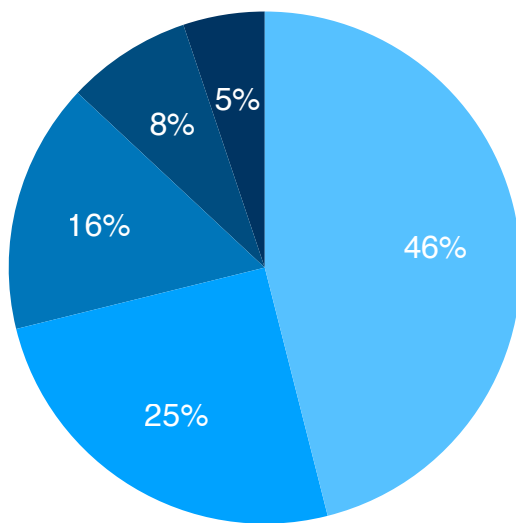
Figure 3.3 A chart visualising the percentage of itinerary categories within *Resurrection Man.*
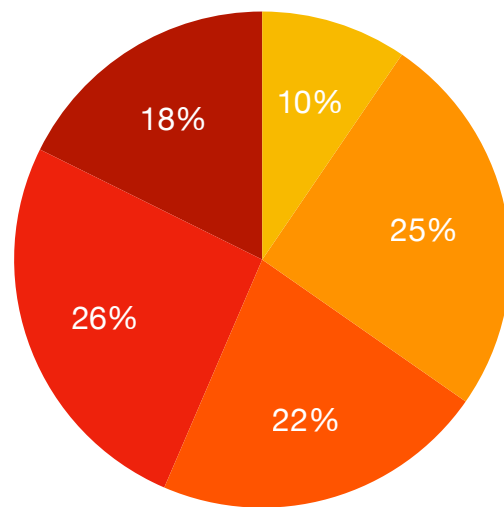
Figure 3.4 A chart visualising the percentage of itinerary categories within the Journalistic Corpus.

Literary characters navigate through a distinctly different terrain. Unlike journalistic reports, which predominantly depict individuals' journeys by clearly outlining their starting and ending points (accounting for approximately 46% of instances), *Resurrection Man* rarely pinpoints precise itineraries. Instead, the fictional space is intricately woven topologically, intertwining space with the characters' dynamics. Spatial landmarks are identified in relation to other characters' residences or significant locations, leaving the exact city location blurred. Accordingly, Belfast is mentioned only five times in the whole novel, while streets and buildings from Belfast abound.

An additional noteworthy observation highlights the pivotal role that bars and pubs play as crucial starting and ending points for character's journeys. The amount of

indicators regarding pubs and bars surpasses similar journeys in journalistic texts by more than a third. This emphasizes McNamee's intentional choice to underscore the violent environment and intense discourse that paramilitaries engage in within these drinking establishments. The strategy is designed to immerse the reader in the realm of paramilarities, exposing the reader to discourses, fear, conflicts, and intimate encounters that unfold in these settings. The intention is to evoke empathy from the reader towards the characters' inner selves, both on a personal and psychological level – an aspect previously underscored as important by Bennett (2014). In this fictional world, which thus becomes an almost timeless representation of the human being, the abstract nature of the McNamee's Belfast chronotope, alluding to a sublimated experience of pain as proposed by McGlynn (2016), is quantitatively supported by findings that indicate specific locations intentionally remain vague, shaping the fictional universe by associating spaces with characters rather than urban-centred geographical features.

3.6.5. Data analysis #5: Belfast Street Names

We previously observed that specific references to Belfast appearing in *Resurrection Man* are less prevalent than anticipated when compared to other textual genres addressing the events involving the Shankill Butchers. Conversely, there is a copious number of named street references, seemingly serving a distinct narratological purpose: they significantly contribute to reinforce the roles and functions of the characters.

To examine the relationship between the street names and the charter system, for each of the 71 mentions of street names, constituting 34 different street names, we can identify the characters who traverse those streets or at least acknowledge their existence and name them.

Out of the 34 named street names, Victor Kelly is associated with 14 (amounting to 41.17%); Ryan is linked to 8 (23.25%); the generally mentioned group of the resurrection man is connected to 8 street names (23.25%); Coppinger is linked to 5 street names (14.70%); Margaret, Ryan's former wife, is associated with 4 (14.70%);

Heather is connected to three street names (8.8%); three victims are named and associated with a single street name (2.94%).

It is apparent that Victor's familiarity with the city, as indicated by his detailed references and accounts of travels within it throughout the storyline, surpasses that of any other character. In fact, we can draw a parallel between his acquaintance with the city and his influence over it, a connection emphasized by the subsequent excerpts:

> 1) In 1969 the streets began to come alive for Victor. They appeared in the mouths of newsreaders, obscure and menacing, like the capitals of extinct civilizations. Delphi Avenue. He got a delivery job driving a lorry. During the day he would memorize a street, the derelict sites, no right turns, areas strangely compassionate under street-lights. He'd listen to the BBC in the cab. Unity Flats, Kashmir Road. The names took on an air of broken glass, bullet holes circled by chalk, burnt timber doused by rain. He felt the city become a diagram of violence centred about him. (McNamee 1997, p. 8)

> 2) They had become untrustworthy, concerned with unfamiliar destinations, no longer adaptable to your own purposes. He read the street names from signs. India Street, Palestine Street. When he spoke them they felt weighty and ponderous on his tongue, impervious syllables that yielded neither direction nor meaning. Sandy Row, Gresham Street. (McNamee 1997, p. 162)

In the initial passage, the narrator establishes a connection between Victor's eagerness to familiarize himself with the city's cartography and a propensity for violence. By committing the rundown locales to memory as potential venues for realizing his desire of death, Victor reshapes the city's mapping to align with his capacity for harm. Along this line of thought, throughout the entire novel, a consistent theme emerges – the correlation between Victor's knowledge of Belfast and his ability to exert dominance. Victor possesses the skill to pinpoint locations for the abduction of a 'Taig' with a seemingly innate understanding of this sinister geography.

This hypothesis gains additional support from the second excerpt, which describes the moment when Victor begins to lose his influence over the city and succumbs to drug

addiction. In this passage, street names lose their significance. In essence, if the sole meaning of a place is tied to the potential for violence, losing the ability to exert force translates into a loss of spatial awareness. Consequently, the names of the streets become mere empty containers of syllables.

The structural utilization of street names to illustrate the evolution of a character represents a semiotic approach entirely absent from journalistic writings, to which we compared *Resurrection Man*. Journalistic texts often overlook the consideration of structuring dynamics or internal connections within a discursive system, as they primarily focus on presenting information without extensively incorporating symbolic or metaphorical layers into language. This creates a significant distinction between literary discourse and other forms of textual expression: fiction can be arguably considered a form of discourse that introduces additional challenges when processed by algorithms, as machine are instructed to process data and consequently struggle to recognize the absence of overarching conceptual structures – absences and gaps that, as we endeavored to illustrate in the analyses within this third chapter, could serve as the focal point for achieving effective interpretation.

3.7. Conclusions: the importance of systematisation

In this chapter, we utilized a singular novel, *Resurrection Man*, as a case study to elucidate how, regardless of the complexity of the algorithms, the human interpretive part remains foundational for all five steps involved in Computational Literary Criticism.

We employed Named Entity Recognition (NER), automatic part-of-speech tagging, utilized Artificial Intelligence-oriented Language Models, leveraged gazetteer approaches, and employed other techniques for Information Retrieval mediated by the Natural Language Processing (NLP) field. However, the focal point has always been the systematization of a singular question within a comprehensive framework that could potentially lead to the discovery of something new or, at the very least, foster a fruitful dialogue with existing academic research on our literary object under investigation.

In this context, artificial intelligence can support literary investigation and aim for the retrieval of specific categories, but these categories must be inserted into a structured interpretation. Mere data retrieval does not contribute to a systemic understanding of the literary phenomena under investigation. This was particularly evident in the case at hand. Even though we could have decided to operationalize and automate the entire process, some categories emerged *during* the analysis, and it would have not been possible to anticipate, for instance, which itinerary categories we would find in the two corpora without reading all the passages entirely and devise a new schema for the itineraries of *Resurrection Man*. Human guidance is fundamental, both in all four steps of Computational Literary Criticism not involving software application and for the step of software application itself.

Above all, computers are unable to process something not present in the text. The process of reasoning about absences (as we did, for instance, with the absence of the 'Belfast' occurrence or the fact that some characters did not move from one place to another) still confers unparalleled primacy to the human mind.

In conclusion, as of now, not even the most advanced Large Language Models, such as the recently developed LLama (Touvron, 2023) or Mistral (Jiang 2023), which seem capable of performing 'intelligent' tasks, displaying some kind of emotional intelligence (Wang et al. 2023) or general understanding abilities (Bubeck et al. 2023), cannot grasp all the nuances of every communicative situation. And in fact, artificial intelligence lacks contextual competence, as noted by Davis & Marcus (2015) and Floridi (2023).

The debate about whether machines will one day be able to infer and categorize everything - even absences – is indeed an open question. However, at least for now, we must adhere to the fact that, for various reasons, it is the human governor that can make the machine produce something relevant, rather than the other way around.

Appendixes

Appendix 1 – List of consulted Computational Literary Criticism works

The complete list of works that fall under the umbrella of 'Computational Literary Criticism', selected according to the criteria stated in Chapter 1.2, is as follows.

	Authors
[1]	Ardrey, C. (2020). Visualising Voice: Analysing spoken recordings of nineteenth-century French poetry. *Digital Scholarship in the Humanities*, 35(4), 737–758.
[2]	Beausang, C. (2022). Diachronic delta: A computational method for analyzing periods of accelerated change in literary datasets. *Digital Scholarship in the Humanities*, 37(3), 644–659.
[3]	Bénard, E., and Frontini, F. (2018). Les Sganarelle De Molière: Un Nom, Des Syntaxes ? *Digital Humanities Quarterly,* 12(1).
[4]	Benotto, G.. (2021). Can an Author Style Be Unveiled through Word Distribution? *Digital Humanities Quarterly*, 15(1).
[5]	Best, M. (2005). 'Is this a vision? is this a dream?': Finding New Dimensions in Shakespeare's Texts. *Digital Studies/le Champ Numérique*, (11).
[6]	Blanke, T. (2018). Predicting the Past. *Digital Humanities Quarterly*, 12(2).
[7]	Bode, K. (2018). *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press.
[8]	Bonch-Osmolovskaya, A., Skorinkin, D. (2017). Text mining War and Peace: Automatic extraction of character traits from literary pieces. *Digital Scholarship in the Humanities*, 32(suppl_1), i17–i24.

[9]   Borin, L., Dannélls, D., Olsson, L.-J. (2014). Geographic visualization of place names in Swedish literary texts. Literary and Linguistic Computing, 29(3), 400–404.

[10]   Bradley, J. (2003). Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy': a Model for Text Analysis Tools. Literary and Linguistic Computing, 18(2), 185–207.

[11]   Brandão, S. C. S., Frota, W. N. (2017). On the path to a methodology for the critique of digital literature. Digital Scholarship in the Humanities, 32(2), 225–233.

[12]   Brill, O., et al. (2020). FAST: Fast and Accurate Synoptic Texts. Digital Scholarship in the Humanities, 35(2), 254–264.

[13]   Broadwell, Peter, Jack W Chen, and David Shepard. (2019). Reading the Quan Tang Shi: Literary History, Topic Modeling, Divergence Measures. Digital Humanities Quarterly, 13(4).

[14]   Brooke, J., Hammond, A., Hirst, G. (2017). Using models of lexical style to quantify free indirect discourse in modernist fiction. Digital Scholarship in the Humanities, 32(2), 234–250.

[15]   Bruster, D., Smith, G. (2016). A new chronology for Shakespeare's plays. Digital Scholarship in the Humanities, 31(2), 301–320.

[16]   Burgers, Johannes. (2021). Imagining the Continuously Present Past: Visualizing William Faulkner's Narratives and Digital Yoknapatawpha. Digital Humanities Quarterly, 15(2).

[17]   Can, T., Cangır, H. (2022). A warring style: A corpus stylistic analysis of the First World War poetry. Digital Scholarship in the Humanities, 37(3), 660–680.

[18]   Čermáková, A., Mahlberg, M. (2021). Gender inequality and female body language in children's literature. Digital Scholarship in the Humanities, 36(Supplement_2), ii72–ii77.

[19]   Clement, T. (2013). Distant Listening or Playing Visualisations Pleasantly with the Eyes and Ears. Digital Studies/le Champ Numérique, 3(2).

[20]   Clement, T. E. (2008). 'A thing not beginning and not ending': using digital tools to distant-read Gertrude Stein's The Making of Americans. Literary and Linguistic Computing, 23(3), 361–381.

[21]   Coffee, N., Koenig, J-P., Poornima, S., Forstall, C. W., Ossewaarde, R., Jacobson, S. L. (2013). The Tesserae Project: intertextual analysis of Latin poetry. Literary and Linguistic Computing, 28(2), 221–228.

[22]   Cumming, P. E. (2007). Digital 'Reception': Hearing 'Stereo' in Matilda through Children's Web-Based Reader Response. Digital Studies/le Champ Numérique, (10).

[23]   Daniil Skorinkin, Boris Orekhov. (2023). Hacking stylometry with multiple voices: Imaginary writers can override authorial signal in Delta. Digital Scholarship in the Humanities, fqad012.

[24]   David Thomson. (2023). Whitman Tracked Between Editions, Rossetti as a Complex Subversive, and the Collective Sense of Authorship: A Mixed Methods Accounting of a Hyperlinked "Calamus". Digital Humanities Quarterly, 17(1).

[25]   DeForest, M., Johnson, E. (2001). The Density of Latinate Words in the Speeches of Jane Austen's Characters. Literary and Linguistic Computing, 16(4), 389–401.

[26]   Dimmit, L., Kirilloff, G., Warren, C., Wehrwein, J. (2015). Exploring the Intersection of Personal and Public Authorial Voice in the Works of Willa Cather. Digital Scholarship in the Humanities, 30(suppl_1), i36–i42.

[27]   Dobson, James. (2021). Interpretable Outputs: Criteria for Machine Learning in the Humanities. Digital Humanities Quarterly, 15(2).

[28]   Edmondson, Chloe. (2017). An Enlightenment Utopia: The Network of Sociability in Corinne. Digital Humanities Quarterly, 11(2).

[29]   Elliott, J. (2017). Whole genre sequencing. Digital Scholarship in the Humanities, 32(1), 65–79.

[30]   Escobar Varela, M. (2016). The Essay/ontology Workflow, Challenges in Combining Formal and Interpretive Methods. Digital Scholarship in the Humanities, 31(1), 84–94.

[31]   Estill, L., & Meneses, L. (2018). Is Falstaff Falstaff? Is Prince Hal Henry V?: Topic Modeling Shakespeare's Plays. Digital Studies/le Champ Numérique, 8(1), 1.

[32]   Fan, L.-T. (2018). On the Value of Narratives in a Reflexive Digital Humanities. Digital Studies/le Champ Numérique, 8(1), 5.

[33]   Fink, P. (2006). The Evolution of Order in the Chapter Lengths of Trollope's Novels. Literary and Linguistic Computing, 21(3), 275–282.

[34]   Finlayson, M. A. (2017). ProppLearner: Deeply annotating a corpus of Russian folktales to enable the machine learning of a Russian formalist theory. Digital Scholarship in the Humanities, 32(2), 284–300.

[35]   Forsyth, Richard, and David Holmes. (2018). The Writeprints of Man: A Stylometric Study of Lafayette's Hand in Paine's 'Rights of Man'. Digital Humanities Quarterly, 12(1).

[36]   Francesca Frontini, Mohamed Amine Boukhaled, Jean-Gabriel Ganascia. (2017). Mining for characterizing patterns in literature using correspondence analysis: an experiment on French novels. Digital Humanities Quarterly, 11(2).

[37]   Freebury-Jones, D. (2019). "Fearful Dreams" in Thomas Kyd's Restored Canon. Digital Studies/le Champ Numérique, 9(1), 3.

[38]   Fuller, S., & O'Sullivan, J. (2017). Structure over Style: Collaborative Authorship and the Revival of Literary Capitalism. Digital Humanities Quarterly, 11(1).

[39]   Ganascia, J.-G., Glaudes, P., Del Lungo, A. (2014). Automatic detection of reuses and citations in literary texts. Literary and Linguistic Computing, 29(3), 412–421.

[40]   García, Nuria Alonso et al. (2017). A Pedagogy for Computer-Assisted Literary Analysis: Introducing GALGO (Golden Age Literature Glossary Online). Digital Humanities Quarterly, 11.

[41]   Gemma, M., Glorieux, F., Ganascia, J.-G. (2017). Operationalizing the colloquial style: Repetition in 19th-century American fiction. Digital Scholarship in the Humanities, 32(2), 312–335.

[42]   Herbelot, A. (2015). The semantics of poetry: A distributional reading. Digital Scholarship in the Humanities, 30(4), 516–531.

[43]   Herrmann, J Berenike. (2017). In a Test Bed with Kafka. Introducing a Mixed-method Approach to Digital Stylistics. Digital Humanities Quarterly, 11(4).

[44]   Hershkowitz, I. (2023). The status of the Jewish temple in modern Hebrew literature (1848–1948): a big-data analysis. Digital Scholarship in the Humanities.

[45]   Hoover, D. L. (2002). Frequent Word Sequences and Statistical Stylistics. Literary and Linguistic Computing, 17(2), 157–180.

[46]   Hoover, D. L. (2003). Multivariate Analysis and the Study of Style Variation. Literary and Linguistic Computing, 18(4), 341–360.

[47]   Hoover, D. L. (2017). The microanalysis of style variation. Digital Scholarship in the Humanities, 32(suppl_2), ii17–ii30.

[48]   Hu, Q., Liu, B., Thomsen, M. R., Gao, J., Nielbo, K. L. (2021). Dynamic evolution of sentiments in Never Let Me Go: Insights from multifractal theory and its implications for literary analysis. Digital Scholarship in the Humanities, 36(2), 322–332.

[49]   Ilsemann, H. (2008). More statistical observations on speech lengths in Shakespeare's plays. Literary and Linguistic Computing, 23(4), 397–407.

[50]   Inaki, A., Okita, T. (2006). A Small-Corpus-Based Approach to Alice's Roles. Literary and Linguistic Computing, 21(3), 283–294.

[51]   José Calvo Tello, and Juan Cerezo Soler. (2018). La Conquista De Jerusalén ¿de Cervantes? Análisis Estilométrico Sobre Autoría En El Teatro Del Siglo De Oro Español. Digital Humanities Quarterly, 12(1).

[52]   Junwu Tian, Shuyue Liu. (2022). A multidimensional and digital humanistic analysis of style in Amy Tan's novels. Digital Scholarship in the Humanities, 38(3), 1281-1295.

[53]   Karlińska, A. (2021). The art of nerves: A quantitative and qualitative analysis of drama at the turn of nineteenth and twentieth century. Digital Scholarship in the Humanities, 36(1), 122–137.

[54]   Kim, D. W. (1998). Finding the Reader in Literary Computing. Digital Studies/le Champ Numérique, (6).

[55]   Kim, S., Tak, J. Y., Kwak, E. J., Lim, T. Y., Lee, S. H. (2021). Implications of vocabulary density for poetry: Reading T. S. Eliot's poetry through computational methods. Digital Scholarship in the Humanities, 36(2), 371–382.

[56]   Klaussner, C., Nerbonne, J., Çöltekin, Ç. (2015). Finding Characteristic Features in Stylometric Analysis. Digital Scholarship in the Humanities, 30(suppl_1), i114–i129.

[57]   Koolen, C., van Cranenburgh, A. (2018). Blue eyes and porcelain cheeks: Computational extraction of physical descriptions from Dutch chick lit and literary novels. Digital Scholarship in the Humanities, 33(1), 59–71.

[58] Kubis, M. (2021). Quantitative analysis of character networks in Polish 19th- and 20th-century novels. Digital Scholarship in the Humanities, 36(Supplement_2), ii175–ii181.

[59] Kusi, D. T. (2022). A stylometric reading of William Wordsworth and W. B. Yeats's poetry: Ambivalent cravings in romantic and modernist consciousness. Research Journal in Advanced Humanities, 3(2), 28-46.

[60] Labbé, C., Labbé, D. (2006). A Tool for Literary Studies: Intertextual Distance and Tree Classification. Literary and Linguistic Computing, 21(3), 311–326.

[61] lamondon, M. R. (2009). Poetic Waveforms, Discrete Fourier Transform Analysis of Phonemic Accumulations, and Love in the Garden of Tennyson's Maud. Digital Studies/le Champ Numérique, 1(3).

[62] Land, K. (2020). Predicting Author Gender Using Machine Learning Algorithms: Looking Beyond the Binary. Digital Studies/ le Champ Numérique, 10(1).

[63] Le, X., Lancashire, I., Hirst, G., Jokel, R. (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. Literary and Linguistic Computing, 26(4), 435–461.

[64] Lee, J., Kong, Y. H., Luo, M. (2018). Syntactic patterns in classical Chinese poems: A quantitative study. Digital Scholarship in the Humanities, 33(1), 82–95.

[65] Lee, James, and Jason Lee. (2017). Shakespeare's Tragic Social Network; or Why All the World's a Stage. Digital Humanities Quarterly, 11(2).

[66] Lee, S. H., Tak, J. Y., Kwak, E. J., Kim, S., Lim, T. Y. (2021). Revisiting Sylvia Plath's and Anne Sexton's confessional poetry: Analyzing stylistic differences and evolution of poetic voice(s) through computational text analysis. Digital Scholarship in the Humanities, 36(4), 950–970.

[67] Li, D., Zhang, C., Liu, K. (2011). Translation Style and Ideology: a Corpus-assisted Analysis of two English Translations of Hongloumeng. Literary and Linguistic Computing, 26(2), 153–166.

[68] Li, M., Mansor, Z., Band, S. S. (2021). Data visualization technique to study the conceptual metaphors in Divan of Hafiz and Bustan of Sa'adi. Digital Scholarship in the Humanities, 36(4), 971–979.

[69] Liu, M., Wu, Y., Jiao, D., Wu, M. S., Zhu, T. (2019). Literary intelligence analysis of novel protagonists' personality traits and development. Digital Scholarship in the Humanities, 34(1), 221–229.

[70] Liu, X., Jin, M. (2022). A corpus-based approach to explore the stylistic peculiarity of Koji Uno's postwar works. Digital Scholarship in the Humanities, 37(1), 168–184.

[71] Maciej Janicki, Kati Kallio, Mari Sarv. (2023). Exploring Finnic written oral folk poetry through string similarity. Digital Scholarship in the Humanities, 38(1), 180–194.

[72] Mahmoudi, M. R., Abbasalizadeh, A. (2019). On comparing and clustering the alternatives of love in Saadi's lyric poems (Ghazals). Digital Scholarship in the Humanities, 34(1), 146–151.

[73] Mandell, L. (2013). How to Read a Literary Visualization: Network Effects in the Lake School of Romantic Poetry. Digital Studies/le Champ Numérique, 3(2).

[74]   Masías, V. H., Baldwin, P., Laengle, S., Vargas, A., Crespo, F. A. (2017). Exploring the prominence of Romeo and Juliet's characters using weighted centrality measures. Digital Scholarship in the Humanities, 32(4), 837–858.

[75]   Matt Morgenstern. (2023). A Keyword Analysis of "Climate Change" in Contemporary Literary Studies, 2000-2022. Digital Humanities Quarterly, 17(1).

[76]   McCarty, W. (1996). Finding Implicit Patterns in Ovid's Metamorphoses with Tact. Digital Studies/le Champ Numérique, (2).

[77]   McCarty, W. (2003). Depth, Markup and Modelling. Digital Studies/le Champ Numérique, (8).

[78]   McKenna, C. W. F., Antonia, A. (2001). The Statistical Analysis of Style: Reflections on Form, Meaning, and Ideology in the 'Nausicaa' Episode of Ulysses. Literary and Linguistic Computing, 16(4), 353–373.

[79]   Michal Škrabal, Karel Piorecký. (2022). The Corpus of Contemporary Czech Poetry: A database for research on contemporary poetic language across media. Digital Scholarship in the Humanities, 37(4), 1240–1253.

