“Digital Comparative Literature”

ICLA Short-term Research Committee

The objective of the Short-term Research Committee on “Digital Comparative Literature” (DCL) is that of addressing a double challenge: the one posed by the growing research field of Digital Humanities (Schreibman, Siemens, and Unsworth 2004; Jannidis, Kohle, and Rehbein 2017) to the discipline of Comparative Literary Studies (Tötösy de Zepetnek 1995; 1996; 2013; 2017; Li 2022; Eve 2022), together with the one—complementary and equally enriching—posed by Comparative Literature to Digital Humanities. In fact, while on the one hand the development of techniques like computer-assisted distant reading and the digitization of large textual archives have opened new perspectives for the study of literature, the terrain of Comparative Literary Studies remains one of the roughest for digital research. Suffice it to mention the issue of multilingualism, as computational analysis tools are still mainly tailored for the English language and multilingual literary corpora are still hard to find (with the only, still isolated exceptions of the *Eltec* collection and the *DraCor* corpora).

To face this double challenge, DCL will leverage the complementary research paths of its members, all active in between comparative literary studies and digital humanities, focusing on different aspects of their intersection.

Overall, five main areas of research can be identified to describe the range of action of DCL. First are **distant reading techniques and computational literary studies**. One of the most stimulating debates in the Humanities, but perhaps more crucial in the field of Literary Studies, concerns the usefulness (and legitimacy) of computational methods of text analysis or computer-assisted reading (Wilkens 2015). What could be seen as a serious threat to a certain order of knowledge, can offer new opportunities for the renewal of humanistic methods, helping to empirically compare working hypotheses with the data collected and taking into account not only the literary works themselves, but also the intellectual and historical context in which they emerged. For this reason, computer-aided analysis should be viewed not as a replacement, but as a natural extension of traditional humanistic approaches. However, it seems important to develop a more dense theoretical discussion on the epistemological implications of using computational tools to process large collections of texts and data (e.g. Jockers 2013), calling on contributions from various academic perspectives, without losing sight of the evaluation of case studies that can help to rethink the access to the archive and enrich the analysis of literary and cultural texts. By allowing a considerable amount of data to be gathered, by mobilizing computational tools and also by proposing new ways of processing

information through statistical or algorithmic methods and visualization strategies, this interdisciplinary field can bring significant changes in terms of the assumptions, methods and objectives of literary and cultural research.

While some skepticism still persists, we firmly believe that it can be overcome through the implementation of a paradigm allowing the coexistence of traditional and computational methods. Recent projects like the *Distant Reading for European Literary History* COST Action, together with the works of authors like Andrew Piper, Katherine Bode, Hoyt Long, Richard Jean So, Alan Liu, Ted Underwood, among many others, increasingly point to these convergence efforts.

A second area of research concerns the **possibilities and limitations in the creation and usage of large digital archives** (see for example the discussion following studies like Michel et al. 2011). Today, it is no longer possible to rethink the scope of Comparative Literary Studies without a deep discussion about the broader implications of structural and content diversity in digital archives (Villanueva and Shiri 2021), and without considering the relevance of metadata standards, user engagement, and the impact of archives on preserving and disseminating cultural heritage (Dahlgren and Hansson 2020). The advent of digital technologies has revolutionized archival practices, so the aim is to generate new approaches and identify new research problems in various areas of knowledge, namely at the literary, artistic and cultural level, from a computational and data-driven perspective, highlighting the way in which codification, algorithmization and narratives of transferability are capable of revealing new dimensions in the experience of the archive. To deal with highly complex environments such as contemporary digital environments, whether due to the amount of data they move or the heterogeneity of practices they call upon, has brought to the forefront the debate on the relevance of using digital technologies for a more sophisticated approach (therefore qualitatively better and not just impressive from a quantitative point of view) of cultural phenomena.

At the present moment, we face issues that cannot be resolved using technology alone and this fact requires a decisive contribution from humanistic culture. Some of the new questions that permeate our technological world are, in reality, old questions, albeit with different formulations and other implications, which have been at the center of human concerns for a long time. In this sense, seeking to reduce the mission of Digital Humanities to the mere application of digital tools and resources is a simplistic attitude, as the way in which most researchers who recognize themselves as agents (and certainly also as builders) of this field are committed to debating and testing new models of understanding culture and the world.

