

Preface

Pablo Valdivia Martin
University of Groningen – Netherlands

When discussing with colleagues and students about the change in the paradigm that we are witnessing in the Humanities, we often find it challenging to define the fundamental elements of our discussion. In this regard, it is more important than ever to find common ground and a baseline for starting the dialogue in the Humanities from wherever we, terminologically, are. One of the goals of this book is to provide a shared territory where it will be easier to move, get inspired, and move forward together. Therefore, we must ask ourselves critical questions and offer tentative working frameworks. Despite commonly and regularly using the term Digital Humanities, it sometimes seems difficult to agree on what we call Digital Humanities. Thus, under the context of this volume, I suggest a working definition of Digital Humanities as an interdisciplinary field that applies computational methods and tools to study human culture and society. It encompasses various disciplines, such as literature, history, art, music, linguistics, philosophy, and more. Digital Humanities aims to enhance our understanding of human expression and experience through analyzing, visualizing, and preserving digital data.

Additionally, when I refer to the term Corpus Studies, also crucial in this book, I opt for a broad definition encompassing a large and structured collection of texts or other forms of data that are representative of a language or a domain. Corpus Studies is essential for Digital Humanities because it provides the raw material for various types of analysis, such as text mining, sentiment analysis, topic modeling, stylometry, and more. Corpus Studies can also help us discover new patterns, trends, and insights not readily observable in individual texts or sources.

Furthermore, Language Technologies, another notion pillared in this volume, are understood in the context of these pages as a branch of artificial intelligence that deals with the processing and generation of natural language. Language Technologies enable us to interact with computers using natural languages, such as speech recognition, machine translation, and chatbots. Language Technologies also facilitate analyzing natural language data, such as natural language understanding, generation, information extraction, summa-

rization, and many more, which are well assessed and reflected in the pages of the present volume.

This book presents examples and applications of how these scientific areas can enrich our knowledge and appreciation of human culture and society. Moreover, this book will inspire new generations of scholars to explore the possibilities and challenges of Digital Humanities in their research and teaching practices.

Therefore, the research present in the chapters of this volume contributes to exploring new avenues regarding the cross-/inter-/multi-disciplinary intersections between the Digital Humanities, Computational Cultural and Literary Studies, and Computational Linguistics. From its very conception, this book results from a joint effort between the University of Antioquia and the University of Groningen and a firm belief in the cross-cutting domain nature of cultural and literary studies and how interdisciplinary approaches to everyday challenges, as recently brought up to the light by the UNESCO “Knowledge Driven Actions (2022), it an essential toolkit for the engineering of our future.

Every chapter has been rigorously evaluated by academic peers who are experts in one of the varied fields of knowledge in this volume. This book will be a valuable resource for researchers, students, and anyone interested in the broadly so-called “digital turn” and the Humanities. I thank the authors who contributed to this book and the academic peers who reviewed their work. I would also like to thank our colleagues at the University of Antioquia and the University of Groningen for their support in bringing this project to fruition. Digital Humanities, Corpus, and Language Technologies are rapidly growing fields that have the potential to revolutionize research across various disciplines. New technologies have opened up new perspectives for research, allowing scientists to analyze data in previously impossible ways.

The first part of this book is devoted to Digital Humanities. This section includes chapters on digital storytelling, data visualization, and text mining. These contributions demonstrate how Digital Humanities can enhance research in various fields, from literature to history to anthropology. For example, one chapter discusses how digital storytelling can be used to teach history. The authors argue that students can better understand historical events and their significance using multimedia elements such as images, videos, and audio recordings. Another chapter discusses how data visualization can be used to analyze literary texts. The authors demonstrate how visualizing patterns in language use can reveal insights into literary style and authorship.

The second part of this book focuses on linguistic corpora construction. A corpus is a collection of texts for linguistic analysis. Corpus-based research has become increasingly

popular in linguistics because it allows researchers to analyze large amounts of data. This section includes contributions to corpus annotation, corpus design, and corpus-based language teaching. Another chapter discusses how corpus-based research can study language change over time. The authors demonstrate how analyzing changes in word frequency over time can reveal insights into linguistic evolution. While another contribution discusses how corpus-based language teaching can improve second language acquisition. The authors argue that exposing learners to authentic language use through corpora can develop more naturalistic language skills.

This book's third part explores projects with corpus analysis and natural language processing as the main areas of interest. Computational linguistics studies how computers can process natural language data, while natural language processing is the application of computational techniques to analyze and understand human language. This section includes contributions to machine translation, named entity recognition, and text classification. For example, one of the chapter studies how machine learning can improve sentiment analysis. The authors demonstrate how training a machine learning algorithm on a large corpus of annotated data can improve its ability to classify sentiment accurately in new texts. Other scholars made substantial advancements in how named entity recognition can extract information. This book overviews current Digital Humanities, Corpus, and Language Technologies research. It demonstrates how these fields can enhance research across various disciplines. The conversation is now open. The data revolution has already changed everything. How would this inform the Humanities of tomorrow? This very question remains open, and yet its overwhelming and unattainable challenge is one of the most scientific quests that our generation must provide an answer to. The pages of this book are a modest but robust effort to create and find new paths.

Prof. dr. Pablo Valdivia

Academic Director Netherlands Research School for Literary Studies (OSL)

Chair-Full Professor European Culture and Literature – University of Groningen