[80]   Muiser, Iwe, Mariet Theune, Jong, De, Ruud, Nigel Smink, Rudolf Berend Trieschnigg, Djoerd Hiemstra, and Theo Meder. (2018). Supporting the Exploration of Online Cultural Heritage Collections: The Case of the Dutch Folktale Database. Digital Humanities Quarterly, 11(4).

[81]   Muralidharan, A., Hearst, M. A. (2013). Supporting exploratory text analysis in literature study. Literary and Linguistic Computing, 28(2), 283–295.

[82]   Murrieta-Flores, P., Donaldson, C., & Gregory, I. (2017). GIS and literary history: Advancing digital humanities research through the spatial analysis of historical travel writing and topographical literature. Digital Humanities Quarterly, 11(1).

[83]   Mustazza, Chris. (2018). Machine-aided Close Listening: Prosthetic Synaesthesia and the 3D Phonotext. Digital Humanities Quarterly, 12(3).

[84]   Muzny, G., Algee-Hewitt, M., Jurafsky, D. (2018). Dialogism in the novel: A computational model of the dialogic nature of narration and quotations. Digital Scholarship in the Humanities, 32(suppl_2), ii31–ii52.

[85]   Nissan, E. (2017). In the Garden and in the Ark: The belles lettres, aetiological tales, and narrative explanatory trajectories—The concept of an architecture combining phono-semantic matching, and NLP story-generation. Digital Scholarship in the Humanities, 32(4), 859–886.

[86]   O'Sullivan, J., Bazarnik, K., Eder, M., & Rybicki, J. (2018). Measuring Joycean Influences on Flann O'Brien. Digital Studies/le Champ Numérique, 8(1), 6.

[87]   Opara, K. R. (2015). Grammatical rhymes in Polish poetry: A quantitative analysis. Digital Scholarship in the Humanities, 30(4), 589–598.

[88]   Pasanek, B., Sculley, D. (2008). Mining millions of metaphors. Literary and Linguistic Computing, 23(3), 345–360.

[89]   Patterson, K. (2000). The "Anna Jameson and her Friends Database": Mapping Anna Jameson's Associative Links with the Victorian Intellectual Community. Digital Studies/le Champ Numérique, (7).

[90]   Peng, C., Jung, J. J. (2021). Interpretation of metaphors in Chinese poetry: Where did Li Bai place his emotions? Digital Scholarship in the Humanities, 36(2), 421–429.

[91]   Queiroz, A. I., Fernandes, M. L., Soares, F. (2015). The Portuguese literary wolf. Digital Scholarship in the Humanities, 30(3), 388–404.

[92]   Ramsay, S. (2003). Reconceiving Text Analysis: Toward an Algorithmic Criticism. Literary and Linguistic Computing, 18(2), 167–174.

[93]   Ramsay, S. (2011). Reading Machines: Toward an Algorithmic Criticism, University of Illinois Press.

[94]   Reiter, N., Frank, A., Hellwig, O. (2014). An NLP-based cross-document approach to narrative structure discovery. Literary and Linguistic Computing, 29(4), 583–605.

[95]   Rizvi, P. (2021). Shakespeare and principal components analysis. Digital Scholarship in the Humanities, 36(4), 1030–1041.

[96]   Rizvi, Zahra, Rohan Chauhan, A Sean Pue, and Nishat Zaidi. (2022). Minimal Computing for Exploring Indian Poetics. Digital Humanities Quarterly, 16(2).

[97]   Rockwell, G. (2003). What is Text Analysis, Really? Literary and Linguistic Computing, 18(2), 209–219.

[98]   Rockwell, G. & Bradley, J. (1998). Eye-ConTact: Towards a New Design for Text-Analysis Tools. Digital Studies / Le champ numérique, 1.

[99]   Rotari, G., Jander, M., Rybicki, J. (2021). The Grimm Brothers: A stylometric network analysis. Digital Scholarship in the Humanities, 36(1), 172–186.

[100]   Ruecker, S., Radzikowska, M., Michura, P., Fiorentino, C., & Clement, T. (2008). Visualizing Repetition in Text. Digital Studies/ le Champ Numérique, (12).

[101]   Ruiz Fabo, P., Sabel, H. B., Cantón, C. M., Blanco, E. G. (2021). The Diachronic Spanish Sonnet Corpus: TEI and linked open data encoding, data distribution, and metrical findings. Digital Scholarship in the Humanities, 36(Supplement_1), i68–i80.

[102]   Rybicki, J. (2006). Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz's Trilogy and its Two English Translations. Literary and Linguistic Computing, 21(1), 91–103.

[103]   Rybicki, J. (2016). Vive la différence: Tracing the (authorial) gender signal by multivariate analysis of word frequencies. Digital Scholarship in the Humanities, 31(4), 746–761.

[104]   Schöch, Christof. (2017). Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. Digital Humanities Quarterly, 11(2), §1-53.

[105]   Schwan, Hannah, Janina Jacke, Rabea Kleymann, Jan-Erik Stange, and Marian Dörk. (2019). Narrelations — Visualizing Narrative Levels and Their Correlations with Temporal Phenomena. Digital Humanities Quarterly, 13(3).

[106]   Siemens, R. (2009). Playing 'Shame': One Technique for Introducing Text Analysis to the Literary Studies Classroom. Digital Studies/le Champ Numérique, (9).

[107]   Simpson, Erik; Brown, Hannah L.P.; Sabb, Lana; Shortell, Olly; Lee, James. (2022). Networked Cross-Dressing: A Digital Refashioning of Shakespearean Gender Subversion. Digital Humanities Quarterly, 16(3).

[108]   Sinclair, S. (2003). Computer-Assisted Reading: Reconceiving Text Analysis. Literary and Linguistic Computing, 18(2), 175–184.

[109]   Sotov, A. (2009). Lexical Diversity in a Literary Genre: A Corpus Study of the Ṛgveda. Literary and Linguistic Computing, 24(4), 435–447.

[110]   Steger, S. (2013). Patterns of Sentimentality in Victorian Novels. Digital Studies/le Champ Numérique, 3(2).

[111]   Stewart, L. L. (2003). Charles Brockden Brown: Quantitative Analysis and Literary Interpretation. Literary and Linguistic Computing, 18(2), 129–138.

[112]   Sturgeon, D. (2018). Unsupervised identification of text reuse in early Chinese literature. Digital Scholarship in the Humanities, 33(3), 670–684.

[113]   Tehrani, J., Nguyen, Q., Roos, T. (2016). Oral fairy tale or literary fake? Investigating the origins of Little Red Riding Hood using phylogenetic network analysis. Digital Scholarship in the Humanities, 31(3), 611–636.

[114]   Teresa Dobson, Sandra Gabriele, Omar Rodriguez-Arenas, Stan Ruecker, Stéfan Sinclair, Shawn DeSouza-Coelho, Alexandra Kovacs, Daniel So. (2013). Visualising Theatre Historiography: Judith Thompson's White Biting Dog (1984 and 2011) in the Simulated Environment for Theatre (SET). Digital Studies/le Champ Numérique, 3(2).

[115]   van Dalen-Oskam, K. (2013). Names in novels: An experiment in computational stylistics. Literary and Linguistic Computing, 28(2), 359–370.

[116]   Van Zaanen, M., Trollip, B., Ramukhadi, P. M., Mlambo, R. (2020). Identifying relations between characters in Afrikaans, Tshivenda, and Xitsonga books. In Annual Conference of the Alliance of Digital Humanities Organizations (ADHO) (pp. 20-25).

[117]   van-Oskam, K., van Zundert, J. (2004). Modelling Features of Characters: Some Digital Ways to Look at Names in Literary Texts. Literary and Linguistic Computing, 19(3), 289–301.

[118]   Wadsworth, F. B., Vasseur, J., Damby, D. E. (2017). Evolution of vocabulary in the poetry of Sylvia Plath. Digital Scholarship in the Humanities, 32(3), 660–671.

[119]   Wang, Y., Garjami, J., Tsvetkova, M., Hau, N. H., Pho, K.-H. (2021). Statistical approaches in literature: An application of principal component analysis and factor analysis to analyze the different arrangements about the Quran's Suras. Digital Scholarship in the Humanities, 36(1), 164–171.

[120]   Weidman, S. G., O'Sullivan, J. (2018). The limits of distinctive words: Re-evaluating literature's gender marker debate. Digital Scholarship in the Humanities, 33(2), 374–390.

[121]   Weidman, Sean, and Aaren Pastor. (2021). Modernism and Gender at the Limits of Stylometry. Digital Humanities Quarterly, 15(4).

[122]   Weingart, S., Jorgensen, J. (2013). Computational analysis of the body in European fairy tales. Literary and Linguistic Computing, 28(3), 404–416.

[123]   Wikle, O. (2020). Listening at a Distance: Reading the Sound World of Gothic Literature with Topic Modeling and Text Analysis. Digital Studies/le Champ Numérique, 10(1).

[124]   Yoo, J. J. (2021). Poetry in action: Networks of literary communication and the cultural leverage in the eighteenth-century Seoul. Digital Scholarship in the Humanities, 36(3), 766–781.

[125]   Yu, B. (2008). An evaluation of text classification methods for literary study. Literary and Linguistic Computing, 23(3), 327–343.

[126]   Abdibayev, A., Igarashi, Y., Riddell, A., & Rockmore, D. (2022). Limericks and Computational Poetics: The Minimal Pairs Framework. Computational Challenges for Poetic Analysis and Synthesis, Journal of Computational Literary Studies, 1(1).

[127] Algee-Hewitt, M. (2017). Distributed Character: Quantitative Models of the English Stage, 1550–1900. New Literary History, 48, 751-782.

[128] Algee-Hewitt, Mark, J.D. Porter, and Hannah Walser. (2020). "Representing Race and Ethnicity in American Fiction, 1789-1920." Journal of Cultural Analytics, 5(2).

[129] Allen, Colin, Hongliang Luo, Jaimie Murdock, Jianghuai Pu, Xiaohong Wang, Yanjie Zhai, and Kun Zhao. (2017). "Topic Modeling the Hàn Diǎn Ancient Classics (汉典古籍)." Journal of Cultural Analytics, 2(1).

[130] Alvarado, R., & Humphreys, P. (2017). Big Data, Thick Mediation, and Representational Opacity. New Literary History, 48, 729-749.

[131] Andresen, M., Krautter, B., Pagel, J., & Reiter, N. (2022). Who Knows What in German Drama? A Composite Annotation Scheme for Knowledge Transfer. Annotation, Evaluation, and Analysis Journal of Computational Literary Studies, 1(1).

[132] Argamon, S. E., Goulain, J. B., Horton R., & Olsen, M. (2009). Vive la Différence! Text Mining Gender Difference in French Literature. Digit. Humanit. Q., 3.

[133] Arriaga, E. A., et al. (2013). Modeling Afro-Latin American Artistic Representations in Topic Maps: Cuba's Prominence in Latin American Discourse. Digit. Humanit. Q., 7.

[134] Berglund, Karl, and Mats Dahllöf. (2021). "Audiobook Stylistics: Comparing Print and Audio in the Bestselling Segment." Journal of Cultural Analytics, 6(3).

[135] Bode, K. (2012). Reading by Numbers: Recalibrating the Literary Field. Anthem Press.

[136] Bologna, Federica. (2020). "A Computational Approach to Urban Space in Science Fiction." Journal of Cultural Analytics, 5(2).

[137]   Boot, P. (2023). "A Pretty Sublime Mix of WTF and OMG": Four Explorations into the Practice of Evaluation on Online Book Reviewing Platforms. Journal of Cultural Analytics, 7(2).

[138]   Bourrier, Karen, and Mike Thelwall. (2020). "The Social Lives of Books: Reading Victorian Literature on Goodreads." Journal of Cultural Analytics, 5(1).

[139]   Broadwell, Peter, David Mimno, and Timothy Tangherlini. (2017). "The Tell-Tale Hat: Surfacing the Uncertainty in Folklore Classification." Journal of Cultural Analytics, 2(1).

[140]   Brottrager, J., Stahl, A., Arslan, A., Brandes, U., & Weitin, T. (2022). Modeling and Predicting Literary Reception. A Data-Rich Approach to Literary Historical Reception Journal of Computational Literary Studies, 1(1).

[141]   Burgers, Johannes. (2020). "Familial Places in Jim Crow Spaces: Kinship, Demography, and the Color Line in William Faulkner's Yoknapatawpha County." Journal of Cultural Analytics, 5(2).

[142]   Burrows, J. F. (2002). "The Englishing of Juvenal: Computational Stylistics and Translated Texts." Style, 36, 677-99.

[143]   Chang, Kent K., and Simon DeDeo. (2020). "Divergence and the Complexity of Difference in Text and Culture." Journal of Cultural Analytics, 5(2).

[144]   Cheng, Jonathan. (2020). "Fleshing Out Models of Gender in English-Language Novels (1850 – 2000)." Journal of Cultural Analytics, 5(1).

[145]   Clement, T. E., Tcheng, D., Auvil, L., Capitanu, B., & Monroe, M. (2013). Sounding for Meaning: Using Theories of Knowledge Representation to Analyze Aural Patterns in Texts. Digit. Humanit. Q., 7(1).

[146]   Cordell, R. C. (2013). "Taken Possession of": The Reprinting and Reauthorship of Hawthorne's "Celestial Railroad" in the Antebellum Religious Press. Digital Humanities Quarterly, 7(1).

[147]   Dalvean, M., & Baldwin, P. (2015). Ranking contemporary American poems. Digital Scholarship in the Humanities, 30(1), 6–19.

[148]   Deijl, Lucas van der, Antal van den Bosch, and Roel Smeets. (2019). "The Canon of Dutch Literature According to Google." Journal of Cultural Analytics, 4(2).

[149]   Du, K., Dudar, J., & Schöch, C. (2022). Evaluation of Measures of Distinctiveness. Classification of Literary Texts on the Basis of Distinctive Words Journal of Computational Literary Studies, 1(1).

[150]   Duhaime, D. E. (2016). Textual Reuse in the Eighteenth Century: Mining Eliza Haywood's Quotations. Digit. Humanit. Quarterly, 10.

[151]   Ehrmanntraut, A., Hagen, T., Jannidis, F., Konle, L., Kröncke, M., & Winko, S. (2022). Modeling and Measuring Short Text Similarities. On the Multi-Dimensional Differences between German Poetry of Realism and Modernism, Journal of Computational Literary Studies, 1(1).

[152]   Erlin, Matt, Andrew Piper, Douglas Knox, Stephen Pentecost, Michaela Drouillard, Brian Powell, and Cienna Townson. (2021). "Cultural Capitals: Modeling Minor European Literature." Journal of Cultural Analytics, 6(1).

[153]   Erlin, Matt. (2017). "Topic Modeling, Epistemology, and the English and German Novel." Journal of Cultural Analytics, 2(2).

[154]   Evans, Elizabeth, and Matthew Wilkens. (2018). "Nation, Ethnicity, and the Geography of British Fiction, 1880-1940." Journal of Cultural Analytics, 3(2).

[155]   Fiormonte, D., Ricaurte, P., Chaudhuri, S. (Eds.). (2022). Global Debates in the Digital Humanities. University of Minnesota Press.

[156]   Froehlich, Heather. (2020). "Dramatic Structure and Social Status in Shakespeare's Plays." Journal of Cultural Analytics, 5(1).

[157]   Fyfe, Paul, and Qian Ge. (2018). "Image Analytics and the Nineteenth-Century Illustrated Newspaper." Journal of Cultural Analytics, 3(1).

[158]   Gavin, Michael, and Eric Gidal. (2017). "Scotland's Poetics of Space: An Experiment in Geospatial Semantics." Journal of Cultural Analytics, 2(1).

[159]   Gittel, Benjamin. (2021). "An Institutional Perspective on Genres: Generic Subtitles in German Literature from 1500-2020." Journal of Cultural Analytics, 6(1).

[160]   Gius, E., & Vauth, M. (2022). Towards an Event Based Plot Model. A Computational Narratology Approach Journal of Computational Literary Studies, 1(1).

[161]   Gold, M. K. (2012). Debates in the Digital Humanities. University of Minnesota Press.

[162]   González, José Eduardo, Elliott Jacobson, Laura García García, and Leonardo Brandolini Kujman. (2021). "Measuring Canonicity: Graduate Reading Lists in Departments of Hispanic Studies." Journal of Cultural Analytics, 6(1).

[163]   Haverals, Wouter, and Lindsey Geybels. (2021). "Putting the Sorting Hat on J.K. Rowling's Reader: A Digital Inquiry into the Age of the Implied Readership of the Harry Potter Series." Journal of Cultural Analytics, 6(1).

[164]   Hirsch, B. D. (Ed.). (2012). Digital Humanities Pedagogy: Practices, Principles and Politics. Open Book Publishers.

[165]   Howell, S., Kelleher, M., Teehan, A., & Keating, J. (2014). A Digital Humanities Approach to Narrative Voice in The Secret Scripture: Proposing a New Research Method. Digital Humanities Quarterly, 8(2).

[166]   Hoyt, E., Ponto, K., & Roy, C. (2014). "Visualizing and Analyzing the Hollywood Screenplay with ScripThreads." Digit. Humanit. Q., 8.

[167]   Igarashi, Y. (2015). Statistical Analysis at the Birth of Close Reading. New Literary History, 46(3), 485–504.

[168]   Jockers, M. L. (2013). Macroanalysis: Digital Methods and Literary History. University of Illinois Press.

[169]   Jockers, Matthew, and Gabi Kirilloff. (2016). "Understanding Gender and Character Agency in the 19th Century Novel." Journal of Cultural Analytics, 2(2).

[170]   Kelly, Nicholas M., Nicole White, and Loren Glass. (2021). "Squatter Regionalism: Postwar Fiction, Geography, and the Program Era." Journal of Cultural Analytics, 6(2).

[171]   Klein, L. F., & Gold, M. K. (Eds.). (2016). Debates in the Digital Humanities 2016. University of Minnesota Press.

[172]   Klein, L. F., & Gold, M. K. (Eds.). (2019). Debates in the Digital Humanities 2019. University of Minnesota Press.

[173]   Klein, L. F., & Gold, M. K. (Eds.). (2023). Debates in the Digital Humanities 2023. University of Minnesota Press.

[174]   Koolen, M., Neugarten, J., & Boot, P. (2022). 'This book makes me happy and sad and I love it'. A Rule-based Model for Extracting Reading Impact from English Book Reviews Journal of Computational Literary Studies, 1(1).

[175]   Kraicer, Eve, and Andrew Piper. (2019). "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction." Journal of Cultural Analytics, 3(2).

[176]   Kraus, K. (2009). "Conjectural Criticism: Computing Past and Future Texts."

[177]   Ladd, John R. (2021). "Imaginative Networks: Tracing Connections Among Early Modern Book Dedications." Journal of Cultural Analytics, 6(1).

[178]   Lavin, Matthew J. (2020). "Gender Dynamics and Critical Reception: A Study of Early 20th-Century Book Reviews from The New York Times." Journal of Cultural Analytics, 5(1).

[179]   Le-Khac, Long, and Kate Hao. (2021). "The Asian American Literature We've Constructed." Journal of Cultural Analytics, 6(2).

[180]   Lee, James Jaehoon, Blaine Greteman, Jason Lee, and David Eichmann. (2018). "Linked Reading: Digital Historicism and Early Modern Discourses of Race around Shakespeare's Othello." Journal of Cultural Analytics, 3(1).

[181]   Liddle, Dallas. (2019). "Could Fiction Have an Information History? Statistical Probability and the Rise of the Novel." Journal of Cultural Analytics, 4(2).

[182]   Long, Hoyt, Anatoly Detwyler, and Yuancheng Zhu. (2018). "Self-Repetition and East Asian Literary Modernity, 1900-1930." Journal of Cultural Analytics, 2(2).

[183]   Long, Hoyt. (2021). "Chance Encounters: World Literature Between the Unexpected and the Probable." Journal of Cultural Analytics, 6(3).

[184]   MacArthur, Marit J., Georgia Zellou, and Lee M. Miller. (2018). "Beyond Poet Voice: Sampling the (Non-) Performance Styles of 100 American Poets." Journal of Cultural Analytics, 3(1).

[185]   Mavrody, Nika, Laura B. McGrath, Nichole Nomura, and Alexander Sherman. (2021). "Voice." Journal of Cultural Analytics, 6(2).

[186]   McGrath, Laura, Devin Higgins, and Arend Hintze. (2018). "Measuring Modernist Novelty." Journal of Cultural Analytics, 3(1).

[187]   Mendelman, L., & Mukamal, A. (2021). The Generative Dissensus of Reading the Feminist Novel, 1995-2020: A Computational Analysis of Interpretive Communities. Journal of Cultural Analytics, 6(3).

[188]   Moravec, Michelle, and Kent K. Chang. (2021). "Feminist Bestsellers: A Digital History of 1970s Feminism." Journal of Cultural Analytics, 6(2).

[189]   Moretti, F. (2011). Network Theory, Plot Analysis. NLR 68, March–April.

[190]   Moretti, F. (2013). "Operationalizing." NLR 84, November–December.

[191]   Moretti, F. (2013). Distant Reading. Verso.

[192]   Moretti, F. (2020). The Roads to Rome. NLR 124, July–August.

[193]   Moretti, F., & Sobchuk, O. (2019). "Hidden in Plain Sight: Data Visualization in the Humanities." New Left Review, 118, 86-115.

[194]   Paige, N. (2018). "The Artifactuality of Narrative Form." Poetics Today, 39(1), 41-65.

[195]   Paige, N. (2020). Technologies of the Novel: Quantitative Data and the Evolution of Literary Systems. Cambridge: Cambridge University Press.

[196]   Piper, A. (2015). Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel. New Literary History, 46(1), 63–98.

[197]   Piper, A. (2018). Enumerations: Data and Literary Study. University of Chicago Press.

[198]   Piper, A. (2020). Can We Be Wrong? The Problem of Textual Evidence in a Time of Data (Elements in Digital Literary Studies). Cambridge: Cambridge University Press.

[199]   Piper, A. (2022). Biodiversity Is Not Declining in Fiction. Journal of Cultural Analytics, 7(3).

[200]   Piper, Andrew. (2016). "Fictionality." Journal of Cultural Analytics, 2(2).

[201]   Riddell, Allen, and Michael Betancourt. (2021). "Reassembling the English Novel, 1789–1919." Journal of Cultural Analytics, 6(1).

[202]   Ridolfo, J., & Hart-Davidson, W. (2015). Rhetoric and the Digital Humanities. The University of Chicago Press.

[203]   Ryan, M-L. (1999). Cyberspace Textuality: Computer Technology and Literary Theory.

[204]   Schröter, J., & Du, K. (2022). Validating Topic Modeling as a Method of Analyzing Sujet and Theme. Journal of Computational Literary Studies, 1(1).

[205]   Seminck, O., et al. (2022). The Evolution of the Idiolect over the Lifetime: A Quantitative and Qualitative Study of French 19th Century Literature. Journal of Cultural Analytics, 7(3).

[206]   Shin, H. (2022). Analyzing the Positive Sentiment Towards the Term "Queer" in Virginia Woolf through a Computational Approach and Close Reading. Journal of Computational Literary Studies, 1(1).

[207]   Simeone, Michael, Advaith Gundavajhala Venkata Koundinya, Anandh Ravi Kumar, and Ed Finn. (2017). "Towards a Poetics of Strangeness: Experiments in Classifying Language of Technological Novelty." Journal of Cultural Analytics, 2(1).

[208]   Sinykin, Dan, and Edwin Roland. (2021). "Against Conglomeration." Journal of Cultural Analytics, 6(2).

[209]   Smeets, Roel, Maarten De Pourcq, and Antal van den Bosch. (2021). "Modeling Conflict: Representations of Social Groups in Present-Day Dutch Literature." Journal of Cultural Analytics, 6(3).

[210]   So, Richard Jean, Hoyt Long, and Yuancheng Zhu. (2019). "Race, Writing, and Computation: Racial Difference and the US Novel, 1880-2000." Journal of Cultural Analytics, 3(2).

[211]   Sobchuk, O. (2016). The Evolution of Dialogues: A Quantitative Study of Russian Novels (1830–1900). Poetics Today, 37(1), 137–154.