A third subject we will explore concerns the **transformations of books and reading in a postdigital context**. Some of the most relevant literary and cultural issues and practices in today's society are not linked exclusively to printed textuality, but rather arise associated with technologies and configurations of digital media, with emphasis on multimodal properties (Kress and Leeuwen

2001; Kress 2003). With the advent of digital media and the development of multimedia and hypermedia languages, new practices of literary inscription have emerged, which is why it is necessary to reflect on the impact that such changes have had on the notion of textuality, the profile of technological mediation and modes of reading. For example, by allowing multiple crossings between objects in the physical environment and virtual content, projects using Virtual, Mixed and Augmented Reality allowed the configuration of hypertextual, non-linear reading proposals, requiring a considerable level of ergodic participation and overcoming limitations inherent to the printed book. Also, the success of the Web 2.0 has supported the development of phenomena like Digital Social Reading (Rebora et al. 2021), where (w)readers take a more active role in the so-called Digital Literary Sphere (Murray 2018), producing an unprecedented amount of data about literary reception (ranging from comments and reviews to multimedia products). The complexity of such a landscape inevitably calls for an interdisciplinary effort in the research, which needs to bring together not only Literary Studies and Digital Humanities, but also disciplines such as Psychology and Social Science.

A fourth area of research concerns **data visualization (network analysis, mapping, etc.) and comparative literature**. As noted on multiple occasions, the practise of distant reading implies the substitution of close reading with an analysis of computer-generated visualizations (Jänicke et al. 2015), which sometimes possess "emerging" qualities, revealing aspects of the text that cannot be identified without their support (Moretti 2005; 2011). Besides, visualizations can provide insights into intertextual and intratextual connections and similarities across genre, time and even language boundaries.

Since the publication of Moretti's Graphs, Maps, Trees (2005), naming three types of visualisations, a number of visual methods have been put forward in Literary Studies (Scrivner and Davis 2017), also by adopting many applications developed in other disciplines. Methods for the automated extraction of network data (Named Entity Recognition, coreference resolution) and their evaluation are of particular importance (Fischer et al. 2017). The application of network analysis methods, usually representing character networks, and geographic information systems for mapping literary texts data (Juvan 2015) has even developed into an independent field of research within digital literary studies. In addition, tag clouds, heat maps, timelines, summary plots, clusters, etc., are among the most common data visualization methods for literary analysis. Still, visualizations can also prove misleading, and a deep understanding of their working logic, together with the implications in their usage, is necessary in research (Drucker 2016). At the integration of text analysis and visualization techniques, close interaction between a reader and a text remains essential to literary scholars.

A last subject we will explore is **machine translation, AI, language models and comparative literature**. For the past 20 years, we have been discussing the untranslatability across the East/West

divide from the perspective of the act of cultural translation. Our assumption is: when the reader/writer performs an act of cultural translation, the transculturally encoded hybridized literary texts of double vision will lay bare their truths. When translating from one language to another culturally and linguistically, translators have witnessed that blocked meanings in the contact zone or border zone of untranslatable original language became the residues remaining to be transgressed, transmigrated, transplanted, transported, and translated (Kim 2018). One needs the unblocking technique to reach the authorial "unintended" intention of the original language. The only possible way to overcome the wall of untranslatability across languages and cultures seems to be via a magic door of Wonderland to be opened by the strange Other which emerges in the 21th century.

Cognitive neuroscience and artificial intelligence have undergone revolutionary changes in the past decades, and they now foreground the embodied and environmentally embedded nature of intelligent action. Via computer and information technology, posthumanism has been able to articulate the retreat of the human agent into a larger ecological environment. The current predicament for human agency involves "the sublimation of matter into the digital." In this age of "postmodern sublime," the domain of comparative literature is preoccupied with the subject of machine translation, which is intricately linked to artificial intelligence and language models.

In this context, the relationship between machine learning and deep learning in AI is intriguing. Machine translation is a practical implementation of deep learning and machine learning that entails the generation of a text in the target language automatically from the source language. Deep learning, artificial intelligence, and machine learning are all components of contemporary machine translation. By observing and capturing the translation problems and their solutions in the translation units stored in translation memories and other parallel corpora, data-driven machine translation (statistical machine translation) provides the training data from which modern machine translation systems learn. 2015 marked a turning point in digital comparative literature with the introduction of a Stanford University-developed transformative neural machine translation (NMT) system that learned "idiomatic expressions and metaphors" and "the cultural equivalent in another language." A deep neural network, an artificial neural network comprising numerous layers that incorporate "word embeddings," has transformed multilingual translation models from magic to the real (Kenny 2022; Viola and Spence 2023).

References

Dahlgren, Anna, and Karin Hansson. 2020. "The Diversity Paradox: Conflicting Demands on Metadata Production in Cultural Heritage Collections." *Digital Culture & Society* 6 (2): 239–56. https://doi.org/10.14361/dcs-2020-0212.

Drucker, Johanna. 2016. "Graphical Approaches to the Digital Humanities." In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Raymond George Siemens, and John Unsworth, 238–50. Chichester, West Sussex, UK: Wiley/Blackwell.