[212]   Storey, Grant, and David Mimno. (2020). "Like Two Pis in a Pod: Author Similarity Across Time in the Ancient Greek Corpus." Journal of Cultural Analytics, 5(2).

[213]   Tatlock, Lynne, Matt Erlin, Douglas Knox, and Stephen Pentecost. (2018). "Crossing Over: Gendered Reading Formations at the Muncie Public Library, 1891-1902." Journal of Cultural Analytics, 3(2).

[214]   Tenen, D. (2018). Toward a Computational Archaeology of Fictional Space. New Literary History, 49, 119-147.

[215]   Underwood, T. (2016). "The Life Cycles of Genres." Cultural Analytics, May 23, 2016.

[216]   Underwood, Ted, and Richard Jean So. (2021). "Can We Map Culture?" Journal of Cultural Analytics, 6(3).

[217]   Underwood, Ted, David Bamman, and Sabrina Lee. (2018). "The Transformation of Gender in English-Language Fiction." Journal of Cultural Analytics, 3.

[218]   Underwood, Ted, Patrick Kimutis, and Jessica Witte. (2020). "NovelTM Datasets for English-Language Fiction, 1700-2009." Journal of Cultural Analytics, 5(2).

[219]   Vierthaler, Paul, and Mees Gelein. (2019). "A BLAST-Based, Language-Agnostic Text Reuse Algorithm with a MARKUS Implementation and Sequence Alignment Optimized for Large Chinese Corpora." Journal of Cultural Analytics, 4(2).

[220]   Vierthaler, Paul. (2016). "Fiction and History: Polarity and Stylistic Gradience in Late Imperial Chinese Literature." Cultural Analytics, May 23, 2016.

[221]   Walsh, Melanie, and Maria Antoniak. (2021). "The Goodreads 'Classics': A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism." Journal of Cultural Analytics, 6(2).

[222]   Wilkens, Matthew. (2016). "Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction." Cultural Analytics, Nov. 1, 2016.

[223]   Wilkens, Matthew. (2021). "Too Isolated, Too Insular: American Literature and the World." Journal of Cultural Analytics, 6(3).

[224]   Witmore, M. (2016). Latour, the Digital Humanities, and the Divided Kingdom of Knowledge. New Literary History, 47(2), 353-375.

Appendix 2 – Selected reference CLC works

Here, we present the list with the fifty selected reference critical works, per country, as outlined in paragraph 1.4. The studies were organized alphabetically by country.

1	Algeria	Hazlon S., Chico Y., Al-Muqaddamah A. (2021). The role digital humanities in the development of humanities, *Journal for Human and Social Studies*, Volume (06), Numero (02), Anno (2021), pp. 155-168.
2	Argentina	Menéndez de la Cuesta González, A. (2021). Modelo de análisis cualitativo con ATLAS.ti para novelas postdigitales: Reina y Game Boy. *Revista De Humanidades Digitales*, *6*, pp. 189–204.
3	Australia	Bode, K. (2018). *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press.
4	Austria	Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin. "To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730-1930). In *Digital Humanities 2018: book of abstracts*, Digital Humanities conference, 26–29 June, Mexico City.
5	Belgium	Geybels, L. (2023). Shuffling Softly, Sighing Deeply: A Digital Inquiry into Representations of Older Men and Women, in *Literature for Different Ages. Social Sciences*, 12(3), 112.

| 6 | Brazil | Brandão, S. C. D. S., & Frota, W. N. (2017). On the path to a methodology for the critique of digital literature. *Digital Scholarship in the Humanities*, 32(2), 225-233. |

| 7 | California (US) | Heuser, R., Moretti, F., & Steiner, E. B. (2016). The emotions of London. *Literary Lab*. |

| 8 | Cameroon | Kusi, D. T. (2022). A stylometric reading of William Wordsworth and W. B. Yeats's poetry: Ambivalent cravings in romantic and modernist consciousness . *Research Journal in Advanced Humanities*, *3*(2), 28-46. |

| 9 | Canada | Roberts-Smith, J., Dobson, T., Gabriele, S., Rodriguez-Arenas, O., Ruecker, S., Sinclair, S., DeSouza-Coelho, S., Kovacs, A. & So, D. (2013). Visualising Theatre Historiography: Judith Thompson's White Biting Dog (1984 and 2011) in the Simulated Environment for Theatre (SET). *Digital Studies/Le champ numérique*, 3(2). |

| 10 | Chile | Masías, V. H., Baldwin, P., Laengle, S., Vargas, A., & Crespo, F. A. (2017). Exploring the prominence of Romeo and Juliet's characters using weighted centrality measures. *Digital Scholarship in the Humanities*, 32(4), 837-858. |

| 11 | China | Tian, J., & Liu, S. (2023). A multidimensional and digital humanistic analysis of style in Amy Tan's novels. *Digital Scholarship in the Humanities*, 38(3), 1281-1295. |

12	Colombia	Grillo Gálvez, C. (2019). Sobremesa digital: objetos de texto a dato: proyecto de visualización del registro de objetos en De sobremesa. [online] Uniandes. [Last checked: 14 Nov 2023]
13	Czech Republic	Škrabal, M., & Piorecký, K. (2022). The Corpus of Contemporary Czech Poetry: A database for research on contemporary poetic language across media. *Digital Scholarship in the Humanities*, 37(4), 1240-1253.
14	Denmark	Kapitan, K. A., & Wills, T. (2023). Sagas and genre: A case for application of network analysis to manuscripts preserving Old Norse-Icelandic saga literature. *Digital Scholarship in the Humanities.*
15	Egypt	Hosny, R. (2018). Mapping Electronic Literature in the Arabic Context. *Electronic Book Review*.
16	Finland	Janicki, M., Kallio, K., & Sarv, M. (2023). Exploring Finnic written oral folk poetry through string similarity. *Digital Scholarship in the Humanities*, 38(1), 180-194.
17	France	Dealberto, C., Grandin, J., & Schuwey, C. (2022). L'atlas Molière. *Les Arènes*.
18	Germany	Zehe, A., Konle, L., Dümpelmann, L. K., Gius, E., Hotho, A., Jannidis, F., Kaufmann, L., Krug, M., Puppe, F., Reiter, N., Schreiber, A. & Wiedmer, N. (2021). Detecting scenes in fiction: A new segmentation task. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 3167-3177.

19	Greece	Koidaki, F., Christou, D., Tiktopoulou, K., & Tsoumakas, G. (2021, December). The concept of nation in nineteenth-century Greek fiction through computational literary analysis. In *Proceedings of the Workshop on Natural Language Processing for Digital Humanities*, pp. 75-84.
20	Hong Kong	Fan, L.-T. (2018). On the Value of Narratives in a Reflexive Digital Humanities. *Digital Studies/le Champ Numérique*, 8(1), 5.
21	Illinois (US)	Underwood, T. (2019). Distant Horizons: Digital Evidence and Literary Change. *University of Chicago Press*.
22	India	Rizvi, Z., Chauhan, R., Pue, A. S., & Zaidi, N. (2022). Minimal Computing for Exploring Indian Poetics. *Digital Humanities Quarterly*, 16(2).
23	Indonesia	Rokib, M., Anam, S., Rahmawati, E., Sodiq, S., & Retnaningdyah, P. (2018). Comparing genres of Indonesian literature by utilizing big data. In *IOP Conference Series: Materials Science and Engineering* Vol. 434, No. 1, pp. 012277.
24	Iran	Mahmoudi, M. R., & Abbasalizadeh, A. (2019). On comparing and clustering the alternatives of love in Saadi's lyric poems (Ghazals). *Digital Scholarship in the Humanities*, 34(1), 146-151.

25	Ireland	Howell, S., Kelleher, M., Teehan, A., & Keating, J. (2014). A digital humanities approach to narrative voice in the secret scripture: Proposing a new research method. *Digital Humanities Quarterly*, 8(2).
26	Israel	Brill, O., Koppel, M., & Shmidman, A. (2020). FAST: Fast and accurate synoptic texts. *Digital Scholarship in the Humanities*, 35(2), 254-264.
27	Italy	Rebora, S., Boot, P., Pianzola, F., Gasser, B., Herrmann, J. B., Kraxenberger, M., Kuijpers, M., Lauer, G., Lendvai, P., Messerli, T., Sorrentino, P. (2021). Digital humanities and digital social reading. *Digital Scholarship in the Humanities*, 36(Supplement_2), ii230-ii250.
28	Japan	Murai, H. (2020). Factors of the detective story and the extraction of plot patterns based on japanese detective comics. *Journal of the Japanese Association for Digital Humanities*, 5(1), 4-21.
29	Kazakhstan	Rakymzhan, O., Shalabay, B., Zhumagulova, O., Kazhibaeva, G., & Kairbekova, I. (2022). Existential Loneliness in Kazakh Literature and Modern World Literature: Insights from Metaphors and Frame-Based Analysis. *International Journal of Society, Culture & Language*, 10(3), 103-116.

30	Kyrgyzstan	Asanov, T., Duishembieva, J., Dwyer, A., Isaeva, A., Prior, D., Wegner, A. L., Storsved, B., Hawk, P. & Qu, M. (2021). Early Modern Kyrgyz Oral-derived Narrative Sources. Miami University Libraries. https://emkons.lib.miamioh.edu/ [Last checked 1 Nov 2023]
31	Latvia	Berglund, K., La Mela, M., Zwart, I., & La Monica, C. (2022). DHNB 2022 Conference: Book of Abstracts. In Digital Humanities in *the Nordic and Baltic Countries (DHNB) 6th Conference*, 15-18 March 2022, Uppsala. Centre for Digital Humanities Uppsala/Department of ALM.
32	Lebanon	Mourad, G. (2021). Digital Transformation and the Future of Social Sciences and Humanities. *Istishraf for Future Studies*, 6(6), 237-260.
33	Malaysia	Li, M., Mansor, Z., & Band, S. S. (2021). Data visualization technique to study the conceptual metaphors in Divan of Hafiz and Bustan of Sa'adi. *Digital Scholarship in the Humanities*, 36(4), 971-979.
34	Mexico	Méndez, R. L. (2020). Análisis distante en un corpus bilingüe: el caso de Rosalía de Castro. *Publicaciones De La Asociación Argentina De Humanidades Digitales*, *1*, e007.
35	Netherlands	Smeets, R., De Pourcq, M., & van den Bosch, A. (2021). Modeling Conflict: Representations of Social Groups in Present-Day Dutch Literature. *Journal of Cultural Analytics*, 6(3).

36	New Zealand	Ciccoricco, D. (2021). Through the Portal, and What Michael Joyce Found There. *Configurations 29*(2), 179-199.
37	Niger	Waliya, Y. J. (2018). «C'est tout bon!» de Corinne Albaut comme poésie numérique: Relation entre le texte et l'image. *Calabar Journal of The Humanities*, 13(1), 126-141.
38	Peru	Uenishi, R., Ortega, C., Pérez Martinez, Á., Rodríguez-Serra, M., & Elías, P. (2022). A systematic approach to literature analysis: traveling through stories. *Digital Scholarship in the Humanities*, 37(2), 565-579.
39	Poland	O'Sullivan, J., Bazarnik, K., Eder, M., & Rybicki, J. (2018). Measuring Joycean influences on Flann O'Brien. *Digital Studies/Le Champ Numérique*, 8(1), 6-1.
40	Portugal	Queiroz, A. I., Fernandes, M. L., & Soares, F. (2015). The Portuguese literary wolf. *Digital Scholarship in the Humanities*, 30(3), 388-404.
41	Russia	Skorinkin, D., & Orekhov, B. (2023). Hacking stylometry with multiple voices: Imaginary writers can override authorial signal in Delta. *Digital Scholarship in the Humanities*, fqad012.
42	Singapore	Varela, M. E. (2021). Theater as Data: Computational Journeys into Theater Research. *University of Michigan Press*.

43	South Africa	Van Zaanen, M., Trollip, B., Ramukhadi, P. M., & Mlambo, R. (2020, July). Identifying relations between characters in Afrikaans, Tshivenda, and Xitsonga books. In *annual conference of the Alliance of Digital Humanifies Organizafions (ADHO),* pp. 20-25.
44	South Korea	Kim, S., Tak, J. Y., Kwak, E. J., Lim, T. Y., & Lee, S. H. (2021). Implications of vocabulary density for poetry: Reading TS Eliot's poetry through computational methods. *Digital Scholarship in the Humanities*, 36(2), 371-382.
45	Spain	Gervas, P. (2009). Computational Approaches to Storytelling and Creativity. *AI Magazine*, *30*(3), 49.
46	Sweden	Borin, L., Dannélls, D., & Olsson, L. J. (2014). Geographic visualization of place names in Swedish literary texts. *Literary and linguistic computing*, 29(3), 400-404.
47	Switzerland	Falcetto, B. (2019). Vedere i testi, leggere le immagini: le visualizzazioni per l'"Atlante Calvino". In Tuttestorie di donne (pp. 203-206). il Saggiatore-Fondazione Arnoldo e Alberto Mondadori. https://atlantecalvino.unige.ch [Last checked 2 Nov 2023]
48	Turkey	Can, T., & Cangır, H. (2022). A warring style: A corpus stylistic analysis of the First World War poetry. *Digital Scholarship in the Humanities*, 37(3), 660-680.

| 49 | United Kingdom | McCarty W. (2003). Depth, Markup and Modelling. *Digital Studies/le Champ Numérique*, (8). |
| 50 | United Arab Emirates | Wrisley, D. (2017). Locating Medieval French, or Why We Collect and Visualize the Geographic Information of Texts. *Speculum*. 92. |

Appendix 3 – In-depth analysis of the reference works

Here, we offer a thorough analysis of every step in Computational Literary Criticism, with a specific emphasis on the fifty selected reference papers.

As certain terms may be considered technical, we have included a technical glossary in appendix 4. Each technical term is highlighted in blue. Therefore, should you have an interest in consulting appendix 3, we recommend checking appendix 4 for the corresponding technical terms.

The in-depth examination encompasses all five steps presented in paragraph 1.4, and the breakdown of the analysis is as follows:

Critical work	Abstracted concept	Literary objects to be investigated	Software application	Explicanda	Final interpretation
Algeria	The paper discusses various methods of using computation in the social sciences. When it comes to literature, it primarily focuses on concordances as a means of investigating authorial style.	The paper provides a comprehensive overview of the Digital Humanities field and discusses the connection between literary studies and computational linguistics. It specifically mentions the work of Donald Wayne Foster, which centers around Shakespeare's works	The paper mentions a traditional pipeline for Digital Humanities projects that is also applicable to Literary studies: data collection, digitization, encoding, data extraction and organization, and data interpretation and analysis. There is a particular emphasis on spatial and temporal analysis using geographic information systems (GIS) and data visualization techniques, which are employed to enhance the understanding and representation of the data.	The paper discusses a range of visualization options within the broader context of Digital Humanities. It specifically highlights network theory softwares like Gephi, examines the advantages and disadvantages of visualization in uncovering hidden patterns, and mentions GIS systems.	The paper mentions the theoretical step of data interpretation and analysis, highlighting the necessity of interpreting, for instance, topics extracted with topic modelling or reflecting on findings through logical, statistical, or quantitative analysis, according to the researcher's specialisation. However, it does not provide any case studies.

Argentina	The scholars present a set of categories related to the post-digital age as defined. They investigate the presence of topics such as the internet, social networks, AI/ robotics, among others.	The corpus includes *Game Boy* by David Parkas (2019) and *Reina* by Elizabeth Duval (2020).	The research leverages 'Atlas.ti', a platform that enables digital textual annotation. The researchers select a class of objects/ categories that they want to trace from the text and annotate that same text digitally using Atlas. The software is then responsible for visualizing these findings.	The platform 'Atlas.ti' offers both a web platform for visualizing notes alongside annotations and the capability to visualize data in Sankey diagrams.	The analysis points to the rise of postdigital literature in Spanish, which is characterized by seamlessly incorporating postdigital technologies into narratives, with a notable emphasis on references to social networks. It highlights how the former exhibits a diverse range of temporal references to technology, representing the author's generation's transitional position between the digital and postdigital eras. The narrative's coexistence of digital incipient technology, postdigital elements, and pre-postdigital elements reflects varying adoption speeds and overlapping timeframes.

| Australia | The study aims to investigate the relationship between nationality and the publishing success of novels by reviewing Australian national newspapers. More specifically, it seeks to address the concept of literary anonymity, editorial success, and gender or national attribution through digital means. | The author investigates newspaper fiction from the Trove Archive, which is part of the National Library of Australia. They analyze 21,000 novels, novellas and short stories published in early Australian newspapers published from 1803 to 1982. | The computational aspect of the work primarily focuses on network analysis, intended to explore the data rather than making quantitative claims about it. Instead of using network visualizations, the author chooses to chart the proportions of serialized and reprinted fiction in various categories, such as Australian vs. British or metropolitan vs. provincial. In the final chapter, the critical analysis employs a combination of topic modeling and decision trees to determine the characteristics that distinguish American, Australian, and British fiction in colonial newspapers, if any. | First, the monograph presents tabular data and conceptual schemes to show how the corpora are composed and the pipeline of their curation process (i.e., the steps involved in transforming Trove newspapers into a curated dataset). Secondly, the book presents, on one hand, line charts, histograms, or a combination of the two to compare characteristics of the corpora over time based on categories such as author nationality, gender, reprints, and other features. On the other hand, it employs decision trees to assess the relationship between gender and various topics extracted using topic modeling. | The book's most significant point highlights the importance of curating corpora more precisely, as they often exhibits excessive bias in terms of provenance, tradition, or authorial gender. In addition to this main argument, the monograph also offers non-methodological interpretive insights, revealing findings about the proportions of serialized and reprinted fiction across various categories (e.g., Australian vs. British or metropolitan vs. provincial). The findings illustrate the significance of provincial networks in the publication and circulation of Australian fiction during the investigated period. These networks acted as a counterforce to dominant metropolitan syndicates, which imported British and American fiction into the colony. Furthermore, the author explores gender using diverse machine learning algorithms and discovers that topics align more closely with an author's national identity than with their gender. For instance, the Australian literary voice is most strongly characterized by topics derived from prominent descriptions of non-metropolitan colonial settings, characters, and activities, and it depicts Aboriginal characters more than previous scholarship had recognized. |

Austria	The aim of the paper, in conceptual terms, is to enhance the comprehension of the dimensions of quantitative character analysis and provide diachronic and typological insights into quantitative dominance relations. The paper explores the concept of character along two dimensions: the prominence of the character, measured by the number of words expressed by that same character, and the character's interactions, examined through network-based measures.	For their analysis, the research uses a corpus of 465 German-language dramas published between 1730 and 1930.	The computational methods used in this study involve the quantitative analysis of characters in dramatic texts using various measures and techniques, such as count-based measures (the number of scenes a character appears in, the number of speech acts performed by a character, etc.) and network-based measures (the analysis of the relationships and interactions between characters in the drama).	The results of the analysis aimed to investigate the typology of quantitative character distribution in dramatic texts are visually represented through the following: 1) A Spearman's rho graph illustrating the correlation between count-based and network-based measures; 2) Figures showing the rankings of characters for five specific dramas based on count-based and network-based measures; 3) Network graphs representing the relationships and connections between characters for specific dramas; 4) The percentage of characters in the upper quartile according to their eigenvector centrality, grouped by decades; 5) The percentage of upper quartile characters according to the number of words, grouped by decades; 6) Decile distribution of weighted degree and number of words for specific dramas.	The paper examines the quantitative character dominance relationships in German-language drama from 1730 to 1930 by exploring different count-based and network-based measures. The results suggest that network-based measures are more responsive to clustering patterns in secondary scenes and less dominant characters, while count-based measures are better suited for quantitatively dominant characters. Furthermore, the research observes shifts in character dominance over the years, revealing a rise in the prevalence of quantitatively less dominant characters over time.

Belgium	The investigation focuses on the representation of gender, with a particular focus on the stereotypes associated with older women. The aim is to argue that not only children's literature, but also young adult and adult fiction play a significant role in age socialization for their respective target audiences.	The corpus analyzed consists of 41 contemporary books published between 1975 and 2021. These books were written by Dutch and Flemish authors such as Bibi Dumon Tak, Ed Franck, Guus Kuijer, Bart Moeyaert, Hilde Vandermeeren, and Joke van Leeuwen. They are intended for readers aged between 7 and 12 years old.	The article employs a variety of computational methods from the field of digital humanities. These techniques include digitization with OCR, text annotation using XML, and the identification and classification of characters based on attributes such as name, gender, ethnicity, and age. Once the characters have been retrieved, a parser with Spacy is used to identify relevant information, particularly for the older characters, by analyzing syntactic dependencies.	The analysis results are presented in a tabular form, showing word lists of verbs, possessions, adjectives, and adverbs that are associated with older adult characters in various categories of literature, such as children's literature, young adult literature, and adult literature.	The examination uncovers disparities in how older adult characters are portrayed based on gender. Older women's mobility is depicted less frequently compared to older male characters. Men are often linked to laborious movements, which both confirm and challenge the stereotype of physical decline. Additionally, children's books focus on the longevity of older adults' lives without strongly associating it with mortality. In contrast, young adult novels begin to introduce references to death, and adult novels further establish the link between aging and dying. Notably, portrayals of loneliness in older age are less common in children's books, highlighting a contrast in representation between literature for children and adults.
Brazil	They refer to two sets of approaches, both using linguistic constituents as symbols of stylistic traits. The first approach is heavily based on statistical rules, and this research approach is known as lexicometry, which refers to someone dealing with the ranking of frequencies, corpus richness or variety, length of sentences or fragments, frequency of punctuation marks, etc. The second type of research, like the one that studies style, is grounded in the general rules of style, which revolve around word concordances.	The study mentions several examples of stylometry/ lexicometry and studies that investigate various literary objects, ranging from biblical texts to the works of Charles Dickens and Shakespeare, as well as the Brown Corpus, which contains 126 novels spanning across multiple genres.	The paper discusses methods in research automation in the field of humanities, with a particular focus on stylometry and lexicometry. It highlights the differences between stylometry, which involves studying linguistic style and authorship attribution through statistical analysis (examining aspects such as word choice, sentence structure, and punctuation usage), and lexicometry, which focuses on the quantitative analysis of texts, specifically the frequency and distribution of words and lexical units.	The text discusses the utilization of statistical and mathematical techniques, data mining software, and the significance of documenting each step of the research process. This implies the establishment of a fundamental procedure for visualizing data and interpreting results via data visualization.	Instead of focusing on specific examples, the paper emphasizes the intricacies involved in conducting research in the humanities, especially in literature. This field lacks a consensus among scholars regarding a standardized methodology for studying literary texts, ranging from classic hermeneutics to Computational Literary Criticism. Interestingly, the authors point out that after the data modeling phase, which could lead to visualization, there comes an evaluation phase. If the data or exceptions are considered errors, the critical process returns to the modeling phase.

California (US)	The paper explores the relationship between emotive valence and London settings as depicted in an English fiction corpus spanning from the eighteenth to the nineteenth century.	The corpus comprises around 5,000 English novels published between 1700 and 1900. It encompasses 304 novels from the period 1700-1749, 1,079 novels from 1750-1799, 1,290 novels from 1800-1849, and 2,189 novels from 1850-1899. To assemble the corpus, we utilized texts classified as fiction by the English Short Title Catalogue for the eighteenth century. Additionally, we incorporated texts from the Literary Lab's 'Eighteenth-Century Fictional Marketplace' project. For the nineteenth century, we sourced texts from the Chadwyck-Healey database and a collection made available on the Internet Archive by the University of Illinois. A small number of additional texts were obtained from the Gutenberg Project, Google Books, or Stanford library scans.	They used Named Entity Recognition (NER) to find place names, and then they performed sentiment analysis in linguistic contexts defined as 100 words before and after the London-related place names.	The paper provides: Scatterplots showing geographic place-names and London place-names as a percentage of the words in our corpus; Maps to visualize the urban growth of London in two centuries or to visualize areas of London predominantly connected with a certain emotion; Bubble maps that blend corpus linguistics and geography by showing on a map how many times place-names are mentioned per million words, how many times a part of the city is mentioned depending on the genre or author, or how intensely an emotion-related constituent is mentioned to describe a certain place in London; Creative graphs to visualize the relationship between the frequency of boroughs per million mentions of London toponymies and the population of each borough as a percentage of the total population.	The study's insights reveal a striking contrast between the evolving real London and its relatively stable fictional representation in 19th-century literature. Despite significant changes in the city, literary portrayals predominantly centered on the City and the West End, overlooking areas witnessing substantial population growth. In terms of the semantics of different areas in London and how they were represented in fiction, The City's enduring fictional presence was due to its heterogeneous components, while the West End's representation arose from its homogeneity and exclusivity. Moreover, the study highlights a prevailing emotional neutrality in the narrative, as emotions were often muted when literary action took place in public spaces.
Cameroon	The paper investigates how two authors represent two variants of the same type of consciousness, where awareness of the world is linked to its aesthetic appreciation. More specifically, aesthetics are investigated in poetry as an instrument of vision through which nature and humanity can be appreciated and perfected.	In the given paper, the literary corpus pertains to the collective works of William Wordsworth and William Butler Yeats, which includes their poems and other writings.	The researcher performed stylometric analysis using R-Stylo and utilized Voyant tools to analyze the stylistic traits related to conflicting desires and emotions experienced by the poets William Wordsworth and William Butler Yeats within the context of their romantic sensibilities.	The results of these analyses are visually presented through graphs or charts, showcasing statistical data, word frequency distributions, or stylistic patterns. These presentations aim to highlight stylometric characteristics and display the results of a comparative analysis of the aesthetic themes within the works of Wordsworth and Yeats.	The paper concludes that Wordsworth and Yeats's sublime consciousness, to a certain extent, illustrates the contrasting emotions of the artists and their characteristics, placing them beyond the confines of their era.

Canada	The study addresses the distinction between theatrical texts and other forms of literary texts. In terms of conceptual aspects, their objectives of reflection are the following: 1) Timelines; 2) Characters: The study implements a 'Character View' that provides a list of characters and indicates their presence on stage, whether speaking or silent, at any given point in the text or performance; 3) Theatrical space: It offers a 'Stage View' which includes a scaled, three-dimensional model of a performance setting. This model is populated by abstract avatars that can be manipulated within the space and optionally supplemented with text or images.	They have at their disposal a corpus of motion picture screenplays, composed of character dialogue as well as descriptive and background metadata provided in a film script. The paper presents different case studies, some focusing on a single movie screenplay, while others examine multiple (though not numerous) screenplays.	The researchers have introduced their new tool, *ScripThreads*, which parses screenplays to extract information about characters, scenes, and meta-information. It then generates statistical values for analysis and provides four distinct types of visualizations, each serving its own purpose.	Visually, the ScripThreads software presents data through various types of graphs: 1) Force-Directed Graph: This graph represents the relationships between characters in a screenplay. Characters are represented as nodes, and their connections are shown as edges; 2) Absence Graph: The absence graph visualizes the presence and absence of characters over time. It represents characters as threads or routes on a bus map. When characters are present in a scene, their threads run parallel, and when they are absent, their threads split off; 3) Presence Graph: The presence graph provides a quick overview of when a character is active in a scene. The size of the thread representing a character is wider when the character is active and narrower when inactive. Time is shown on the y-axis, and horizontal lines indicate dialogue between characters; 4) Increasing Graph: The increasing graph represents character activity and storytelling techniques across the narrative. It shows the vertical increase of character threads to indicate their presence in scenes over time.	The article presents ScripThreads, a tool created to analyze screenplays and visually represent character interactions over time, aiming to demonstrate its usefulness. Using the analysis of 'The Big Chill' as an example, the tool illustrates a close grouping of character threads, underlining frequent interactions among the ensemble cast in various settings throughout the film. In contrast, when analyzing 'Grand Canyon,' the tool reveals a more unequal distribution of character presence, with the protagonist Mack appearing in 50% more scenes than any other character, highlighting the film's emphasis on his journey.

| Chile | The paper examines the relationships between dramatic characters using Social Network Analysis (SNA), which involves analyzing social behavior by representing individuals (referred to as 'actors') as nodes and their social relationships as links or bonds. It then compares the literary works of authors, including Shakespeare, with those of other notable authors. | They investigate Shakespeare's play, *Romeo and Juliet*. | The study employed graph theory to construct a character network using turn-taking communication as a foundation. After constructing the character network, the researchers utilized weighted centrality measures to assess the significance of each character. To compare the rankings of characters based on different centrality values, the researchers performed cluster analysis using the K-means algorithm. | The paper presents various network visualizations, including visualizations depicting the turn-taking in Romeo and Juliet. These visualizations are scaled according to the number of bonds, with brighter nodes indicating a higher number of links to other nodes. Additionally, there is another visualization where the thickness of the link represents the strength of the connection between the joined nodes. | After conducting a study that applies Social Network Analysis (SNA) to explore the significance of characters in Shakespeare's play Romeo and Juliet, the findings indicated that Romeo achieved the highest scores in all three centrality measures. The discussion delves into the reasons behind Romeo and Juliet's prominence, highlighting that Juliet's centrality is even higher when considering the type and strength of the network bonds, rather than solely focusing on the quantity of her interactions with other characters. This is accomplished by employing the 'who-talked-after-whom' heuristic, which illuminates the social context of women's roles during Shakespeare's era and confirms the prevailing notion that women were expected to have fewer social engagements. |
| China | Drawing on Douglas Biber's multidimensional analysis, the researchers investigate a set of core lexico-grammatical features to identify stylistic variations. | The corpus consists of Amy Tan's six novels in English. | The authors utilized MAT, which stands for Multidimensional Analysis Tagger, specifically version 1.3, which is a software tool that facilitates the application of Biber's model to each corpus of Tan's novels | The data is presented in tabular and graphical form. The tables display the scores and statistics for each dimension and the novel, based on Biber's model of multidimensional analysis. Meanwhile, the figures, including line charts and a correlogram, provide visual representations of the scores across each dimension. | The analysis reveals that Tan's novels generally that her texts contain linguistic patterns that make them more affective and interactional. The novel *The Kitchen God's Wife* stands out as being more informational. Also, her come out to be works narrative-driven, thus containing considerable reference to past time, third-person animate referents, reported speech, and depictive details. In terms of style, Tan stand out for directness and clarity in language, showing a preference for specific and tangible descriptions rather than abstract concepts. |

Colombia	The research investigates the presence of objects interpreted in the light of their symbolic value, in terms of literary style, religious aspects, and reflection of characters' identity, within the literary poetics of José Asunción Silva.	The corpus comprises the posthumous novel *De Sobremesa* by José Asunción Silva.	The author uses XML markup to identify and categorize 496 objects in the reference literary work, and employs a visual analytics approach to visualize the data using software.	The project provides a visualization for the diary pages with colored points to indicate when an object was found. Then, it offers two visualizations to show the results of the annotation project: 1) a normal sankey diagram, offering a visualization for the objects present on each day of the diary; 2) a circular sankey diagram, showing the entire collection of objects in the literary work.	By analyzing the presence of furniture and decorative items in various cities where the diary entries are written, the paper reveals that Paris and London have the highest number of objects related to their urban context. This leads to a new hypothesis regarding the importation of furniture and luxury goods into Bogotá in the 19th century.
Czech Republic	The paper discusses the initial concepts of categorization upon which the corpus is divided. These include distinctions between printed and web pages, authors' gender, year of publication, and other filters. Following that, the mining tools begin with keyword-in-context tools, which are used to identify foreignisms and multilingualism, as well as any other concept chosen by the user.	They analyze the Corpus of Contemporary Czech Poetry, a database that currently contains approximately 35.5 million running words. The print poetry subcorpus includes about 1.7 million words, taken from 21,478 poems printed in 496 poetry collections by 209 authors. The web component of the corpus comprises over 442,000 poems from six literary forums.	The computation methods used in the described project involve several steps: 1) Firstly, the printed poetry books were scanned and digitized using Optical Character Recognition (OCR) technology; 2) Secondly, the digitized texts were converted into standard XML (Extensible Markup Language) format, enabling a structured representation of the text with tags and metadata; 3) For corpus management, the project utilizes a corpus manager tool called 'KonText.' This tool allows users to search for specific words or phrases, filter the search based on metadata, and analyze concordance lines; 4) To perform automatic rhyme analysis, the project employs a tool called 'Gunstick.' This tool filters and searches for specific rhymes based on word forms or lemmas; 6) Lastly, the project includes visualization tools that provide a summary of the applications and offer visualization options.	The paper presents a bar chart to visualize the composition of the corpus, an area chart for the work on capitalism, displays the KonText interface (expanded view showing the found keyword in the individual poem), pie charts to visualize rhyme analysis and the percentage of each word used as a rhyme, a timeline area chart, and a map of Czech Verse. The map is based on data from CCV, where the map pins represent natural objects such as peaks, mountain ranges, protected areas, volcanoes, deserts, rivers, canals, and streams, as well as settlements such as cities and villages. Circles on the map represent countries.	The paper does not explicitly offer in-depth literary interpretations of particular poems or authors in the Corpus of Contemporary Czech. Instead, it primarily centers on methodological observations, the advancement of data mining tools, and the potential applications of the project for linguistic and literary analysis. The platform and its tools create a fresh arena for critical and theoretical analysis, to be conducted by other researchers in the community.

Denmark	The study delves into the analysis of relationships among literary genres in order to broaden the understanding of the connections between the works, their transmission, and potential modes of reception.	They utilize bibliographical and textual data gathered from four prominent online databases concerning Old Norse-Icelandic manuscripts from the Medieval era: Handrit (providing descriptions of 9,137 manuscripts), the Old Norse Prose Register (featuring descriptions of nearly 5,000 manuscripts), the Stories for All Time project (consisting of 818 manuscript descriptions), and the Skaldic Project (comprising 5,797 verses by 447 skalds preserved in 718 manuscripts).	The software application involves two steps. Firstly, data curation: the manuscript databases, where literary works were categorized by genre, were imported and normalized. This means that no literary work could be labeled as belonging to more than one genre. After the manuscript data was corrected and normalized, a network visualization was performed using the software Gephi to visualize the connections between literary works and genres. This was done by assuming that manuscripts mentioned in the same catalog belong to the same genre.	The data in the study is presented visually through network visualizations. The network consists of nodes and edges, representing literary works and their co-occurrences in catalog manuscripts, to test their relevance to a category. The nodes represent the literary works, and the edges represent the connections between these works. The edges are undirected and have weights indicating the number of manuscripts in which two works appear together.	The research discusses the findings of a network analysis conducted on Old Norse-Icelandic texts in their manuscript context. The analysis reveals the presence of two main groups: one includes romances like legendary sagas (Fornaldarsögur) and chivalric sagas (Riddarasögur), and the other comprises pseudo-historical works such as the sagas of Icelanders (Íslendingasögur) and kings' sagas (Konungasögur). Eddic poems (Rímur) are shown to have connections with both communities, particularly with Hervarar saga. In conclusion, this study challenges the traditional genre distinctions, highlighting the flexibility of categories during manuscript transmission.
Egypt	The study focuses on the lack of Arab and Middle Eastern electronic literature coverage in academic research, by investigating the amount of electronic literary works language distribution in the biggest corpus of Electronic Literature at their disposal, discussing the causes and consequences of such a lack.	The paper focuses on the third volume of the Electronic Literature Collection, which is the official volume by the Electronic Literature Organization. This volume was issued in February 2016 and includes 114 works from 26 countries in 13 languages.	The computational methods mentioned in the essay include data analysis and visualization techniques, such as map of the geographic distribution of authors in the Electronic Literature Collection volume 3.	The paper's analyses are accompanied by the following: 1) Maps that show the origins of the literary works included in the corpus; 2) A pie chart that represents the percentage of authors' nationalities, and an interactive world map that displays the global distribution of e-lit artists. When hovering over each country with the mouse, the name of the country and the number of its e-lit artists are displayed; 3) A graph that illustrates the distribution of e-lit writers in Arabic countries.	Following an extensive analysis of the Electronic Literature Collections, the research paper points out a historical lack of proper representation for Arabic electronic literature in major critical arenas. The study draws attention to a language-based bias in the field of e-literature, where certain languages, such as English, receive substantial coverage, whereas others, such as Arabic, Chinese, Dutch, and Swedish, are given scant attention.

Finland	The paper is exploring in an abstract manner the frequent repetition of similar text pieces across different levels (including entire poems, passages, verses, and collocations). Due to significant variations in spelling, morphology, and composition, identifying these similarities poses a challenge, therefore they present a method for automatically detecting equivalent verses.	The literary object investigated is the Suomen Kansan Vanhat Runot (Old Poems of the Finnish People), a collection of nearly 90,000 oral folk poems written down between 1564 and the early 20th century.	The computation methods described in the paper involve similarity search and clustering techniques for analyzing verses in a corpus of metric poetry. First, they represent each verse as a vector in a high-dimensional space. Then, the approach utilizes vector representations, cosine similarity, and graph clustering to identify similar verses and clusters of equivalent verses within the corpus of metric poetry.	The data regarding the computation of verse similarity is presented in tabular form. The paper also provides a network visualization, where the thickness of the edges represents the degree of similarity, and the color represents the resulting clusters (the same color indicates the same cluster). Additionally, the paper offers a web user interface, allowing researchers to browse through the automatically identified intertextual links, enabling side-by-side visualization of poems for comparison.	The paper mainly focuses on the technical evaluation of verse alignment and does not leave much space for actual literary interpretation. The work serves as a foundation for others to interpret: their platform allows researchers to look for similar verse types and variations, recognizing texts that share some literary connection.
France	The monograph leverages Schuwey's expertise in French Literature and marketing techniques, as well as fashion in French classical literature, and tackles several key concepts about the editorial system, such as the salary differences between editorial-related jobs, among others. Additionally, it explores the materiality of theaters in the 17th century and the gender of dramatic characters. The investigated concepts are related, so to speak, to both the societal and infrastructural environments in which aspects of the literary events took place and to literary aspects of Moliere's pieces.	The literary works under investigation are Molière's comedies.	The software utilisation concerns mainly visualisation, and consists of the digital production of 150 infographic maps representing a diversity of aspect of Moliére comedies. A abundant quantity of bibliographical is stored with traditional methods, and the represented with digital tools that construct clear infographic.	Atlas Molière presents itself as a visual encyclopedia on the life, work, and era of Molière, with 150 cards and infographics that draw from the expertise of Clara de Alberto in information design and Jules Grandin in digital cartography. The work is a relevant example of leveraging a clear division within a DH project between literary and visualization aspects.	The monograph provides insights into the editorial system surrounding the production and staging of Moliére's comedies, offering a reevaluation of literary works by focusing on the external environment. For instance, by examining extra-textual aspects of literature, such as salary differences between theatre-related jobs or the geographical position of theatrical traditions, they gain insights about Moliére's entrepreneurial ability in managing collaborators and leveraging the actual theatrical environment proficiently.

| Germany | The study addresses the possibility of identifying scenes in literary texts by drawing on Gerard Genette's conceptualization of scenes. The concept was further subdivided into four main elements: time, space, action, and characters. The group of researchers asserts that by examining the concept of scenes, they can effectively explore the concept of fictional plots. | The corpus they describe consists of German-language dime novels (550 k tokens). | The authors describe how they used software to evaluate scene segmentation, a task that involves dividing a text into different scenes, as defined by Genette. To assess the difficulty of scene segmentation, the authors evaluated multiple baselines on their dataset. They employed two unsupervised standard segmentation techniques (TextTiling and TopicTiling) and then a supervised baseline with BERT (Bidirectional Encoder Representations from Transformers), a pre-trained language model. They fine-tuned the BERT model for binary scene segmentation by training it on a training sample consisting of a target sentence, context, and label. | The research presents the following charts: 1) A box plot to illustrate the difference in length between scenes and non-scenes; 2) A line chart that shows the distribution of scenes along the text flow, indicating the relationship between the number of sentences passed in a novel and the number of scenes found up to that point; 3) Several box plots displaying the distribution of locational (left) and temporal (right) adverbials at the beginning of a scene, as well as their correlation with annotated reasons for a scene change; 4) A line chart depicting the distribution of explicit character references within a scene based on their positions; 5) A table presenting the results of applying BERT to identify borders between two scenes and between scenes and non-scenes. | Instead of providing literary interpretations of the corpus of interest, the study explores the limitations of standard unsupervised methods, such as the BERT model, in identifying scene changes in novels for text segmentation. The study concludes that these subtasks of Natural Language Processing still require future development. |

Greece	The research delves into the construction of national consciousness and the formation of the concept of the nation-state in Greece. It investigates how the idea of the nation is developed and shaped during the 19th century.	The research delves into a catalog of 91 prose fiction novels, corresponding to 90 novelists and an anonymous title. These 91 novels are all written in the Greek language and were published inside and outside of Greece or Greek-speaking territories between 1811 and 1901.	The computational methods used in the study involve a combination of statistical and linguistic analyses, as well as topic modeling. The statistical analyses focus on determining the frequency and occurrence of the concept of the nation, while the linguistic analyses examine the collocations around each keyword related to nations.	The paper provides: 1) Bar charts with the distribution of word counts in the 19th century; 2) Line charts to visualize the frequencies of different keywords related to the concept of the nation over time; 3) Word clouds to showcase the collocations and associations of specific nation-related keywords.	By employing keyword-based retrieval to trace direct references to the concept of the nation, the researchers observed a notable upward trend during the latter decades of the century. The phrasing used to describe the nation showed a preference for self-identification as 'Hellenes' rather than 'Romios' or 'Graecus'. The thematic pattern revealed in the collocation analysis is closely related to the process of state-building, showing elements of romantic nationalism, combining concepts such as justice, idea, freedom, glory, honor, soul, heart, life, hope, belief, and Union.
Hong Kong	The paper investigates the possibility of abstracting and retrieving the figurative meaning of narratives through digital tools, adopting a narratological framework and discussing what Katherine Hayles describes as «the inexplicable, the unspeakable, the ineffable» aspects of narrative literature.	After discussing other research that focuses on Computational Criticism, which involves different corpora, the specific paper solely focuses on Mark Z. Danielewski's *House of Leaves* (2000) as an example of a literary text that challenges our understanding of 'narrative' or 'book' through its incorporation of computerized culture and representation.	The paper discusses the potential utilization of a relational database to map Mark Z. Danielewski's *House of Leaves*, which, as a printed hyper-text, presents at least four possible narrative trajectories. A suitable relational database should represent the layers of meaning within its narrative content, incorporating literary elements such as metaphor and trope. Additionally, it should represent the novel's distinct approaches and occurrences of mediating the same narrative idea, space, or event. Ultimately, it should also illustrate the multiple ways in which these approaches and occurrences overlap and interact with one another.	The paper first provides a description of network visualization devised by another study and then presents their own data through a relational database representation. It also provides an example of XML-tagged text that serves as a foundation for the database.	The paper delves into the intersection of narratives and databases within Computational Literary Studies. It draws a parallel between the layers of meaning found in figurative texts and the management of content in non-relational databases. The primary aim of the paper is to showcase the potential of a recursive feedback loop between the text and the digital, while also recognizing the significance of the uncertain as a literary category. For instance, Danielewski's House of Leaves features a scene in which characters enter a labyrinthine hall: this event is presented in the novel through various mediations like documentary scenes, textual footnotes, and a comic book depiction, and attempting to represent such multi-linearity in a relational database would lead to numerous tables with empty values.

Illinois (US)	The monograph examines the lifespan of genres, conceptualizing genre and acknowledging their evolution, as well as exploring the influence of prestige (defined as cultural prestige, academic recognition, economic success, and other related aspects) and gender. However, it does not delve into the binary division debate due to technical limitations.	The monograph draws from three major archives: HathiTrust, Chicago Text Lab Corpus, and Eighteenth Century Collections Online (ECCO-TCP). Then, the author presents several case studies that leverage a subcorpus each, selected by genre or other literary categories, obtained from the three archives.	The author primarily utilizes logistic regression within the machine learning framework, employing an approach known as 'perspectival modeling.' This approach involves manipulating the dataset, which has been labeled by human annotators, to predict various literary categories (such as genre or author gender) as well as mathematical and linguistic variations in texts or excerpts.	The monograph presents the following: 1) Scatterplots used to investigate the corpora and its literary characteristics. These scatterplots show various aspects, such as the presence of volumes in a library over a span of more than two centuries, the occurrence of certain topics in the corpus, or the time duration (whether a genre narrates a week or several years). They also examine the time pace of different genres; 2) Scatterplots utilized to assess the predictive ability of machine learning models, particularly their capacity to detect different genres; 3) Line charts employed to demonstrate linguistic changes in genres per decade. They also explore prestige-related characteristics, for example, the ratio of bestsellers written by reviewed authors compared to those by random authors, among other literary attributes; 4) Area charts employed to investigate literary characteristics such as gendering of words.	Beyond addressing theoretical critics of distant reading, the author delves into several findings from their corpus. For example, the research reveals that judgments of prestige throughout history, based on reviews, have increasingly encouraged the understanding of literariness as having temporal immediacy and concreteness. In contrast to traditional views of literary history that emphasized abrupt changes, the study demonstrates a gradual accumulation of literary styles, leading to a differentiation of fiction from biographical accounts through a distinct language known as literariness. In another case, the author investigates gender distinctions and discovers that around 1850, a machine learning model could more easily distinguish characters as male or female. However, over time, gender distinctions blurred and multiplied, resulting in a convergence. As a consequence, certain words like 'grinned' and 'smiled' switched gender associations throughout history. Similarly, words like 'eyes' and 'hair' became more associated with femininity, 'pocket' with masculinity, while 'read' remained gender-neutral throughout history.

India	The paper theoretically refers to 'digital sound studies' and focuses on poetic rhythmical prosody to retrieve metrical differences and various types of poetic rhythms in Urdu-Hindi poetry.	The paper does not explicitly mention the specific texts that the authors are working on. However, it provides an overview of their approach and the tools they use for working with Hindi and Urdu digital texts.	The computational part is composed as follows. Firstly, given the lack of a vast publicly available and curated corpus for Indian languages, they explore the use of Optical Character Recognition (OCR) to transcribe printed text into digital form. Secondly, they discuss the utilization of software and tools such as GitHub, Jupyter Book, and Sphinx for collaborative markup, leveraging the framework of minimal computing, to eventually obtain an annotated corpus that could be used as the foundation for poetic analyses.	The paper includes examples and code snippets that provide specific instances and demonstrate the usage of the tools mentioned in the text within the context of creating multilingual and annotated critical texts for Indian languages.	The conclusion they arrive at is that embracing a minimal computing approach using plain text allows authors and researchers to enjoy greater creative and scholarly freedom over their work. Beyond gaining insights into the texts of interest, the reflection also explores the potential of digital technologies for critical editions.
Indonesia	The paper investigates the relationship between characters by studying the undirected co-occurrence network at the 'adegan' (scene) level in Javanese 'shadow puppetry'.	The analyzed corpus comprises works from the Indonesian classic period or Malay literature, ranging from 1900 to 2000. It was identified by defining a subcorpus from the Google Books Ngram database, which consists of over 40 million books. The subcorpus was selected based on twenty Indonesian poetic or prosaic genres, including pantun, epik, and syair, among others.	The researchers gather data by searching for semantic words connected to Indonesian literary genres, encompassing book titles, subtitles, and related topics. Subsequently, they utilize the Google Books 'ngram' viewer to arrange and analyze the data. This viewer enables the researchers to present a comparative graphical model of the genres.	The paper utilizes the Google Ngram viewer to generate line charts with differently colored lines for each literary genre or literary characteristic related to a genre, whose prominence is being investigated.	After conducting a comparative analysis of Indonesian poetry genres, the paper ultimately observes that the genre with the highest number of citations is the mantra genre. However, it is important to note that the term 'mantra' refers not only to the Indonesian tradition but also to the Indian tradition and, due to the polysemic nature of the word, it is excluded from the analysis, resulting in the identification of the most popular poetry genres as Pantun, Epik, and Syair. Within the prose category of Indonesian literature, Hikayat stands out as the most prominent genre.