Eve, Martin Paul. 2022. *The Digital Humanities and Literary Studies*. The Literary Agenda. Oxford, New York: Oxford University Press.

Fischer, Frank, Mathias Göbel, Dario Kampkaspar, Christopher Kittel, and Peer Trilcke. 2017. "Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts." In *DH2017 Book of Abstracts*.

Jänicke, S, G Franzini, M F Cheema, and G Scheuermann. 2015. "On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges." In *Eurographics Conference on Visualization (EuroVis)*, 83–103.

Jannidis, Fotis, Hubertus Kohle, and Malte Rehbein, eds. 2017. *Digital Humanities: eine Einführung*. Stuttgart: J.B. Metzler Verlag.

Jockers, Matthew Lee. 2013. *Macroanalysis: Digital Methods and Literary History*. Topics in the Digital Humanities. Urbana: University of Illinois Press.

Juvan, Marko. 2015. "From Spatial Turn to GIS-Mapping of Literary Cultures." *European Review* 23 (1): 81–96. https://doi.org/10.1017/S1062798714000568.

Kenny, Dorothy, ed. 2022. *Machine Translation for Everyone*. *Language Science Press*. Language Science Press. https://doi.org/10.5281/zenodo.6653406.

Kim, Youngmin. 2018. "Structured Data in Digital Humanities and World Literature." *The Journal of East-West Comparative Literature* 46 (December):427–57. https://doi.org/10.29324/jewcl.2018.12.46.427.

Kress, Gunther. 2003. *Literacy in the New Media Age*. 0 ed. Routledge. https://doi.org/10.4324/9780203299234.

Kress, Gunther, and Theo van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.

Li, Quan. 2022. "Comparative Literature and the Digital Humanities: Disciplinary Issues and Theoretical Construction." *Humanities and Social Sciences Communications* 9 (1): 1–8. https://doi.org/10.1057/s41599-022-01438-4.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, THE GOOGLE BOOKS TEAM, Joseph P. Pickett, et al. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–82. https://doi.org/10.1126/science.1199644.

Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London ; New York: Verso.

———. 2011. "Network Theory. Plot Analysis." *Stanford Literary Lab Pamphlets* 2:2–12.

Murray, Simone. 2018. *The Digital Literary Sphere: Reading, Writing, and Selling Books in the Internet Era*. Baltimore: Johns Hopkins University Press.

Rebora, Simone, Peter Boot, Federico Pianzola, Brigitte Gasser, J Berenike Herrmann, Maria

Kraxenberger, Moniek M Kuijpers, et al. 2021. “Digital Humanities and Digital Social Reading.” *Digital Scholarship in the Humanities* 36 (Supplement_2): ii230–50. https://doi.org/10.1093/llc/fqab020.

Schreibman, Susan, Ray Siemens, and John Unsworth, eds. 2004. *A Companion to Digital Humanities*. Malden, MA: Blackwell.

Scrivner, Olga, and Jefferson Davis. 2017. “Interactive Text Mining Suite: Data Visualization for Literary Studies.” *Proceedings of the Workshop on Corpora in the Digital Humanities (CDH 2017)*, 29–38.

Tötösy de Zepetnek, Steven. 1995. “Consensus Ex Machina. Théorie Des Systèmes et Étude de La Littérature.” In *Littérature et Informatique. La Littérature Générée Par Ordinateur*, edited by Alain Vuillemin and Michel Lenoble, 77–86. Arras: Artois PU.

———. 1996. “The Impact of the Electronic Revolution on Literary Research: A Socio-Semiotic Point of View.” *Dedalus: Revista Portuguesa de Literatura Comparada* 6:199–205.

———, ed. 2013. *Digital Humanities and the Study of Intermediality in in Comparative Cultural Studies*. West Lafayette, Ind.: Purdue Scholarly Publ. Services.

———. 2017. “About the Situation of the Discipline of Comparative Literature and Neighboring Fields in the Humanities Today.” *Comparative Literature: East & West* 1 (2): 176–203. https://doi.org/10.1080/25723618.2017.1387398.

Villanueva, Emily, and Ali Shiri. 2021. “Methodological Diversity in the Evaluation of Cultural Heritage Digital Libraries and Archives: An Analysis of Frameworks and Methods.” *Canadian Journal of Information and Library Science* 43 (3): 316–42.

Viola, Lorella, and Paul Spence, eds. 2023. *Multilingual Digital Humanities*. Digital Research in the Arts and Humanities. London ; New York: Routledge/Taylor & Francis Group.

Wilkens, Matthew. 2015. “Digital Humanities and Its Application in the Study of Literature and Culture.” *Comparative Literature* 67 (1): 11–20.