Iran	Different words used as alternatives to love were classified into twelve categories in order to investigate the concept of love, its semantics, and its symbolic meaning in the poetry of the poet of interest.	The study encompasses all the verses in Saadi's lyric poems (ghazals), which are spread across four books.	The research involves the manual classification of words related to love and their metaphorical representation, computing descriptive statistics for their usage, testing interpretive hypotheses using the chi-square test, and conducting clustering analysis using the K-means algorithm. Technically speaking, the Statistical Package for the Social Sciences (SPSS) is used as the computational tool for data analysis.	The results of the analyses are presented in tables that account for the statistics of words and categories used as alternatives to 'love.' There are separate tables for the results of the Chi-square goodness-of-fit test and the K-Means clustering, which pertain to different categories and words.	The objective of this study was to explore various words that Saadi employed as substitutes for love. The alternatives for the word 'love' were identified and categorized into twelve groups. The findings revealed that in Saadi's lyric poems, the terms 'Fire' and 'Pain,' along with the categories 'Sickness,' 'Human,' and 'Heat,' were the most frequently used to represent alternatives for love.
Ireland	Leveraging a narratological approach, the researchers investigate narrative voice to examine how the author's use of this formal feature affects the content of the novel and its reception.	The novel *The Secret Scripture* (2008) by Irish author Sebastian Barry, with comparative reference to Pat Barker's well-known war novel Regeneration (1997).	The software application consists of developing a tool for analyzing and visualizing XML-encoded data that represents narrative structure, trauma, and cultural context in literary texts. This software allows for the encoding and markup of passages, enabling researchers to annotate and compare literary excerpts. The platform is also web-based, facilitating collaborative processes among researchers.	The data is visually presented through a web-based visualization tool designed to render XML-encoded data. The tool supports comparative analysis by simultaneously rendering XML-encoded narrative for two texts.	The research discovers that in The Secret Scripture, an autodiegetic narrator is presented, with Roseanne, the protagonist, telling her own story. This narrative decision fosters a close bond between the narrator and the reader, likening the reader to a trusted confidant. The examination demonstrates how Roseanne's trauma is delicately expressed through pauses and omissions in her storytelling, underscoring her reluctance to voice her traumatic experiences. This mirrors the historical suppression of women's voices in Irish society from the early to mid-twentieth century.

Israel	The study talks about the ideological and cultural roles, as well as their transformations, played by the image of the Jewish temple in Hebrew fiction.	The literary objects under consideration include early editions of Shakespeare's plays and manuscripts of chapters from the Babylonian Talmud – a 6th-century Hebrew-Aramaic compendium of Jewish law that records oral traditions.	In order to construct a 'synoptic' text for identifying transcription errors and editorial discretion in classical texts, researchers employ computational methods to measure the alignment quality. This involves calculating the distance between aligned word pairs. Two types of distances are considered: orthographic distance and conceptual distance. The first one is computed using Levenshtein distance, which measures the dissimilarity between words based on their spelling. The second one is calculated using word2Vec, a method that represents the semantic meaning of words in a mathematical space and allows for learning their mathematical-conceptual distance. After computing the synoptic distance, these scripts are inserted into a specifically-devised algorithm called 'FAST'. This algorithm splits the texts, aligns the segments, and orders them based on their similarity, as computed in the previous steps. As a final post-processing step, the algorithm examines words that are mostly aligned with gaps and checks if combining them makes sense in the context.	The results of the analyses are presented in the form of a bar chart that compares the results of the alignment algorithms. This allows the paper to compare the alignment quality between different methods. Additionally, textual examples from Hamlet or the Talmud are visually provided and highlighted to showcase specific passages that can be aligned and compared.	The research offers valuable insights into a novel method known as 'Fast and Accurate Synoptic Texts' (FAST), which aims to compare synopses and align parallel lectiones of literary works. Instead of delving into the literary interpretation of the corpus at hand (which includes works from Shakespeare and the Talmud), the paper's conclusion aims to demonstrate the method's efficiency and quality advantages over existing approaches.

| Italy | The group of researchers is investigating the concept of digital social reading, drawing from a variety of theories such as literary studies, sociology of literature, new media studies, etc. Within this framework and the conceptualisation of digital social reading, they are examining ten aspects: community, markets, theory and methods, textual style, source, reading/ non-reading orientation, society, literacy, site type, literature, and institution. More specifically, they investigate reviews and comments on online platforms to find references to possible authorities, changes in sentiment in the comments written throughout the text, along with the progression of the story, or changes in sentiment in reviews, among other things. | The paper presents seven case studies, each focusing on different corpora: 1) The first case study investigates comments written by 300,000 users in the margins of twelve English novels on Wattpad. Wattpad is a platform that hosts millions of stories written in over thirty languages, spanning various genres, including literary classics, fan fiction, and original fiction; 2) The second case study examines comments in the margins of paragraphs on Wattpad. It explores how these comments allow researchers to investigate the progression of readers' responses to a story. The study aims to establish links between the verbalization of aesthetic, cognitive, and emotional reactions and specific text passages; 3) The third case study employs questionnaires to investigate the reading and writing behavior of young people on online literature platforms such as Wattpad, Archive of Our Own, or Fanfiktion; 4) The fourth case study focuses on a corpus of anonymized German user-generated content from an online survey conducted in both German and English. The study primarily targets young users (aged 13 and above; N = 315) to evaluate their responses to the | Researchers used a variety of computational methods. First, they employed network analysis to study the social interactions of readers on Wattpad, focusing on comments written by 300,000 users in relation to twelve English novels. Second, they analyzed the relationship between the text and comments by employing sentiment analysis not only on the novels themselves but also on the dataset of comments left by readers. Third, the researchers applied sentiment analysis to the reviews on the literature-related platform and then performed a statistical analysis to examine the association between the star ratings and the sentiment expressed in the reviews. Fourth, they extracted references to possible authorities, such as traditional critics, newspapers, prizes, television programs, the book trade (publishers, booksellers, libraries), authors, teachers, websites, and private contacts from online reviews using collections of search terms and regular expressions. Fifth, they identified the difference in style between professional critics/journalists' reviews of literature and Digital Social Reading platforms using stylometric analysis performed with machine learning algorithms. They also conducted a word count to extract vocabulary related to criticism. Lastly, they devised an annotation tool to label passages related to story world absorption and then | The data from the different steps of the analysis are presented as follows: 1) A network diagram that represents the connections between texts and categories in the categorization of Digital Social Reading (DSR) research. The edges in the network diagram indicate the connections between different texts and categories; 2) Graphs that show the emotional arcs, which are the results of sentiment analysis, for six classic novels. The emotional arcs depict the sentiment of the story and comments; 3) An association plot is used to depict the relationship between the sentiment of reviews and the ordinal scale ratings, which range from one to five stars; 4) A bar chart illustrates the authorities mentioned in the collection of downloaded online reviews; 5) A visual representation showcasing the importance of features for the classification of book reviews. | The study presents various insights from different case studies. It examines social reading and web-based platforms, finding that as the linguistic and cultural complexity of texts increases, readers tend to interact more. In terms of commenting practices, teenage readers prefer witty characters, emotional conflicts, and familiar cultural references. However, there are only minor differences in practices and motivations across different age groups. Regarding online reviews, users tend to excessively use negative sentiment for ratings below five stars, not just for the lowest ratings. When comparing the style of criticism between non-professionals and professionals, the study reveals that stylistic features are more effective in classifying reviews than content-based features. This suggests that how reviewers express themselves is more distinguishing than the content itself. Finally, concerning the experience of story world absorption, emotional engagement is most frequently used by those reviewing romance novels, while attention is more common among thriller reviewers. |

Japan	The study identifies a set of 37 plot elements related to detective novels (investigation, reasoning, confession, etc.); a set of ten tricks used to conceal the murder (time camouflage, place camouflage, alibi, fake criminals, etc.); nine types of criminal motives (vengeance, love, wealth, work, etc.), and a set of features describing the relationships between criminals and victims (friend, family, partner, rich person, etc.). In total, the study presents sixty-six investigated concepts, where the underlying theory is narratology, and the main references are Propp, Greimas, and Genette.	The literary objects being investigated are 134 renowned Japanese comic stories from the detective series Case Closed.	The computational part takes place after rigorously formalizing a detective novel plot ontology. This ontology is solely constructed based on plot elements such as investigation, reasoning, confession, etc., as well as tricks used in the cases such as disguising the purpose, means, or time. Additionally, motives in the cases are categorized into themes like love, wealth, work, etc., and the relationships between criminals and victims are categorized as witnesses of past cases, friends, family, etc. Once this robust ontology is developed, a factor analysis is conducted to identify correlations among the sixty-six plot parameters.	The data is presented in the form of tables and diagrams. The tables present data related to plot and character parameters, and bold font is used to indicate cells where the strength of the relationship between a parameter and a factor is relevant for the analysis. The diagrams represent plot transition networks, where nodes represent plot elements, and edges represent the transitions between them. The size of the nodes corresponds to the frequency of each plot element, and the connections between nodes indicate frequent transitions between plot elements.	The researchers employed factor analysis on 42 parameters abstracted from the detective stories genre, identifying eleven factors associated with fundamental story patterns. These factors encompass themes such as 'eliminating the witness,' 'vengeance for a past death case,' 'a request for a greedy type of crime,' 'vigilance against theft,' and more. Additionally, by representing plot transitions between various plot elements as networks, they observed changes in the structures of these plot transition networks based on the factors present in the detective stories, thus confirming the significance of these factors.
Kazakhstan	The study explores how existential loneliness is portrayed across different cultures, including its stylometric signature, underlying conceptual metaphors, and the impact of each linguistic metaphor on defining existential loneliness.	The paper tackle a corpus including George Eliot's *Silas Marner*, Emily Brontë's *Wuthering Heights*, Dostoevsky's *Crime and Punishment* and *The Brothers Karamazov*, Guy de Maupassant's *stories*, Thomas Hardy's *Jude the Obscure*, Herman Melville's *Moby Dick*, Jack London's *Martin Eden*, Marquez's *One Hundred years of Solitude*, McCarthy's *Outer Dark*, Pushkin's *Eugene Oneguine*, Steinbeck's *Winter of Our Discontent*, and Wilde's *The Canterville Ghost*.	The project offers a digital workspace where the user can simultaneously visualize manuscripts, digitally annotate, and comment on excerpts.	The visualization techniques used in this study include figures and tables to present the results of the frame-based analysis and linguistic metaphor analysis. The visualization of the data involves mapping corpus-based findings to FrameNet, quantifying the frequency of using specific frames in each literary work, and identifying frame-to-frame relationships.	The paper concludes that loneliness is a recurring and significant theme in literature. By extracting metaphors from a diversified corpus that includes works from various cultures and nationalities (e.g. Cervantes, Dostoevsky, Saken Zhunusov), the researchers were able to demonstrate that this emotion is expressed in a similar way, transcending cultural boundaries.

Kyrgyzstan	The research provides a platform where manuscripts of literary works can be visually annotated. It offers an example of poetic manuscripts, where the annotated passages relate to the codicological information of the literary works.	The paper centres around the Early Modern Kyrgyz Oral-derived Narrative Sources (EMKONS), a digital workspace that provides facsimiles, text transliterations, editions and translations, as well as commentaries on manuscripts of early modern Kyrgyz oral-derived narrative sources.	The computational methods involve user interaction mechanisms for opening documents and images from the sidebar of the digital platform. Additionally, digital annotations and links can be accessed by clicking on highlighted portions.	The platform offers the possibility of multiple manuscript visualizations with colored annotations, where each color carries a different meaning: pink for manuscript annotation, red for editorial note, green for page link, and blue for in-manuscript emendation.	Being a platform for digital literary annotation on digitized manuscripts, the project aims to serve as a foundation for other researchers' analyses, but it does not provide exemplary literary insights into any manuscript. The case study provided is a digital annotation on a singular manuscript.
Latvia	Within the framework of cognitive metaphor theory, the study explores the research of source and target domains in metaphoric tropes. The source domain refers to the figurative concept used to express meaning, while the target domain represents the intended meaning. Specifically, the testing focuses on conceptual metaphors associated with temperature, such as strong emotions are heat, vitality is heat, passion is heat, death is coldness, and loneliness is coldness.	They are processing a corpus of 20th-century Latvian poetry (1920-1999), which includes approximately 400 poetry books.	The computational methods used in this study involve the application of Word2vec and BERT models to identify and analyze metaphorical expressions to investigate the relationships between source domain terms (such as 'fire' and 'ice') and target domain terms.	The data has been visualized traditional graphs to explore the associations between terms from the source domain and concepts in the target domain of Latvian poetry. The aim is to depict the frequency and prominence of both the source and target terms while showcasing their relationships.	The results focus on how frequently the intended meaning of metaphors (target domain) is explicitly expressed, acknowledging that it is often implied rather than stated. Due to the limited effectiveness of topic modeling in poetry, the research adopts Word2vec, which shows better performance in identifying connections between various source domain terms. The most encouraging initial outcomes are obtained by utilizing a pre-trained BERT model that incorporates a corpus of Latvian language texts.

Lebanon	The paper does not provide a single example but discusses, among other trader topics of Digital Humanities, a comment pipeline for Digital Literary Studies that needs to draw from a so-defined 'meta-epistemology.' Therefore, it implies that research needs to concentrate on abstract conceptualizations that are algorithmically comprehensible	The paper does not provide an actual case study but rather discusses the relationship between preservation and digitization as the foundation for utilizing software in literature.	The paper provides a general description of the framework in which Digital Humanities operate. It mentions several software approaches and tools, also relating literary studies within the broader framework of humanities, such as markup and annotation through XML, Information Retrieval (IR) Techniques and their link with ontologies, automatic tagging. When debating the usage of Artificial Intelligence algorithms, it also highlights the difficulties faced by the latter in efficiently performing with the Arabic language due to a lack of training data.	The paper discusses the general pipeline of a Digital Humanities project and highlights the annotation of diverse media, including images and videos. It also explores the technologies that serve as the foundation for the visualization process, such as digitization and database construction.	The paper discusses various topics, including cultural values, traditions, heritage preservation, and their connection to language-specific technologies that require refinement. Additionally, it explores theoretical reflections, such as the impact of the digital culture revolution on human perception of things and relationships, leading to changes in the approaches to studying social and human sciences. However, the primary emphasis lies in integrating technology into education and preserving culture, with a particular focus on the Arabic context, rather than delving into literary interpretation or specific literary examples.
Malaysia	The paper explores and interprets the conceptual metaphors in mystical literature, specifically those related to love.	The dataset comprehends the texts of Divan of Hafiz and Bustan of Sa'adi', two famous Persian poets.	The study employs a combination of clustering analysis and data visualization techniques to extract information about the frequencies of love-related lemmas and their alternatives in the reference literary works. They also apply the chi-square test to assess the statistical significance of their findings.	The paper presents four word clouds. Two of the word clouds are used to display words and categories that serve as alternatives to love in the Divan of Hafiz. The other two are utilized to visualize words and categories used as alternatives to love in the Bustan of Sa'adi.	The research delves into the various interpretations of the concept of 'love' in the works of two Persian poets, Hafiz and Sa'adi. In Divan of Hafiz's poetry, the term 'way' and the category 'human' are commonly employed to express the notion of 'love.' On the other hand, Bustan of Sa'adi frequently utilizes the words 'fire' and 'pain,' along with the categories 'sickness' and 'human,' as alternatives to convey the idea of 'love.' Both poets assert that love serves as a means to attain spiritual perfection and is an inherent trait of humanity, setting humans apart from angels.

Mexico	The researchers investigate Rosalía de Castro's bilingualism. They reflect on the length of the work, the linguistic diversity (assessed through the Type Token Ratio) of individual poetic collections or poems, the most used words per languages (Spanish and Galician), and the frequency of a topic per language.	The poetic work, both in Spanish and Galician, by Rosalía de Castro (1837-1885).	In the initial stage, they undertake an analysis of word frequencies and vocabulary diversity by employing Voyant tools. Secondly, they employ cartographic visualization techniques to visualise the geographical references made within the text.	They provide: 1) Line charts to visualize the most frequent words in the corpus; 2) Bubble charts to compare pronoun usage; 3) Line charts to show the frequency of keywords in contexts or linguistic correlations; 4) Maps to visualize the mentioned toponymies.	The analysis compares Galician and Spanish in Rosalía de Castro's works. It shows that in her Galician poems, she comfortably takes on the role of the poetic self 'eu' as the protagonist. However, in Spanish, she seems more distant and less familiar. Additionally, the analysis highlights the strong religious influence in Spanish culture, evident through the frequent use of words like 'Dios' (God) and 'vida' (life) in her Spanish writings.
Netherlands	The study tackles the conceptualization of conflict as a narrative mechanism, drawing from narratology, conflict studies, and network theory.	They start from a corpus of 170 present-day Dutch novels and then focus on two specific novels: Bart Koubaa's *De Brooklynclub* (2012) and Leon de Winter's *VSV* (2012).	They claim that the analyses they conducted are founded on an extensive set of metadata (including gender, country and place of descent, country and place of residence, age, level of education, profession or daily activities) pertaining to 2,137 characters identified by human annotators. The data collection occurred in multiple phases over a period spanning approximately from 2014 to 2018. They utilized programming pipelines such as BookNLP and verified the accuracy of the automatically annotated information by manually reviewing the passages. Additionally, machine learning algorithms, specifically a multiple linear regression, were employed to predict the composite conflict scores of characters based on their gender, ancestral origin, and education..	The data is presented in a tabular format for analyzing conflicts between two characters. Additionally, a bar chart is used to compare the absolute distribution of social (im)balance across all enemy/friend triads in the corpus (friend-friend-friend, friend-friend-enemy, friend-enemy-enemy, and enemy-enemy-enemy).	The paper highlights conflict as a fundamental element in narrative fiction and its significance in shaping the dynamics of literary works. Among the many examples of this kind of computational analysis, the paper provides observations such as the fact that the nameless protagonist in Bart Koubaa's De Brooklynclub belongs to the lower socio-economic class, working in an abattoir, a restaurant, a bar, and as an elevator operator. In contrast, Mayer, a real-estate giant, represents the upper class, symbolizing a clash between the higher and lower classes. Overall, the paper aims to reflect on how societal tendencies are reflected in literary worlds through the narratological concept of conflict.

New Zealand	In relation to the figure/concept of 'the portal' in Michael Joyce, an important American author and critic of electronic literature, the author attempts to extract the network of characters in a novel called *Disappearance*, where characters, so to speak, dis/ appear. This signifies that, in this instance, the concept of address becomes an idiosyncratic concept associated with the individual author.	The literary objects to be studied are four of Michael Joyce's digital literary works: *The Six Portals*, *Disappearance*, *Remedia*, and *Othermindedness*.	The computational analysis involves identifying and extracting named entities, such as characters, from the text using machine reading techniques. The process consists of the following steps: 1) Preprocessing the text by performing lemmatization and removing stop-words. 2) Employing Spacy for named entity recognition. 3) Identifying characters through a combination of named entity recognition (NER) and manual curation. This approach allows for the social network analysis, which aims to locate co-occurrences of named entities in the text and represent relationships between characters. Finally, the analysis is visualized using node-based graphs in Gephi, where the nodes represent characters, and the lines connecting them depict their co-occurrences.	The visualization technique used is character network analysis, where characters are represented as nodes, and the connections between them are represented by lines. The size of the nodes indicates the prominence of the characters, and the thickness of the lines represents the strength of their relationships. The paper also provides character timeline plots illustrating the overlapping and shifting identities of characters throughout the narrative.	The author primarily highlights that Joyce's protagonists face a challenge in freeing themselves from their current circumstances, even with access to transcendent portals. This struggle reflects their inadequacy as travelers in interworld transit, symbolizing the human condition, where individuals often find themselves grappling with life's challenges, even when presented with opportunities for change or escape. Moreover, Joyce's Disappearance system of characters, with characters shifting identities, adds a level of complexity to representing literary systems through network analysis.

Niger	The study investigates the relationships between text and pictures in digital poetry, focusing on the fact that e-poetry arises from a layered polysemantics resulting from intertextuality.	The study focuses on the textual and pictographic analysis of the visual poem *C'est tout bon!* by Corinne Albaut, as an example of digital poetry.	While computational analysis is not explicitly mentioned, the paper focuses on digital poetry, which inherently involves hypertextuality and digital transmediality. As a result, the critical analysis extends to co-computationally generated objects and becomes computational.	The visual aspect of the study focuses on the subject under consideration, which is digital and hypertextual poetry. The only visualization provided is an example of a digital poem.	The research revolves around exploring the correlations between the analyzed digital poem and other poetic forms, such as 'poésie lettriste,' 'poésie concrète,' and 'poésie sonore,' with a focus on the potential of sound and visual elements. While the title 'C'est tout bon!' ('It's all good!') embodies the spirit of France, representing abundance and emotional happiness, the poem predominantly employs masculine rhymes, which mirror the prevailing patriarchal tendencies in French society. However, one feminine rhyme stands out, indicating a distinct preference among French women for specific dishes.
Perù	The paper extracts geospatial data such as cities, urban planning, buildings like monuments or administrative structures, and transportation infrastructure such as roads, railways, and airports, etc.	They present a case study on Pedro Benvenutto Murrieta's book *Quince Plazuelas, Una Alameda y Un Callejón: Lima en los Años de 1884 a 1887*, a non-fictional literary travel narrative.	The software involves three steps. First, they fine-tune a machine learning model on the Spanish language to perform Named Entity Recognition (NER) on their corpus and identify place names and locations. Then, they georeference the identified locations and use GIS-related software to visualize the data. Finally, they use automatic Part of Speech tagging to identify the linguistic traits around place names.	The results are presented in several ways: 1) A tabular format to present the model performance scores reported by Spacy; 2) A Loss Curves graph displays the model training and validation loss curves; 3) Literary text with highlighted parts to identify locations in orange and people in grey; 4) A tree structure shows the model's dependency parsing and POS tagging results for an example sentence; 5) An Entity Relation Map visualizes the geolocated location entities represented with their current name, original name, and related person entities. It's interesting to note that some of the visual representations refer to the machine learning training process, while others pertain to the reshaping of the literary text.	After analyzing the locations and the characters associated with them in Pedro Benvenutto's work, the paper discovered that Lima has the most significant number of character relationships. This finding confirms Lima's centrality in the book. Additionally, each location was geolocated and visualized on an Entity Relation Map, showcasing its current name, original name, and related characters to emphasize its importance and present-day connections.

Poland	The paper explores the stylometric similarities between James Joyce and Flann O'Brien, demonstrating which works from the latter's body of work are the most Joycean in terms of style	They compare the works of Flann O'Brien (such as *Swim-Two-Birds*, *New Statesman*, and *The Hard Life*) with a corpus that includes a selection of English-language Irish modernists, encompassing Joyce, Beckett, Bowen, Stevens, and others.	The paper utilizes *Stylo*, an R package, to perform stylometry and mathematically identify the linguistic features of Joyce's works. The ultimate goal is to compare the stylistic differences between Joyce and Flann O'Brien computationally.	The data is presented through several figures: 1) A dendrogram representing stylistic similarities between Irish modernist writers; 2) A graph using a tree-like structure to show stylistic clusters; 3) Several graphs showing the results of rolling Delta analysis of Joyce's Ulysses, A Portrait of the Artist as a Young Man, and Dubliners compared to other modernists' works. Another analysis compares certain Flan O'Brien works with famous excerpts from Joyce to test their similarity. The X-axis represents the segments of the test text being compared, while the Y-axis represents the stylistic affinity or similarity between the test text and the other texts being compared; 4) An alternative visualization of the stylistic clusters represented in the dendrogram; 5) Line charts presenting average sentence lengths of literary works to test their similarity.	The analysis highlights Flann O'Brien's writing style similarities to Joyce's works, focusing on specific passages in Ulysses and A Portrait of the Artist as a Young Man. O'Brien's stylistic resemblances are attributed to his use of an archetypal Dublin dialect and his incorporation of parody and pastiche, demonstrating his active awareness of and engagement with the literary tradition and Joyce's influence. By using stylometric analysis, they identify the most similar passages in different literary works and provide interpretations of singular findings.

| Portugal | The research investigates the literary representations of wolves from a temporal and spatial perspective, blending the frameworks of ecocriticism and digital humanities. | The article includes excerpts from various genres of canonical Portuguese literary works, including novels, stories, chronicles, diaries, and poetry. These excerpts have been sourced from the 'Atlas das Paisagens Literárias de Portugal Continental database', which consists of 423 recorded literary works. | The software application is used solely to identify excerpts in which the wolf (either referred to by its common name or a related term) is mentioned. It then assists the researcher in closely analyzing those excerpts, which are manually grouped into twelve categories (Encounters, Animal ethics, Landscape occurrence, persecution, biological knowledge, domestication, fear, uses, association with unease, comparison, folklore, supernatural) based on the context in which the lemmas are used. | The paper provides: 1) A chart representing the proportion of literary works that mention the wolf in each time period; 2) A chart showing the number of wolf literary representations by time period of publication (Tp) and time period of narrative (Tn); 3) A graph presenting the percentage of occurrence of the 12 categories and three groups by which the wolf representations were categorized; 4) A bar chart showing the relative frequency of each category per literary representation in the three time periods of narrative (Tn); 5) Two box plots displaying the mean richness and the mean diversity of the conceptualized wolf categories for each different time period; 6) A scatter plot of the Correspondence Analysis results; 7) Three Portugal GIS-oriented representations illustrating the comparative distribution and frequency of works with wolf references, the distribution per category, in different geographical regions, and the distribution of categories per time period. | The paper explores the portrayal of the literary wolf in Portuguese literature from the late 19th century to the present. It eventually notes a decline in mentions of wolves over time and their association with rural themes. Additionally, it uncovers that ethical considerations towards wolves have been present since the late 19th century, and the geographic distribution of literary representations aligns with the historical range of wolves in Portugal. |

Russia	The paper focuses on the application of Burrows's Delta in literary texts to explore the concept of multiple narratological voices. Specifically, it investigates how a single author can assume different voices with distinct stylistic traits. It appears that texts written by the same author, but adopting different literary identities, can deceive Burrow's Delta.	They present two experiments. The corpus for the first experiment consisted entirely of poems written by Pessoa, published either under his own name or under the names of Ricardo Reis, Alberto Caeiro, or Álvaro de Campos. In the second phase of the experiment, they added texts by other contemporary authors to Pessoa who wrote poetry in Portuguese (Manuel de Arriaga, Machado de Assis, Florbela Espanca, Carlos Drummond de Andrade, António Feijó, Guerra Junqueiro, António Gomes Leal, and António Pereira Nobre).	The researchers employed stylometry with a twofold aim: 1) To analyze the writings of Fernando Pessoa with the goal of determining if there were distinctive stylistic patterns or voices associated with each of his heteronyms (such as Ricardo Reis, Alberto Caeiro, and Álvaro de Campos); 2) To compare Pessoa's writings with those of other authors in order to determine if stylometric analysis could distinguish between Pessoa and his heteronyms. Ultimately, computational tools were employed to visualize the network of stylometric similarities between texts written by Pessoa and his pseudonyms.	The article includes cluster dendrograms that illustrate the stylometric similarity between different texts or authors. Additionally, there are heatmaps that utilize colors to depict the stylometric distances between texts or authors, along with histogram distributions that display the frequency of stylometric distances between texts or authors.	The stylometric analysis consistently confirms that each of Pessoa's heteronyms possesses its own distinctive voice, separate from both other authors in the corpus and Pessoa himself. Additionally, the experiments showed that while Emile Ajar's works, one of Gays' heteronyms, exhibited stylistic autonomy, it was less distinct from Romain Gary's works compared to Pessoa's heteronyms. This suggests that there are varying degrees of stylistic autonomy among different author-pseudonym relationships.

Singapore	After stating that theater performances can be viewed as data, the monograph tackles four dimensions: 1) 'words as data': it measures changes in vocabulary usage (measured by investigating the most common words over time), 2) 'relationships as data': the scholar investigates the actors taking part in a company and their colleagues, 3) 'motion as data': it studies both motion-related words from dramatic texts and analyzes choreographic movements for a visual representation of theatrical texts, and 4) 'location as data': the geographic distribution of theater performances.	The paper investigates dramatic texts (not only playscripts but also theater reviews and program booklets), but also drama-related data, such as the links between fictional characters and the collaborative networks of artists and producers, the movement of performers and objects on stage, and locations (the coordinates of performance events, venues, and touring circuits).	The monograph utilizes a wide variety of computational methods, including: 1) Dimensionality reduction: This technique reduces the dimensions of textual data in order to identify patterns and analyze relationships. It employs methods such as principal component analysis (PCA); 2) Time series analysis: This method visualizes changes in word usage over time, enabling the tracking of trends and interpretation of cultural shifts; 3) Measurement of linguistic differences: Quantitative measures, such as type-token ratio and delta, are employed to compare texts and identify variations; 4) Topic modeling: This technique identifies groups of words that frequently occur together, revealing thematic elements, character patterns, and recurring actions or settings in theatrical texts; 5) Digital platforms for computer-assisted literary analysis: Specifically, the monograph utilizes tools like Voyant and CATMA; 6) Network analysis: This approach investigates both the co-presence of actual actors in the same production and the connections between characters within the dramatic text; 7) Data visualization: Various visual forms, such as graphs, charts, and interactive visualizations, are employed to represent the data; 8) Embodied visual analytics: This includes software tools that analyze the movement of actors and hardware devices, such as Microsoft Kinect, that can record movements; 9)	The monograph provides: 1) Line charts to investigate changes in word usage or to explore the relationship between the roles of characters in plays and the actors/ people involved in the production and cast of a company; 2) Scatterplots to demonstrate the relationships between various components of network-theoretical measures (such as Weighted Degree, Betweenness Centrality, and Eigenvector Centrality) of Javanese and Indian characters. It also utilizes regression lines to compare the average number of performances per month with factors such as population density, land area, and total population per regency; 3) A line chart providing an overview of the changes in video frames. This analysis involves converting a theater video into a series of static images and calculating the difference image for each pair of subsequent images to evaluate the degree of motion; 4) Choropleth maps to illustrate the distribution of performances across the region, among other things; 5) Series plots to display the total number of performances per month for the years 2016, 2017, and 2018.	The monograph presents an analysis of three aspects: words, relationships, and motion as data. It brings forth numerous key points, and among the ones provided, there are several examples: 1) Vocabulary changes in reviews by The Flying Inkpot, a prominent Singapore-based theater critic group, indicate reduced reliance on the term 'audience,' suggesting increased confidence among critics in expressing their views without relying on audience-centered language; 2) Network analysis shows that while nearly half of the characters are Javanese, Indian-derived characters have stronger connections within the co-occurrence network, implying they appear more frequently. The chapter also delves into colonial-era Dutch scholarship and postcolonial concerns, shaping the interaction between Javanese and Indian characters; 3) Data analysis of Wayang kulit (a traditional puppet-shadow play) performances across regions reveals varying patterns, challenging common assumptions about the distribution of cultural practices and emphasizing the need for a more nuanced understanding of regional representation.

South Africa	The research focuses on the relationships between characters, which are identified through the co-occurrence of their names in sentences.	The investigated objects are one Afrikaans novel and two scanned dramas, one in Tshivenḓa and one in Xitsonga.	Computationally speaking, the study initially utilizes the NCHLT Optical Character Recognition (OCR) system for South African languages to extract text from the Tshivenḓa and Xitsonga scans. Then, they employ the CTexTools2 NER system, which performs Named Entity Recognition (NER) on South African languages. Finally, they attempt - and discuss the challenges of doing so for underrepresented languages - to identify personal relationships through character name co-occurrence in sentences and visualize these relationships using Gephi.	The paper presents several network visualizations of name co-occurrence at the sentence level using Gephi. The thickness of the lines represents the frequency with which two names (edges) occur together.	According to this study, the various methods of addressing individuals in Tshivenḓa plays create challenges for Named Entity Recognition (NER), as it becomes harder to track characters and their connections, especially in concise texts where multiple characters do not appear within a single sentence. Although Afrikaans novels have more sentences with multiple characters, Tshivenḓa and Xitsonga plays exhibit fewer such instances. To enhance the comprehension of relationships in plays, the study suggests including the speaker as a character in every sentence.
South Korea	The concept being analyzed is the relative frequency of words automatically tagged as nouns and adjectives in English and American poetry. This analysis is conducted to investigate the Vocabulary Density, which is the ratio of the total number of words in the document to the number of unique words in the document.	The paper analyzes 80 poems pieces from forty-seven poets available on Project Gutenberg, a catalog spanning from the eighteenth century to the early twentieth century.	The authors employed several computational methods. First, they downloaded and preprocessed the text data of the poem pieces. This involved tokenization, lowercasing, and stop-word removal. The researchers then calculated the Vocabulary Density and performed Part-of-Speech (PoS) Tagging using the Apache OpenNLP Maxent tokenizer (Hornik 2016) from the 'openNLP' package. These processes provided insights into the usage of different word categories in the poems and how they varied over time.	The paper presents a table displaying the results of the vocabulary density analysis, which compares T.S. Eliot's poetry to that of other writers. Additionally, the same data is visualized through colorless scatterplots, where points are grouped and labeled to emphasize the similarities and differences between modern and romantic poetry.	Through conducting computational research on vocabulary density, the analysis uncovers that Eliot's poetry exhibits the highest density compared to the other modernist poets studied. He achieves this by incorporating neologisms, enabling him to convey pioneering ideas. One notable example is Eliot's use of allusions and fragments in The Waste Land, a linguistic pattern which represents a symbol of the fragmented and disillusioned state of post-World War I society. Additionally, the study confirms that Eliot's vocabulary richness and innovative language departed from Romanticism, where concrete nouns were employed to depict objective reality.

Spain	The paper reports on different methods that have been tried to model computational storytelling, and then assesses creativity. Each of these methods captures creativity in a different way. In general, they attempt to reflect on chronology, focalization, causality, and the distinction between fabula and discourse. More specifically, some of them address the relationships between characters (competition, dominance, familiarity, affection, trust, deceit, and indebtedness), event chains (backward chaining, forward chaining), character goals, and other plot dimensions.	Since the paper investigates the value of computer-generated text, the literary objects being investigated are the stories generated by the software each time. In the discussed examples of softwares generating stories (or providing writers with a starting point for a story), the genres range from stories about King Arthur and his knights of the round table to stories that model the generation of scripts for a series of TV soap opera episodes. In fact, it is not the genre of the story that is being addressed, but rather the generative process of these stories.	The article discusses various computational methods used in storytelling systems. These methods include using probabilistic rules for literary character behavior and event generation, reasoning from events to consequences and vice versa, reformulating author goals throughout story development, utilizing plot fragments and planning for plot generation, and employing autonomous intelligent agents. Generally speaking, they present an ontology of how a storyline is built and aim to computationally generate a chain of events.	Given that this paper could be placed at the edge of Computational Literary Criticism, the only visualization provided is a visually represented narrative plan. This plan outlines the sequence of events and actions that take place in a story. One of the described software programs is built upon this ontology to generate literary stories.	The examination of existing storytelling systems provides insights for literary studies regarding the significance of distinguishing between story invention and story narration. It is essential to study these aspects separately to comprehend their unique demands and interactions. Furthermore, the research emphasizes the often neglected role of the audience as active participants in the storytelling process, which should be given more consideration, aligning with modern literary theory that emphasizes the reader's importance.
Sweden	The study focuses on geographical locations found in older literary texts.	Although the article does not specifically mention the literary works under investigation, it does refer to their utilization of the Spraakbanken, the Swedish Language Bank, which includes a vast corpus that encompasses literary works as well. Additionally, the article provides two specific examples of texts: *The Wonderful Adventures of Nils* by Selma Lagerlöf and *Det går an* by Carl Jonas Love Almqvist.	The article discusses computation methods that center around utilizing digital maps, geographical information systems (GIS), and diverse language processing tools. These methods aim to extract place-related information using Named Entity Recognition (NER) tools and enhance accuracy by incorporating a place name lexicon specific to certain time periods.	The only visualization provided is a static map of a selected set of places mentioned in the work Det går an.	Since the article describes the development of a geographical information system as a part of a visualization solution for an archive of historical Swedish literary texts, it delves into the technicalities of enhancing NER algorithms and correcting typing errors. However, it is expected that the GIS system will also serve as a foundational point for potential literary interpretations that are not directly provided in the paper.

Switzerland	The group of researchers investigated many different concepts, grouping them into three levels: 'space,' 'doubt,' and 'form.' For each of these levels, they provided a conceptual itinerary that includes three substeps: phenomenon (introducing the most concrete and tangible aspect of the itinerary), process (attempting to grasp the dynamic behind each phenomenon), and problem (examining the main issue leading to each phenomenon, thus initiating the process). For any concept that they present in relation to Calvino's work, the project offers a brief theoretical explanation to help the reader/user understand.	The corpus includes all the narrative works by Calvino: over two hundred texts written and published between 1943 and 1985, which are divided, in the project, according to chronological and editorial criteria.	The software application in this project primarily focuses on information design. Instead of concentrating on information retrieval through Natural Language Processing, the project utilizes the expertise of classically trained literary theorists to visualize the data using a specially designed software. However, to assist the humanists in obtaining the relevant information for visualization, the technical team from Politecnico di Milano's *DensityLab* provided the annotators with two software tools – *Explorer* and *Wanderer*. *Explorer* allows the highlighting of specific parts of the text, provides the position of the highlighted text, allows the specification of properties based on a data schema defined by the researcher, and enables the export of this information in tabular format. On the other hand, *Wanderer* enables the simultaneous selection of two text portions and the highlighting of connections between them.	The work offers a web platform created as part of the Atlante Calvino project, which combines literary analysis with information design in several interactive ways that cannot be fully explained here. Among the numerous visualizations, the main one represents Calvino's literary corpus as a territory, where each text is depicted by a graphic element, and the individual text is the smallest unit of visualization. The size of each graphic element is proportional to the length of the text it represents. The visualization employs color to introduce two different paths for exploration. In 'Chronology mode', the chromatic scale ranging from green to purple represents the chronological order of first publication for each text. In 'Volumes mode', the color represents the volumes published by the author, including collections of short stories and other types of texts. The project is particularly creative and complex, benefiting from the expertise of the team members from the Politecnico of Milano's DensityLab, which specializes in information design.	The project aims to offer insights into Italo Calvino's literary works, revealing patterns, trends and themes in his writing style. It leaves the door open for personal interpretations while also providing each visualization with a 'reading tips' section, where researchers give their personal interpretations of the visualizations. For instance, thanks to the visual component of the project, it becomes evident that terrestrial locations consistently appear throughout Calvino's works, challenging the notion that Calvino's interest in realism was confined to the early stages of his career.

Turkey	The researchers adopted a top-down corpus-driven approach that moved from key semantic domains to lexical items and then to their collocates. First, they identified the key semantic domains of the genre using a corpus tool. In their corpus, 'Anatomy and physiology,' 'religion and the supernatural,' and 'senses' were the key semantic domains. They then focused on addressing the unusual abundance of 'binary oppositions.' Secondly, they analyzed these key semantic domains to highlight the frequently occurring lexical items within each domain. They also explored the surrounding context of those items to investigate stylistic features.	They delve into the poetry of the Great War by compiling a corpus of 487 poems from 110 poets, all of whom had firsthand experience of the Great War, either as combatant soldiers or as personnel in the army services. The majority of the poems in the corpus were taken from four anthologies of poetry: *Soldier Poets: Songs of the Fighting Men*, *More Songs by the Fighting Men*, *Minds at War: The Poetry and Experience of the First World War*, and *Poetry of the First World War*.	The authors employ computational methods with the aim of analyzing key semantic domains, examining key lexical items within those domains, exploring collocational patterns, and determining the strength of associations between words. They mainly leverage Wmatrix, Lancsbox, and make use of a few linguistic scores such as Mutual Information (MI).	The paper presents collocational networks, with the central point (in red) representing the reference word, and the other points (in grey) representing the words that occur most frequently nearby, from a linguistic perspective. Additionally, collocations, key semantic domains, and the frequency of occurrence of certain words are visualized in tabular form.	The analysis uncovers the presence of binary oppositions in the poetry, such as light and darkness, life and death, sadness and happiness, calmness and fear/shock, and heat and cold. These oppositions emphasize the emotional turmoil endured during the war. Additionally, there is a profusion of key semantic domains related to the senses, including hearing, taste, sight, smell, and touch. These sensory references enable readers to sympathize with the physical and emotional impact of the war. Other findings include the abundance of metaphysical reflections questioning death and life. It was discovered that words like 'heart' and 'eyes' are frequently used metaphorically to convey emotions, while words associated with body parts are used literally to depict the physical realities of war.
United Kingdom	The study revolves around the concepts of 'personification' and 'metamorphosis'.	The paper applies its theory to Ovid's *Metamorphoses*.	The software being used involves a relational model and a database to record instances of personification. The software allows for the correlation of different occurrences of the same entity (personification) and assigns an ontological type to each occurrence. The ontology, which is an important part of the model, expresses how personification works on the microscopic level.	The author mentions the use of an Excel spreadsheet and charts to visualize the results generated from the database. The charts display the cumulative result of effects derived from different linguistic factors, such as context, local factors, verbs, and participles. It also uses different colors to represent different types of factors, allowing for a visual understanding of the personification phenomenon.	By formalizing personification in Metamorphoses through a relational database, the authors identified five key classes of linguistic factors contributing to that phenomenon: context, local factors, subject verbs, object verbs, and participles. The analysis, rather than solely focusing on Metamorphoses, emphasizes that personification in literature is a complex and non-linear phenomenon, whose nuances have to be understood and explored with a modeling approach, while also relying on human judgment.

| United Arab Emirates | The concept under examination is 'geographic information,' which the author contends refers to textual mentions of various types of place names across different levels of detail within continuous prose or poetic narratives. These encompass landmarks, settlements, regions, and countries, both real and fictional. | Broadly speaking, the paper focuses on a corpus of over 250 texts from various periods and genres from the Medieval French tradition, spanning approximately 1250 to 1500, known as the. The corpus includes works such as the oeuvre of Christine de Pizan, *the Roman de Rou*, Joinville's *Vie de saint Louis*, and Gerbert de Montreuil's *Roman de la Violette*. | The computational aspect of the project involves extracting spatial data from texts using not only a gazetteer approach but also manually curated spatial data from medieval French literary and non-literary texts. Subsequently, this information is stored in a database, facilitating comparative analysis of texts utilizing spatial data. Finally, a Geographic Information System (GIS) tool is employed to visualize the data. | The author utilizes various tools and techniques to present geographical information extracted from medieval French texts. These include: 1) Creating maps using Google Fusion Tables, ArcGIS, or Google Earth to display the specific locations mentioned in individual works; 2) Generating maps that showcase the top fifty locations mentioned in a collection of medieval French texts, sized according to their frequency. These maps can be visualized using CARTO or QGIS, depending on the type of geographical data in the corpus; 3) Employing a choropleth map to illustrate the geographies over time. This map can be visualized using CARTO; 4) Aggregating the Visualizing Medieval Places (VMP) data and incorporating a time slider to enable customized temporal views; 5) Overlaying the Visualizing Medieval Places (VMP) project data onto a base layer of major European cities around the year 1200. This visualization can be accomplished using QGIS; 6) Providing a comparative heat map view that juxtaposes the VMP data of places mentioned in medieval French texts with the data from the Pelagios Latin geographic tradition. | In general, the paper focuses on the use of GIS system to map out the locations of literary works to discover new insights. For instance, when examining the reference corpus of Medieval French Literature, one discovery is the infrequent mention of Europe east of the Rhine and the limited number of texts that reference it. Another discovery concerns the difference between the thirteenth-century original verse Roman de la Violette and its fifteenth-century prose rewriting. They vary in terms of the geographic distribution of the settings, demonstrates how the localization of an older narrative changes under new regimes of power. |

Appendix 4 – Technical Glossary

Autonomous agents – Autonomous Intelligent Agents are computational entities or characters within a storytelling system that possess the ability to make decisions and take actions based on their internal states, goals, emotions, and perception of the environment. These agents operate autonomously, meaning they have a degree of computational independence and can act without direct control or intervention from a human or central authority.

BERT – Bidirectional Encoder Representations from Transformers (BERT) is a powerful language model capable of comprehending and generating contextualized word representations. This ability allows it to retrieve sentences that are conceptually and linguistically similar to the input sentences used during training (Devlin 2019). In this particular case, they utilize this machine learning algorithm by providing it with labeled excerpts as 'scenes', enabling the retrieval of new excerpts that are similar to the provided ones.

Biber's model – Biber's model is a linguistic model that consists of sixty-seven linguistic features used to analyze texts (Biber 1995).

Black Box (machine learning) – In machine learning, a 'black box' refers to a model or system that provides predictions or outputs without revealing its internal workings. It operates as an opaque entity, making it difficult to understand how it arrives at its conclusions. Understanding black box models is a topic of ongoing research to improve transparency and trust in Artificial Intelligence systems (Loyola-González 2019).

BookNLP – BookNLP is a natural language processing pipeline that scales to books and other long documents (in English), including: Part-of-speech tagging; Dependency parsing; Named entity recognition; Character name clustering (e.g., 'Tom', 'Tom Sawyer', 'Mr. Sawyer', 'Thomas Sawyer'); Quotation speaker identification; Pronominal coreference resolution; Supersense tagging (e.g., 'animal', 'artifact', 'body', 'cognition', etc.) (Bamman, Underwood 2014).

Brown Corpus – The Brown University Standard Corpus of Present-Day American English (or simply the Brown Corpus) is an electronic collection of text samples of American English. It was released in 1967 as the first significant structured corpus encompassing various genres. Compiled at Brown University in Rhode Island, it is a comprehensive language corpus comprising 500 English samples, amounting to approximately one million words.

Chi-squared (statistics) – A chi-squared test is a statistical hypothesis test used to examine whether two categorical

variables are independent in influencing the test statistic.

Chicago Text Lab – The Chicago Text Lab corpus is a substantial collection of contemporary American fiction, covering the period from 1880 to 2000. It comprises nearly 9,000 novels that were selected based on their number of library holdings as recorded in WorldCat.

Cosine similarity – Cosine similarity is a mathematical measure used to determine the similarity between two vectors, often representing textual or semantic features of documents, sentences, or words.

Correspondence Analysis – Correspondence Analysis is a statistical technique used to analyze and interpret data organized into categories. It is frequently used to explore relationships between categories as geometric patterns in a low-dimensional space.

Decision trees (machine learning) – Machine learning decision trees (DTs) are predictive non-parametric supervised learning method used for classification and regression, that mimics the structure of a tree. They make decisions by dividing data into smaller subsets based on specific features or attributes. Each division creates a 'branch' in the tree, leading to a final prediction or outcome at the 'leaves.'

Eighteenth Century Collections Online (ECCO) – The Eighteenth Century Collections Online includes English-language and foreign-language titles printed in the United Kingdom during the 18th century, along with thousands of significant works from the Americas. The database comprises over 32 million pages of text and more than 205,000 individual volumes in total.

Factor analysis – Factor analysis is a statistical method used to identify underlying patterns (factors) among observed variables in a dataset. It helps to simplify complex data and understand the relationships between variables by grouping them into common factors that explain the variance and structure in the data.

Fine-tuning (machine learning) – Fine-tuning, in the context of machine learning, refers to the process of taking a pre-trained model and further training it on a specific task or dataset. Instead of starting the training process from scratch, which can be time-consuming and resource-intensive, fine-tuning allows to leverage the knowledge and learned patterns from the pre-trained model.

FrameNet – The FrameNet project is building a lexical database of English that is both human- and machine-readable, based on annotating examples of how words are used in actual texts. From the student's point of view, it is a dictionary of more than 13,000 word senses, most of them with annotated examples that show the meaning and usage. See: https://framenet.icsi.berkeley.edu/about

Force-directed algorithms – Force-directed algorithms are among the most flexible methods for calculating layouts of simple undirected graphs. Also known as spring embedders, such algorithms calculate the layout of a graph using only information contained within the structure of the graph itself, rather than relying on domain-specific knowledge (Kobourov 2012).

Gazetteer – A gazetteer is a book or part of a book that contains a list of names of places, usually accompanied by additional information. It is often used as a reference to retrieve place names from other texts, employing the so-called 'gazetteer approach'.

Gephi – Gephi is an open-source visualization and exploration software for various types of graphs and networks, widely used within the framework of Digital Humanities.

Graph clustering – Graph clustering, also known as community detection, is a technique used to identify groups or clusters of nodes within a graph. In the context of analyzing verses in a corpus of metric poetry, graph clustering can be employed to find clusters of equivalent verses.

HathiTrust Digital Library – HathiTrust Digital Library is a vast collaborative repository of digital content from research libraries, which includes material digitized through initiatives such as Google Books and the Internet Archive, as well as content digitized by individual libraries. It encompasses approximately 17 million books.

Information retrieval – Information retrieval refers to the process of obtaining relevant information from a vast collection of data or documents. It encompasses techniques and algorithms that empower users to search, retrieve, and access information according to their specific needs or queries, often utilizing search engines or database systems.

K-means (statistics) – The K-means algorithm is an iterative process aiming to divide the dataset into K predefined distinct and non-overlapping subgroups (clusters), with each data point assigned to a single group.

Lemmatisation – Lemmatisation in linguistics is the process of grouping together the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma, or dictionary form. In computational linguistics, lemmatisation is the algorithmic process of determining the lemma of a word based on its intended meaning.

Leverstein distance – In information theory, linguistics, and computer science, the Levenshtein distance is a string metric for measuring the difference between two sequences (Levenshtein 1966).

Logistic regression – Logistic regression is a statistical algorithm used for binary classification tasks. It predicts the probability of an event occurring by fitting a logistic function, which is a mathematical formula that maps input values to a range between 0 and 1, to the input data. In the case of the study at hand, the author provides the model with literary works labeled by literary genre, date of publication, or other literary categories, in order to predict the same category for other works or assess the variance of a new literary work compared to the input work.

Multidimensional Analysis Tagger – The Multidimensional Analysis Tagger (MAT) is a program that replicates Biber's (1988) Variation across Speech and Writing tagger for the multidimensional functional analysis of English texts, generally applied for studies on text type or genre variation.

Multiple linear regression (machine learning) – Multiple linear regression is a statistical method that helps us understand and analyze relationships between multiple variables. In the given case, if we are trying to understand the factors that contribute to someone's conflictual centrality, we can examine whether gender, societal status, or other variables influence the relationship between a character and their propensity for conflict.

Named Entity Recognition – Named Entity Recognition (NER) is a process used in natural language processing (NLP) to automatically identify and classify specific named entities in texts. 'Named entities' refer to real-world objects such as people, organizations, locations, dates, and other important entities.

Natural language processing (NLP) – Natural language processing is an interdisciplinary subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.

Network analysis – Network analysis is a method of studying the relationships between entities in a network. It involves analyzing the connections, or links, between the entities, as well as the characteristics of the entities themselves.

Ontology (computer science) – In computer science, ontology refers to a formal representation of knowledge that defines the relationships and concepts within a specific domain. It provides a structured way to organize information and facilitates the understanding and processing of data by both humans and machines.

Optical Character recognition – Optical Character Recognition (OCR) is a technology that converts scanned or photographed text into machine-readable

characters, allowing computers to recognize and interpret the text, enabling easier data retrieval and manipulation.

Parsing – 'Parsing' is the process of analyzing a string of symbols, either in natural language, computer languages or data structures, conforming to the rules of a formal grammar. In the context of natural language, it involves labeling words with their respective grammatical categories.

Part of speech tagging – Part of speech tagging (POS tagging) is a process in natural language processing (NLP) where each word in a sentence is assigned a specific grammatical category or 'part of speech' based on its usage and context. These categories include nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, and more.

Pelagios – The Pelagios Network is an initiative that establishes connections between online information by employing shared references to specific locations and time periods. Comprised of researchers, scientists, and curators, the Pelagios Network utilizes Linked Data methodologies and tools to explore historical contexts and can be used as a source of open-source geographical information. See: https://pelagios.org

Regular expressions – Regular expressions (RE) are a powerful tool in computer science that can also be useful for literary theorists. In simple terms, a regular expression is a sequence of characters that forms a search pattern. It is a way to describe and match patterns within text.

Relational database – A relational database is a (most commonly digital) database based on the relational model of data, as proposed by Codd (E. F. Codd. 1970). It can be likened to a library catalog, where books are organized on shelves and labeled with categories and subcategories. Each book (data) is stored in a table (shelf) with columns (labels) and rows (individual entries). This structure allows for clear connections and analysis between related information, simplifying the understanding of interrelationships within complex data structures.

Relational model – The relational model (RM) is an approach to managing data that involves structuring and organizing data in a way that is compatible with a relational database.

Sentiment analysis – Sentiment analysis is a commonly used technique in natural language processing (NLP). It involves determining the emotional valence of a piece of text, whether it is positive, negative, or neutral, by analyzing the occurrence of emotionally-related words within a given context, to a certain degree of complexity.

Stop words (Computer Science) – Stop words are the words in a stop list that are filtered out before or after the processing of natural language data (text) because

they are insignificant. Examples of stop words are 'the,' 'is,' and 'and.'

Stylometry – Stylometry is a branch of literary theory and computational linguistics that focuses on the quantitative analysis of literary style. It involves applying statistical and computational methods to study linguistic features and patterns within texts, with the aim of identifying and distinguishing different authors or analyzing the evolution of an author's writing style over time. While the concept is straightforward, stylometry becomes more advanced when considering its mathematical background. Unlike traditional approaches to style, computational stylistics recognizes the equal importance of both function words (e.g., 'the,' 'of,' 'in') and meaningful content words in measuring fundamental stylistic affinities.

Supervised learning (machine learning) – Supervised learning is a machine learning approach that utilizes labeled datasets to train models and make predictions or classifications. In supervised learning, the dataset consists of input data (in this case, excerpts that can be labeled as 'scenes') and corresponding output data (labels for excerpts that enable scene identification). The goal is to learn a function that can accurately predict the output for new, unseen narratological scenes.

TextTiling – TextTiling is a technique for subdividing texts into multi-paragraph units that represent passages, or subtopics

(Hearst 1997). The main idea behind TextTiling is to identify topic shifts or changes in a document by analyzing patterns in textual features such as vocabulary and syntax.

Topic modeling – In statistics and natural language processing (NLP), a topic model is a statistical model used to discover the abstract 'topics' present in a collection of documents. Topic modeling is a widely employed text-mining tool for uncovering hidden semantic structures within a body of text. In simpler terms, it is a method for identifying the simultaneous occurrence of lemmas in a cluster, allowing that cluster to be labeled as a topic. For instance, if the lemmas 'love,' 'desire,' 'passion,' and 'sex' co-occur in a paragraph, such a cluster could be defined as 'passionate love.'

TopicTiling – TopicTiling is an unsupervised segmentation technique used to divide a text into coherent segments based on the underlying topics or themes present in the text. It utilizes topic modeling, specifically Latent Dirichlet Allocation (LDA), to identify the topics discussed in the text and then applies a tiling algorithm to segment the text based on these topics.

Unsupervised learning (machine learning) – 'Unsupervised learning', also called 'unsupervised machine learning', involves using algorithms to examine and group unlabeled datasets. These algorithms can identify hidden patterns and clusters within the data without

relying on human guidance or labels. It allows for the discovery of meaningful insights and structures in data autonomously. In other words, the algorithm explores the data on its own without any guidance about what it should be looking for.

Voyant tools – Voyant Tools is a web-based reading and analysis environment for digital texts. https://voyant-tools.org

Wmatrix (software) – Wmatrix is a software tool designed for corpus analysis and comparison. It offers a user-friendly web interface that incorporates various natural language (NLP) processing tools. It includes a multilingual semantic tagger and incorporates standard corpus linguistic techniques such as frequency lists and concordances. Furthermore, it expands upon the keywords approach by incorporating key grammatical categories and essential semantic domains.

Word2Vec – Word2Vec is a popular algorithm used for natural language processing tasks, specifically for transforming words into numerical vectors. It aims to capture the semantic meaning of words based on their context in a large corpus of text. The main idea is that words appearing in similar contexts tend to have similar meanings. The algorithm learns these relationships by analyzing a given text dataset and creating vector representations, or embeddings, for each word (Mikolov 2013).

XML-encoded data – Extensible Markup Language (XML) is a markup language and file format for storing, transmitting, and reconstructing arbitrary data. XML-encoded data means data, in the cases at hand natural language, marked-up with XML.

Bibliography

Reference edition for the analyzed novels

[1]   Bateman, C. (1995). *Divorcing Jack.* Harper Collins.

[2]   Bateman, C. (1996). *Of wee sweetie mice and men.* HarperCollins.

[3]   Bateman, C. (1998). *Maid of the Mist.* HarperCollins.

[4]   Berry, F. (2021). *Northern Spy*. Penguin Publishing Group.

[5]   Bradby, T. (1998). *Shadow Dancer.* Bantam Press.

[6]   Breslin, J. (2012). *World Without End, Amen: A Novel.* Open Road Media.

[7]   Burns, A. (2018). *Milkman*. Faber & Faber.

[8]   Clancy, T. (1988). *Patriot Games.* Penguin Publishing Group.

[9]   Clarke, S. (2013). *Soldier E: Sniper Fire in Belfast.*   Bloomsbury Publishing.

[10]   Davies, M. (1999). *The Drumbeat of Jimmy Sands*. HarperCollins

[11]   Dickinson, P. (2015). *The Green Gene: A Crime Novel.* Open Road Media.

[12]   Doyle, R. (1999). *A star called Henry.* Jonathan Cape.

[13]   Easterman, D. (2014). *Day of Wrath*. Brown Book Group.

[14]   Edwards, R. D. (2001). *The Anglo-Irish Murders.* HarperCollins.

[15]   Healy, D. (1994). *A goat's song.* Harvill.

[16]   Herron, S. (2014). *The Whore-Mother.* M. Evans.

[17]   Higgins, J. (1987). *A Prayer for the Dying.* Signet.

[18]   Higgins, J. (1994). *On dangerous ground.* Michael Joseph.

[19]   Higgins, J. (1996). *Angel of Death.* Penguin.

[20]   Higgins, J. (1996). *Drink with the Devil.* G.P. Putnam's.

[21]   Higgins, J. (2000). *Day of Reckoning.* HarperCollins.

[22]   Higgins, J. (2005). *Without Mercy.* HarperCollins Publishers Limited.

[23]   Higgins, J. (2008). *The Violent Enemy.*   Harper Collins Publishers.

[24]   Hutson, S. (1997). *Knife Edge.* Brown Book Group Limited.

[25]   Lingard, J. (2003). *Across the Barricades: A Kevin and Sadie Story.* Penguin Random House Children's UK.

[26]   Lingard, J. (2003). *The Twelfth Day of July: A Kevin and Sadie Story.* Penguin Random House Children's UK.

[27]   Lingard, J. (2017). *Into Exile.* Penguin Random House Children's UK.

[28]   Madden, D. (1996). *One by one in the darkness.* Faber & Faber.

[29]   McCabe, P. (1998). *Breakfast on Pluto.* Londra: Pan Macmillan.

[30]   McDonald, I. (1996). *Sacrifice of Fools.* Gollancz.

[31]   McLaim R.W.. (1996). *Eureka Street.* .Minerva.

[32]   McNamee, E. (1995). *Resurrection Man: A Novel.* Picador.

[33]   O'Brien, E. (1994). *House of splendid isolation.* Farrar Straus Giroux.

[34]   Park, D. (1994). *The rye man.* Jonathan Cape.

[35]   Park, D. (1996). *Stone kingdoms.* Phoenix House.

[36]   Park, D. (2015). *Stone Kingdoms.* Bloomsbury.

[37]   Patterson, G. (1999). *The International.* Anchor.

[38]   Petit, C. (1996). *The Psalm Killer.* Macmillan.

[39]   Silva, D. (1999). *The Marching Season: A Novel.* Random House.

[40]   Strong, T. (1994). *The Tick Tock Man.* Heinemann.

[41]   Weber, K. (1999). *The Music Lesson.* Phoenix House.

List of works for the testimonies of The Troubles

[42]   Smyth, M., & Fay, M. T. (Eds.). (2000). *Personal accounts from Northern Ireland's Troubles: Public conflict, private loss.* Pluto Press.

[43]   McKittrick, D. (2001). *Lost Lives: The Stories of the Men, Women and Children who Died as a Result of the Northern Ireland Troubles.* Mainstream.

[44]   White, S. (2011). *Before the Bandits.* Lagan Press.

[45]   Grimaldi, F., North, S. (1998). *Blood in the Street*. Guildhall Press.

[46]   Myers, K. (2006). Watching the Door: A Memoir, 1971-1978. Lilliput Press.

[47]   Spencer, G. (2005). *Omagh: voices of loss*. Appletree Press.

[48]   Rolston, B., & Gilmartin, M. (2000). *Unfinished business: State killings and the quest for truth*. Irish Books & Media.

[49]   Restorick, R. (2000). *Death of a soldier: a mother's search for peace in Northern Ireland*. Blackstaff Press.

[50]   Pringle, P., & Jacobson, P. (2000). *Those are Real Bullets, Aren't They?: Bloody Sunday, Derry, 30 January 1972*. Fourth Estate.

[51]   O'Connor, F. (1993). *In search of a state: Catholics in Northern Ireland*. Blackstaff Press.

[52]   MacKay, S. (2000). *Northern Protestants: an unsettled people*. Blackstaff Press.

[53]   McClean, R. (1983). *The Road to Bloody Sunday*. Ward River Press.

[54]   McCann, E. (1974). *War and an Irish Town*. Penguin.

[55]   Lindsay J. (1998). *Brits Speak Out: British Soldiers' Impressions of the Northern Ireland Conflict*. Dufour Editions.

[56]   Kerr, A., Carlin, D., Hippsley, P. (1996) *Perceptions: Cultures in Conflict*. Guildhall.

[57]   Kennally, D., & Preston, E. (1971). *Belfast, August 1971: A case to be answered*. Independent Labour Party.

[58]   Healy, E., Harris, H. (eds.)(2001). *Strong about it All--: Rural and Urban Women's Experiences of the Security Forces in Northern Ireland*. North West Women's / Human Rights Project Publications.

[59]   McAliskey, B. D. (1969). *The Price of My Soul*. Knopf.

[60]   Daly, E. (2000). *Mister, are You a Priest?* Four Courts Press.

List of works for compiling the journalistic corpus of The Shankill Butchers

[61]   Dillon, M. (1999). The Shankill Butchers: the real story of cold-blooded mass murder. Psychology Press.

[62]   BBC News (2004). "Murder link to Shankill Butchers", in *BBC News*, Thursday 4 November 2004, http://news.bbc.co.uk/2/hi/uk_news/ northern_ireland/3983053.stm [Last checked 16 dec 2023]

[63]   BBC news (2006). "Loyalist bar killing 'horrific'", in *BBC news*, 1 August 2006. http://news.bbc.co.uk/2/hi/uk_news/northern_ireland/ 5233840.stm [Last checked 16 dec 2023]

[64]   BBC news (2011). "Shankill Butchers held Belfast in grip of terror", in *BBC news*, 28 March 2011. https://www.bbc.com/news/uk-northern-ireland-12858930 [Last checked 16 dec 2023]

[65]   BBC news (2020). "Winston Rea 'criticised Shankill Butchers' in Boston tapes", in *BBC news*, 26 November 2020. https://www.bbc.com/news/ uk-northern-ireland-55094032 [Last checked 16 dec 2023]

[66]   Belfast Telegraph (1975). "Slaughter in back alley", in *Belfast Telegraph*, Tuesday 25 November 1975.

[67]   Browne M. (2002), "On the first anniversary of Martin O'Hagan's murder", in *London Freelance Journal*, 3 January 2003, http:// www.londonfreelance.org/ohagan7.html [Last checked 16 dec 2023]

[68]   Carroll J. (2007). "Tales from the Dark Side", in *The irish Times*, Feb 2 2007. https://www.irishtimes.com/culture/tales-from-the-dark-side-1.1193266 [Last checked 16 dec 2023]

[69]   Dillon M. (2009). "A legacy of hatred", In Belfast Telegraph, Sun 24 May 2009.

[70]   Gorman R. (2023). "Sick new UVF memorial honouring notorious Shankill Butchers gang causes outrage", in *Sunday World*, Sun 23 Jul 2023.   https://www.sundayworld.com/news/irish-news/sick-new-uvf-memorial-honouring-notorious-shankill-butchers-gang-causes-outrage/a1850609390.html [Last checked 16 dec 2023]

[71]   Independent (2000). "Exposure sealed fate of notorious activists", in *Independent*, Thursday 24 August 2000. https://www.independent.co.uk/news/uk/this-britain/exposure-sealed-fate-of-notorious-activists-710345.html [Last checked 16 dec 2023]

[72]   Jordan H. (2011) "Lenny Murphy – Local Hero, Shankill Road Style", from Jordan H. (2011) *Milestones in Murder: Defining Moments in Ulster's Terror War.* Random House.

[73]   McAleese D. (2009). "Shankill Butcher William Moore found dead in flat was due to face questioning", in *Belfast telegraph*, 18 May 2009.

[74]   Preston A. (2023). "Shankill Butcher Eddie McIlwaine spotted putting up UVF flag", 25 June 2023, in *Irish news*. https://www.irishnews.com/news/northernirelandnews/2023/06/25/news/shankill_butcher_eddie_mcilwaine_spotted_putting_up_uvf_flag-3380086/ , [Last checked 16 dec 2023]

[75]   Reid K. (2023). "Shankill Butcher Eddie McIlwaine spotted marching in Whiterock parade", in *Belfast telegraph*, Sat 24 Jun 2023. https://m.belfasttelegraph.co.uk/news/northern-ireland/shankill-butcher-eddie-mcilwaine-spotted-marching-in-whiterock-parade/a644644071.html [Last checked 16 dec 2023]

[76]   Sengupta K. (2000). "Spate of savage killings blamed on loyalist feud", in *The Independent*, 15 July 2002, https://web.archive.org/web/20100926194149/http://www.independent.co.uk/news/uk/this-britain/spate-of-savage-killings-blamed-on-loyalist-feud-706364.html   [Last checked 16 dec 2023]

[77]   Tinney A. (2017). "Lenny Murphy's brother was real boss of Shankill Butchers gang, says new book", in *Belfast Telegraph*, Friday 16 Jun 2017.

[78]   Walker G. (1996). "Shankill Butcher is free", in *Belfast Telegraph*, Sat 26 Oct 1996

General bibliography

[79]   Algee-Hewitt, M. (2017). *Canon/Archive: Studies in Quantitative Formalism from the Stanford Literary Lab*, n+1 Foundation.

[80]   Algee-Hewitt, M. (Forthcoming). *The Afterlife of Aesthetics: Literature, the Sublime, and the Art of Criticism in the Long Eighteenth Century*, Cambridge University Press.

[81]   Allington, D., Brouillette, S., Golumbia, D. (2017). *Neoliberal Tools (and Archives): A Political History of Digital Humanities,* at https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/ [checked on 31st July 2023]

[82]   Anderson, C. (2008). *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired: https://www.wired.com/2008/06/pb-theory/ [checked on 1st of Agust 2023]

[83]   Aristotle (1933). *Metaphysics. Vol. 1*. Harvard University Press.

[84]   Ashley, D. R., Herrmann, V., Friggstad, Z., & Schmidhuber, J. (2022). On Narrative Information and the Distillation of Stories. *arXiv preprint* arXiv:2211.12423.

[85]   Bakke, K. M., & Rickard, K. (2023). *"Ten Pound Touts": Post-conflict Trust and the Legacy of Counterinsurgency in Northern Ireland*. United Nations University World Institute for Development Economics Research.

[86]   Bal M. (2009). *Narratology: Introduction to the Theory of Narrative*, University of Toronto Press.

[87]   Bamman D., Underwood, T., & Smith N. A. (2014). *A Bayesian Mixed Effects Model of Literary Character*, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 370–379, Association for Computational Linguistics.

[88]   Barassi, V. (2020). *Child data citizen: How tech companies are profiling us from before birth*, MIT Press.

[89]   Bardon, J. (2011). *The plantation of Ulster: War and conflict in Ireland.* Gill & Macmillan Ltd.

[90]   Benito-Santos, A., & Sánchez, R. T. (2020). A data-driven introduction to authors, readings, and techniques in visualization for the digital humanities. *IEEE Computer Graphics and Applications*, 40(3), 45-57.

[91]   Berry, D. M. (2022). Critical digital humanities. *The Bloomsbury Handbook to the Digital Humanities*, 125.

[92]   Bia, A., Pedreño, A. (2001). The Miguel de Cervantes Digital Library: the Hispanic Voice on the Web, *Literary and Linguistic Computing*, Volume 16, Issue 2, June 2001, Pages 161–177.

[93]   Biber, D. (1985). *Investigating macroscopic textual variation through multifeature/multidimensional analyses*, Linguistics, 23(2), 337–360.

[94]   Bleich, D. (2019). *Subjective criticism*. John Hopkins University Press.

[95]   Boal, F. W. (2002). Belfast: walls within. *Political Geography*, *21*(5), 687-694.

[96]   Bode, K. (2019). *A World of Fiction: Digital Collections and the Future of Literary History*, University of Michigan Press.

[97]   Bode, K. (2020). *Why You Can't Model Away Bias*, Modern Language Quarterly, 81 (1), 95–124.

[98]   Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). *Enriching word vectors with subword information*. Transactions of the association for computational linguistics, 5, 135-146.

[99]   Borooah, V. K., & Knox, C. (2017). *Inequality, segregation and poor performance: the education system in Northern Ireland.* Educational Review, 69(3), 318-336.

[100]   Bottiroli, G. (2018). *Return to Literature. A Manifesto in Favour of Theory and against Methodologically Reactionary Studies (Cultural Studies etc.)*. Comparatismi, (3), 1–37.

[101]   Bourdieu P. (2004). *Science of Science and Reflexivity*, Polity.

[102]   Brennan T., *The Digital-Humanities Bust* (2017). The chrnoicle of higer education, October 15.

[103]   Bridgman, P. W. (1927). *The logic of modern physics* (Vol. 3). Macmillan.

[104]   Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. https://doi.org/10.48550/arxiv.2303.12712

[105]   Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR.

[106]   Bushell, S., Butler, J., Hay, D., Hutcheon, R., & Butterworth, A. (2021). *Chronotopic Cartography: Mapping Literary Time-Space.* Journal of Victorian Culture, 26(2), 310-325.

[107]   Castells, M. (2000). *The Rise of the Network Society*, Blackwell.

[108]   Cawthon, N., and Vande Moere, A. (2007). *The effect of aesthetic on the usability of data visualization,* 11th International Conference Information Visualization (IV'07). IEEE.

[109]   Chatzimparmpas, A., Martins, R. M., Jusufi, I., & Kerren, A. (2020). *A survey of surveys on the use of visualization for interpreting machine learning models*. Information Visualization, 19(3), 207-233.

[110]   Chaudhuri, Arindam & Mandaviya, Krupa & Badelia, Pratixa & Ghosh, Soumya. (2017). *Optical Character Recognition Systems for Different Languages with Soft Computing*. Springer International Publishing.

[111]   Codd, E. F. (1970). *A relational model of data for large shared data banks*. Commun. ACM 13, 6, 377–387.

[112]   Cohen, M. (1999). *The Sentimental Education of the Novel*, Princeton University Press.

[113]   Collitz, K. H. (1931). Verbs of motion in their semantic divergence. *Language*, 7(1), 5-112.

[114]   Conathan, L. (2011). *Archiving and language documentation*. In P. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages*, Cambridge University Press.

[115]   Conrad K. (2014) *What the Digital Humanities Can't Do*, The Chronicle Review (The Chronicle of Higher Education), 8 Sept.

[116]   Croce, B. (1922). *Estetica come scienza dell'espressione e linguistica generale: teoria e storia*, Laterza.

[117]   D'Ignazio, C., Klein L.F. (2020), *Data Feminism,* MIT Press.

[118]   Da, N. Z. (2019). *The computational case against computational literary studies*, Critical inquiry, 45(3), 601-639.

[119]   Darby, J. (1995). *Conflict in Northern Ireland: A background essay*, in Facets of the conflict in Northern Ireland, Palgrave Macmillan UK, pp. 15-23.

[120]   Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9), 92-103.

[121] de Bustos, J. C. M., & Izquierdo-Castillo, J. (2019). Who will control the media? The impact of GAFAM on the media industries in the digital economy. *Revista Latina de Comunicación Social*, (74), 803-821.

[122] De Mauro, A., Greco, M., & Grimaldi, M. (2016). *A formal definition of Big Data based on its essential features*. Library review, 65(3), 122-135.

[123] Derrida, J., and Prenowitz, E. (1995). *Archive Fever: A Freudian Impression*, Diacritics, vol. 25, no. 2, 1995, pp. 9–63.

[124] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. ArXiv, abs/1810.04805.

[125] Dick, E. A., Ballard, M., Alwan-Walker, H., Kashef, E., Batrick, N., Hettiaratchy, S., & Moran, C. G. (2018). *Bomb blast imaging: bringing order to chaos*. Clinical Radiology, 73(6), 509-516.

[126] Dillon, M. (1999). *The Shankill Butchers: the real story of cold-blooded mass murder.* Psychology Press.

[127] Dobson, J. E. (2019). *Critical digital humanities: the search for a methodology*. University of Illinois Press.

[128] Dobson, J.E. (2019). *Critical Digital Humanities: The Search for a Methodology*, University of Illinois Press.

[129] Drucker, J. (2014). *Graphesis: Visual forms of knowledge production*,Vol. 6, Harvard University Press.

[130] Drucker, J. (2017). *Why Distant Reading Isn't*. PMLA : Publications of the Modern Language Association of America, 132(3), 628–635.

[131] Duever, M., & McGinn, E. (2020). *Teaching GIS in a Digital Humanities Environment.* Journal of Map & Geography Libraries, 16(3), 229-238.

[132] Dyrstad, K., & Hillesund, S. (2020). *Explaining support for political violence: grievance and perceived opportunity*. Journal of conflict resolution, 64(9), 1724-1753.

[133]   Eagleton, T. (2006). *Criticism and ideology: A study in Marxist literary theory,* Verso.

[134]   Einstein, A. (1905). *On the electrodynamics of moving bodies*, Annalen der physik, 17.10, 891-921.

[135]   Estill, L. (2019) *Digital Humanities' Shakespeare Problem*, Humanities, 8, 45.

[136]   Etzkowitz, H. (2022). *Entrepreneurial university icon: Stanford and Silicon Valley as innovation and natural ecosystem*, Industry and Higher Education, 36(4), 361–380.

[137]   Eve, M. P. (2019). *Close reading with computers: textual scholarship, computational formalism, and David Mitchell's Cloud atlas* (p. 272). Stanford University Press.

[138]   Fadem, M. E. R., & Maureen, E. (2015). *The Literature of Northern Ireland: Spectral Borderlands*. Palgrave Macmillan.

[139]   Fair, J. D. (1972). *The Anglo-Irish Treaty of 1921: Unionist Aspects of the Peace.* Journal of British Studies, 12(1), 132-149.

[140]   Fairclough, N. (2013). *Critical discourse analysis: The critical study of language.* Routledge.

[141]   Feehan, J. (2015). *Bobby Sands: And the Tragedy of Northern Ireland.* Open Road Media.

[142]   Felski, R. (2015). *The limits of critique*. University of Chicago Press.

[143]   Ferguson E. L., Hegarty M. (1994). *Properties of cognitive maps constructed from texts, in Memory & Cognition*, Volume 22, Issue 4, pp. 456–457.

[144]   Fish, S. (2012), *Mind Your P's and B's: The Digital Humanities and Interpretation*, Opinionator Blog, New York Times, 23 Jan.

[145]   Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*, OUP Oxford.

[146]   Floridi, Luciano (2023). AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models, *Philosophy and Technology*.

[147]   Fludernik, M. (2002). *Towards a "natural" narratology*, Routledge.

[148]   Foley, R., & Murphy, R. (2015). Visualizing a spatial archive: GIS, digital humanities, and relational space.

[149]   Foucault, M. (1976). *The archaeology of knowledge.* Harper & Row.

[150]   Frodeman, R., Klein, J. T., & Pacheco, R. C. D. S. (Eds.). (2017). *The Oxford handbook of interdisciplinarity*, Oxford University Press.

[151]   Frye, N. (1957). *Anatomy of Criticism: Four Essays*, Atheneum.

[152]   Galina Russell, I. (2014) *Geographical and linguistic diversity in the Digital Humanities*, Literary and Linguistic Computing, Oxford University Press, 29, 307-316.

[153]   Gandomi, A., & Haider, M. (2015). *Beyond the hype: Big data concepts, methods, and analytics*. International Journal of Information Management, 35(2), 137–144.

[154]   Genette, G. (1980). *Narrative discourse: An essay in method* (Vol. 3). Cornell University Press.

[155]   Gibbons, A. (2012). Multimodality, cognition, and experimental literature (Vol. 3), Routledge.

[156]   Gitelman, L. (ed.)(2013). *Raw data is an oxymoron*. MIT press.

[157]   Gold, M. (2012). *Debates in the Digital Humanities 2012*, University of Minnesota Press.

[158]   Gold, M. & Klein, L. (2016) *Debates in the Digital Humanities 2016*, University of Minnesota Press.

[159]   Gormley-Heenan, C., & Aughey, A. (2017). *Northern Ireland and Brexit: Three effects on 'the border in the mind'*. The British Journal of Politics and International Relations, 19(3), 497-511.

[160]   Graham, E. (2017). *Introduction: Data visualisation and the humanities*, English Studies, 98 (5), 449-458.

[161] Grimshaw, M. (2018). *Towards a manifesto for a critical digital humanities: critiquing the extractive capitalism of digital society*. Palgrave Communications, 4(1), 1-8.

[162] Hall, S. (1997). *Representation: Cultural Representations and Signifying Practices*, SAGE Publications.

[163] Hamon, P. (2004). *What is a Description?* in Bal, M. Narrative Theory: Critical Concepts in Literary and Cultural Studies, 1, 309-340.

[164] Harvey, D. (1973). *Social Justice and the City*. University of Georgia Press.

[165] Hawking S. & Penrose R., (1970). *The singularities of gravitational collapse and cosmology*, Proc. R. Soc. A314529–548.

[166] Hawking, S., & Penrose, R. (2010). *The nature of space and time*, Princeton University Press.

[167] Hayles, N. K. (2007). *Electronic literature: What is it?*. in "Doing Digital Humanities: Practice, Training, Research", 197-226.

[168] Hearst. M. A. (1997). *TextTiling: segmenting text into multi-paragraph subtopic passages*. Computational Linguistic 23, 1, March, 33–64.

[169] Heidegger, M. (2010) *Being and time*, Suny Press.

[170] Hepworth, J. (2021). *'The Age-Old Struggle': Irish republicanism from the Battle of the Bogside to the Belfast Agreement, 1969-1998*. Oxford University Press.

[171] Herman D., Story logic (2002). *Problems and possibilities of narrative*, University of Nebraska Press.

[172] Herman, P. C. (2015) Terrorism and the Critique of American Culture: John Updike's Terrorist. *Modern philology.* 112 (4), 691–712.

[173] Hickmann, M., Engemann, H., Soroli, E., Hendriks, H., & Vincent, C. (2017). Expressing and categorizing motion in French and English. *Motion and space across languages: Theory and applications*, 61-94.

[174] Hopkinson, M. (2004). *Green Against Green–The Irish Civil War: A History of the Irish Civil War, 1922–1923*. Gill & Macmillan Ltd.

[175]   Horneber, D., Laumer, S (2023). *Algorithmic Accountability*. Bus Inf Syst Eng.

[176]   Human, S., & Cech, F. (2021). A human-centric perspective on digital consenting: The case of gafam. In *Human Centred Intelligent Systems: Proceedings of KES-HCIS 2020 Conference* (pp. 139-159). Springer Singapore.

[177]   IBM Data Processing Division (1973). *Jesuit Father Uses Computer to Analyze Works of St. Thomas Aquinas*, Modern Data. 6 (9): 39–40.

[178]   Jänicke, S. (2016). *Valuable research for visualization and digital humanities: A balancing act*. In Workshop on Visualization for the Digital Humanities, IEEE VIS (Vol. 7).

[179]   Jankowska, M. A. (2021). *Digital Humanities in Academic Library: Creating Narratives through Spatial Data*. Biblioteka, (25 (34)), 201-213.

[180]   Jessica Barness, Amy Papaelias (ed.s)(2015). *Critical Making: Design and the Digital Humanities*, University of Cincinnati, School of Design.

[181]   Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. de las, Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., & Sayed, W. E. (2023). Mistral 7B. https://doi.org/10.48550/arxiv.2310.06825

[182]   Jockers, M. (2013). *Macroanalysis: Digital Methods and Literary History*, University of Illinois Press.

[183]   John Doyle. (2018). *Reflecting on the Northern Ireland Conflict and Peace Process: 20 years since the Good Friday Agreement*, Irish Studies in International Affairs.

[184]   Johnson, N. C. (1999). The Cartographies of Violence: Belfast's Resurrection Man. *Environment and Planning D: Society and Space*, 17(6), 723-736.

[185]   Keen, S. (2006). *A Theory of Narrative Empathy*. Narrative, 14(3), 207–236.

[186]   Kennedy-Andrews, E. (2003). *Fiction and the Northern Ireland Troubles since 1969: (De-)Constructing the North*. Four Courts Press.

[187]   Kenney, M. (2000). *Understanding Silicon Valley: The anatomy of an entrepreneurial region*, Stanford University Press.

[188]   Keynes, J. M. (1936). *The General Theory of Employment Interest and Money*, Macmillan.

[189]   Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2021). *The age of AI: and our human future*. Hachette.

[190]   Klein, L. & Gold, M.(eds.). *Debates in the Digital Humanities 2019*, University of Minnesota Press.

[191]   Kleymann, R., Niekler A., Burghardt M. (2022). *Conceptual Forays: A Corpus-Based Study of 'Theory' in Digital Humanities Journals*, Journal of Cultural Analytics, vol. 7, no. 4.

[192]   Kneubuehl, B. P. (Ed.). (2011). *Wound ballistics: basics and applications*. Springer Science & Business Media.

[193]   Knoke, D., & Yang, S. (2019). *Social network analysis*, SAGE publications.

[194]   Kobourov, S. G. (2012). *Spring Embedders and Force Directed Graph Drawing Algorithms*, arXiv:1201.3011.

[195]   Kordzadeh, N. & Ghasemaghaei, M (2022). *Algorithmic bias: review, synthesis, and future research directions*. European Journal of Information Systems, 31:3, 388-409.

[196]   Kristeva, J. (1980). *Desire in language: A semiotic approach to literature and art*, Columbia University Press.

[197]   Kuhn, T. S. (1962). *The structure of scientific revolutions,* University of Chicago Press.

[198]   Kumar, A., (2021). *Digital Dante*, in Gragnolati M., Lombardi E., and Southerden F. (eds), *The Oxford Handbook of Dante*, Oxford Handbooks.

[199]   Lareau, A. (2003). *Unequal Childhoods: Class, Race, and Family Life*. University of California Press.

[200]   Lefebvre, H. & Nicholson-Smith, D. (1991). *The production of space*. Vol. 142. Blackwell.

[201]   Lenci, A. (2008). *Distributional semantics in linguistic and cognitive research*. Italian journal of linguistics, 20(1), 1-31.

[202]   Levenshtein, V. I. (1966). *Binary codes capable of correcting deletions, insertions and reversals*, Soviet Physics Doklady 10, no. 8, 707-710.

[203]   Linde C., Labov W. (1975). *Spatial networks as a site for the study of language and thought*, in Linguistic Society of America, Vol. 51, No. 4, Washington, 1975, pp. 924-939.

[204]   Liu, A. (2012). *Where Is Cultural Criticism in the Digital Humanities?*", 490–, in Gold, M. (ed)(2012). *Debates in the Digital Humanities 2012*, University of Minnesota Press.

[205]   Lotman, J. M., & O'Toole, L. M. (1975). *Point of View in a Text.* New Literary History, 6(2), 339-352.

[206]   Loyola-González O. (2019). *Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View*, in IEEE Access, vol. 7, pp. 154096-154113.

[207]   Manovich, L. (2008). *Cultural analytics: Analysis and visualization of large cultural data sets*, The MIT Press.

[208]   Martin Hilbert, Priscila López (2011). *The World's Technological Capacity to Store, Communicate, and Compute Information*. Science332,60-65.

[209]   McAtackney, L. (2011). *Peace maintenance and political messages: The significance of walls during and after the Northern Irish 'Troubles'*. Journal of Social Archaeology, 11(1), 77-98.

[210]   McCarthy, D. (2000). Belfast Babel: Postmodern Lingo in Eoin McNamee's "Resurrection Man". *Irish University Review*, 30(1), 132-148.

[211]   McCarty, W. (2012). *Collaborative research in the digital humanities*, Collaborative research in the digital humanities, 22(5), 1-10.

[212]   McCarty, W., & Deegan, M. (Eds.). (2012). *Collaborative Research in the Digital Humanities*, Routledge.

[213]   McCarty, Willard (2005). *Humanities computing*, Palgrave Macmillan.

[214]   McGann, J. (1994). *The Complete Writings and Pictures of Dante Gabriel Rossetti: A Hypermedia Research Archive*, Text, Vol. 7, pp. 95-105.

[215]   McGann, J. (2004). *Marking texts of many dimensions*, in A companion to digital humanities, 198-217.

[216]   McGlynn, M. (2016). *Narratives of Class in New Irish and Scottish Literature: From Joyce to Kelman, Doyle, Galloway, and McNamee.* Springer.

[217]   McSweeney, B. (1998). *Identity, Interest and the Good Friday Agreement.* Irish Studies in International Affairs, 9, 93-102.

[218]   Michel, Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, Pickett, J., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). *Quantitative Analysis of Culture Using Millions of Digitized Books*, Science. 2011 Jan 14, 331(6014), pp. 176-82.

[219]   Mikolov, T., & Chen, K. & Corrado, G.s & Dean, J.. (2013). *Efficient Estimation of Word Representations in Vector Space*, Proceedings of Workshop at ICLR.

[220]   Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781.

[221]   Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.

[222]   Minkowski, H. (1908–1909), *Raum und Zeit* [Space and Time], Physikalische Zeitschrift, 10, pp. 75–88.

[223]   Mohseni, M., Gast, V., & Redies, C. (2021). *Fractality and variability in canonical and non-canonical English fiction and in non-fictional texts*, Frontiers in psychology, 12, 599063.

[224]   Moretti F. (2022). *Falso movimento. La svolta quantitativa nello studio della letteratura*, Nottetempo.

[225]   Moretti, F. (2013). *"Operationalizing": or, the function of measurement in modern literary theory.*

[226]   Moretti, F. (2013). *Operationalizing*, New Left Review, 84.

[227]   Moretti, F. (2020). *The roads to Rome: Literary Studies, Hermeneutics, Quantification*, New Left Review, 124, pp. 125–36.

[228]   Moretti, F. & Sobchuk, O. (2019). *Hidden in plain sight: Data Visualization in the Humanities*, New Left Review, 118, pp. 86-115.

[229]   Murphy, G. L, & Douglas L. M. (1985). The Role of Theories in Conceptual Coherence. *Psychological review 92.3* (1985): 289–316.

[230]   Murphy, R. (1986). *Walter Long and the making of the Government of Ireland Act, 1919-20*. Irish Historical Studies, 25(97), 82-96.

[231]   Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1), 3-26.

[232]   Nathan, D. (2011). *Digital archiving*, in P. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (Cambridge Handbooks in Language and Linguistics, pp. 255-274). Cambridge University Press.

[233]   Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). A Comprehensive Overview of Large Language Models. *arXiv preprint arXiv:2307.06435.*

[234] Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., & Woodard, D. (2018). *Surveying Stylometry Techniques and Applications*, ACM Computing Surveys, 50(6), 1–36.

[235] Nelson, Gregory S. (2019). *Bias in artificial intelligence*, North Carolina medical journal 80.4, pp. 220-222.

[236] Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., ... & Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1356.

[237] Nussbaum, M. (1986). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*, Cambridge University Press.

[238] Nyhan, J., Terras, M., & Vanhoutte, E. (2013). *Defining Digital Humanities*, Ashgate.

[239] Oppenheimer, A. R. (2008). *IRA, the bombs and the bullets: A history of deadly ingenuity*. Irish Academic Press.

[240] Patterson, H. (2012). *Unionism after Good Friday and St Andrews.* The Political Quarterly, 83(2), 247-255.

[241] Pavel, T. G. (1975). *"Possible worlds" in literary semantics*. The journal of aesthetics and art criticism, 34(2), 165-176.

[242] Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *In Proceedings of the 2014 conference on empirical methods in natural language processing* (EMNLP), pp. 1532-1543.

[243] Pennington, J., Socher, R., and Manning, C. (2014). *GloVe: Global Vectors for Word Representation, i*n Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Association for Computational Linguistics.

[244] Pessach, D., & Shmueli, E. (2023). *Algorithmic fairness*, in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook* (pp. 867-886), Springer International Publishing.

[245] Piatti, B., Reuschel, A. K., & Hurni, L. (2009, November). *Literary geography–or how cartographers open up a new dimension for literary studies*. In Proceedings of the 24th International Cartography Conference (pp. 1-12).

[246] Picard, R. W. (2008). Toward machines with emotional intelligence.

[247] Piotrowski, M. (2018). *Digital Humanities: An Explication*. INF-DH-2018.

[248] Propp, V. (1968). *Morphology of the Folktale*, University of Texas Press.

[249] R. Mittal and A. Garg, "Text extraction using OCR: A Systematic Review," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 357-362.

[250] Ramsay, S. (2011). *Reading Machines: Toward and Algorithmic Criticism*, University of Illinois Press.

[251] Ramsay, S., & Rockwell, G. (2012). *Developing things: Notes toward an epistemology of building in the digital humanities*, Debates in the digital humanities, 75-84.

[252] Reuschel, A. K., Piatti, B., & Hurni, L. (2009). *Mapping literature. the prototype of "A Literary Atlas of Europe"*. In Proceedings of the 24 th International Cartographic Conference.

[253] Risam, R. (2017). *Other Worlds, Other DHs: Notes Towards a DH Accent*, Digital Scholarship in the Humanities, 32.2, 377.

[254] Rockwell, G., & Sinclair, S. (2022). Hermeneutica: Computer-assisted interpretation in the humanities. mit Press.

[255] Russell, B. (1925). *The ABC of Relativity*, Harper & Brothers.

[256] Russell, S. (2021). Human-compatible artificial intelligence. *Human-like machine intelligence*, 3-23.

[257]   Ryan M.L (2003), *Cognitive maps and the construction of narrative space*, in Herman D., *Narrative theory and the cognitive sciences*, Stanford, CSLI Publications.

[258]   Ryan M.L. (2012), *Space*, in De Gruyter W., *The living handbook of narratology*, Hamburg University Press.

[259]   Sard, R. D. (1970). *Relativistic Mechanics: Special Relativity and Classical Particle Dynamics*, Benjamin.

[260]   Saussure, F. d., Sechehaye, A., Riedlinger, A. (1966). *Course in General Linguistics*. McGraw-Hill.

[261]   Scanlan, M. (2001). *Plotting terror: novelists and terrorists in contemporary fiction*. University of Virginia Press.

[262]   Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2008). *A companion to digital humanities,* John Wiley & Sons.

[263]   Schuessler, J. (2017). Reading by the numbers: when big data meets literature. *New York Times*, 30.

[264]   Schumacher, M. K. (2023). *Orte und Räume im Roman: ein Beitrag zur digitalen Literaturwissenschaft*. Springer Nature.

[265]   Shah, H. (2018). *Algorithmic accountability*, Philosophical Transactions of the Royal Society, Mathematical, Physical and Engineering Sciences, 376(2128), 20170362.

[266]   Sheppard, E. (1995). *GIS and society: towards a research agenda*, Cartography and Geographic Information Systems, 22(1), 5-16

[267]   Sinclair, S., Ruecker, S., Radzikowska, M., & Inke, I. N. K. E. (2013). *Information visualization for humanities scholars,* Literary studies in the digital age, an evolving anthology.

[268]   Sinykin D., So R. J., Young J. (2019). *Economics, Race, and the Postwar US Novel: A Quantitative Literary History*, American Literary History, Volume 31, Issue 4, Pages 775–804.

[269]   Slobin, D. I. (1996). Two ways to travel: Verbs of motion in English and Spanish. *Grammatical constructions: Their form and meaning*, 195217.

[270]   Smalheiser, N. R. (2002). *Informatics and hypothesis-driven research*. EMBO reports, 3(8), 702-702.

[271]   Stachel, J. (1998). *Einstein's Miraculous Year: Five Papers that Changed the Face of Physics*, Princeton University Press.

[272]   Stainer, J. (2006). Localism, signification, imagination: de-stabilizing sectarian identities in two fictionalized accounts of 'Troubles' Belfast. *Social & Cultural Geography*, 7(1), 103-126.

[273]   Stamenkovic, D., & Tasic, M. (2013). English verbs of motion and prototype theory. *British and American studies,* 19, 210.

[274]   Stein, H. (1968). *On Einstein-Minkowski Space-Time*, The journal of Philosophy, 65(1), pp. 5-23.

[275]   Svensson, P. (2013). *Humanities computing as digital humanities*, p. 159-186, in Terras, M., Nyhan, J., Vanhoutte, E. (2013). *Defining digital humanities: a reader,* Routledge.

[276]   Tannam, E. (2001). *Explaining the Good Friday agreement: A learning process.* Government and Opposition, 36(4), 493-518.

[277]   Taylor, Peter (1999). *Loyalists*. Bloomsbury Publishing.

[278]   Todorov, T. (2007). *La littérature en péril.* Flammarion.

[279]   Törnberg, P. (2023). How to use LLMs for Text Analysis. *arXiv preprint arXiv:2307.13106*.

[280]   Torop, P., Osimo, B. (2000). *La traduzione totale.* Guaraldi Logos.

[281]   Touvron, H., Lavril, T., Izacard, G., Martint, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models. https://doi.org/10.48550/arxiv.2302.13971

[282] Underwood, T. (2017) A Genealogy of Distant Reading, *Digital Humanities Quarterly*, vol. 11, 2.

[283] Vanier, M. (2009). *Territoires, territorialité, territorialisation: controverses et perspectives*. Presses universitaires de Rennes.

[284] Vatri, A., & McGillivray, B. (2018). The Diorisis Ancient Greek Corpus: Linguistics and Literature, *Research Data Journal for the Humanities and Social Sciences*, 3(1), 55-65.

[285] Wade, T., & Sommer, S. (2006). *A to Z GIS, An illustrated dictionary of geographic information systems*.

[286] Walker, C. P. (1984). Irish Republican Prisoners—Political detainees, prisoners of war or common criminals?. *Irish Jurist (1966-)*, 19(2), 189-225.

[287] Walsh, D. (2000). *Bloody Sunday and the rule of law in Northern Ireland.* Springer.

[288] Wang, X., Li, X., Yin, Z., Wu, Y., & Liu, J. (2023). Emotional intelligence of large language models. *Journal of Pacific Rim Psychology*, 17.

[289] Weinstock, C. B., & Goodenough, J. B. (2006). *On system scalability*, Carnegie Mellon University, Software Engineering Institute.

[290] Westphal, B. (2011). *Geocriticism: Real and fictional spaces*. Springer.

[291] Windhager, F., Salisu, S., & Mayr, E. (2019). *Exhibiting uncertainty: Visualizing data quality indicators for cultural collections*. In Informatics (Vol. 6, No. 3, p. 29). MDPI.

[292] Wing, J.M. (2008) *Computational thinking and thinking about computing*. Philos Trans A Math Phys Eng Sci. Oct 28;366(1881):3717-25.

[293] Wittgenstein, L. (1933), *Tractatus logico-philosophicus*. [Reprinted, with a few corrections], Brace.

[294] Wu, J., Gaur, Y., Chen, Z., Zhou, L., Zhu, Y., Wang, T., Li, J., Liu, S., Ren, B., Liu, L., & Wu, Y. (2023). On decoder-only architecture for speech-to-text and large language model integration. *https://doi.org/ 10.48550/arxiv.2307.03917*

[295] Zhang, N. N. (2002). Movement within a spatial phrase. *Perspectives on prepositions*, 47-63.

[296] Žliobaitė, I (2017). Measuring discrimination in algorithmic decision making. *Data Min Knowl Disc* 31, 1060–1089.