STEVEN E. JONES, *ROBERTO BUSA, S. J., AND THE EMERGENCE OF HUMANITIES COMPUTING: THE PRIEST AND THE PUNCHED CARDS*, ROUTLEDGE, LONDON – NEW YORK 2016, 196 PP., ISBN: 9781315643618.

MARGHERITA FANTOLI
KU LEUVEN

ɷ

The book entitled *Roberto Busa, S.J., and the Emergence of Humanities Computing: The Priest and the Punched Cards* was first conceptualized during a conference on Digital Humanities (DH), more precisely in the drinks line (p. 23). The work of Jesuit priest Father Roberto Busa (1913–2011) is in fact considered a seminal moment in the history of the field that is now known as DH. Busa was the driving force behind the creation and publication of the 56-volume *Index Thomisticus*, [1] which was accomplished using IBM punched card technology and, later on, electronic digital equipment. As a result, Busa is considered a pioneer in the field of Humanities Computing, which is, in turn, considered a precursor to DH. However, the book challenges the notion of linear progress between the two fields and communities. The author, who is well-versed in alternative beginnings of the field,[2] brings a unique perspective to the legacy of Busa, and the thorough historical reconstruction based on archival material makes the book highly engaging for those currently involved in the DH field.

The book comprises five chapters, each taking as a starting point a historical event. An extensive introduction provides the readers with an overview of the book's scope, the materials used, and the author's methodological choices. The monograph's scope is clearly defined as an analysis of the first decade of Busa's project, specifically, the historical context in which the project was shaped. With the foundational value of Busa's work in mind, the book aims to « complicate the myth with history » (p. 3). The chronological boundaries of the work are defined by the year 1949, when Busa first met with IBM CEO Thomas J. Watson Sr., and the end of the 1950s, when the *Index Thomisticus* was well underway and the CAAL

---

[1] ROBERTO BUSA, « The Annals of Humanities Computing: The Index Thomisticus », *Computers and the Humanities*, 14/2 (1980), p. 83–90. The project is available here <https://www.corpusthomisticum. org/it/index.age;jsessionid=948AD0B0FAC46BEF8A6F87E3E08BEF5D> (Accessed January 2023).

[2] STEVEN E. JONES, *The Emergence of the Digital Humanities*, Routledge, New York 2013, <https://doi.org/10.4324/9780203093085> (Accessed in January 2023).

(Centro per L'Automazione dell'Analisi Letteraria) engaged with the Dead Sea Scrolls project.

Most of the material used by Jones was sourced from the Busa Archive at the Catholic University of the Sacred Heart in Milan, accessed with the support of Marco Passarotti, director of the CIRCSE, the Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione.[3] Jones also received materials from the IBM Corporation Archives, Fordham University, and Columbia University. Jones acknowledges that these materials come from self-curated or corporate collections, providing only a partial aspect of reality (p. 7).

From a methodological standpoint, the author incorporates the perspective of media archaeology (referring specifically to Zielinksi and Emerson),[4] which examines the historical phenomena that facilitate the emergence of specific technologies, and platform studies, which focus on the interactions between different components of a technology, including hardware and human agents. In this way, the technical aspects of Busa's work, specifically the punched card workflow, are placed within a broader historical framework.

The first Chapter of the book (« Priest walks into the CEO's office The Meeting between Father Roberto Busa, S.J. and Thomas J. Watson, Sr. of IBM, November 1949 ») takes the reader back to the years following World War II. The Jesuit priest, Busa, traveled to North America with funding provided by a wealthy Italian family, under the condition that he would chaperon their 15-year-old son on his first overseas trip. This travel marks the starting point of the *Index Thomisticus* project. Busa set out with the intention of finding the appropriate machinery to help him create a lemmatized concordance of the works of Thomas Aquinas, but he also actively networked within the scholarly community of Thomists and classical philologists in North America. After being directed from the Library of Congress to MIT and finally to IBM (p. 28), in November 1949, Busa « walked into the CEO's office », where the CEO was Thomas J. Watson. In Busa's account of the meeting,[5] he showed Watson the IBM slogan (« The difficult we do right away; the impossible takes a little longer ») to convince him to give the project a chance. Without underestimating Busa's evident ability to convince people and institutions (cf. Chapter 4), Jones examines the unique political, technological, and historical context that likely influenced Watson Sr.'s decision, particularly IBM's efforts to

---

[3] Readers interested in Father Busa's texts can find part of the material edited and translated in JULIANNE NYHAN, MARCO PASSAROTTI (eds.), *One Origin of Digital Humanities: Fr Roberto Busa in His Own Words*, Springer International Publishing, Cham 2019, <https://doi.org/10.1007/978-3-030-18313-4> (Accessed in January 2023).

[4] Referring specifically to Zielinksi and Emerson, cf. SIEGFRIED ZIELINSKI, *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*, MIT Press, Cambridge, MA 2008 (Electronic Culture: History, Theory, and Practice), and LORI EMERSON, *Reading Writing Interfaces: From the Digital to the Bookbound*, University of Minnesota Press, Minneapolis 2014 (Electronic Mediations).

[5] BUSA, « The Annals of Humanities Computing: The Index Thomisticus », p. 84.

rebuild its network of relations with Italy. Busa was subsequently directed towards the newly-founded IBM World Trade Corporation, under the direction of Watson's second son, Arthur K. ('Dick') Watson. There, Busa's primary point of contact became Paul Tasman, with whom he collaborated for nearly 30 years. Jones introduces in this chapter the concept of the « adjacent possible » (p. 38).[6] The idea is that Tasman's and Busa's approach was guided by the selection of what was easily available in terms of IBM resources and technologies. However, in assessing these 'adjacent possibilities', Busa's vision for the ultimate shape and goal of the project emerges. Jones provides an example by discussing the decision to discard the Microfilm Rapid Selector, which Busa saw at the Washington Department of Agriculture and was inspired by Vannevar Bush's memex (p. 39-41). Busa preferred the punched card technology because it made printing the results more accessible. However, the two technologies ultimately achieved different processing of the text. The Rapid Selector applied, as Jones puts it, « a kind of 'standoff' metadata layer to whole documents », whereas the punched-card system « atomized » the Latin text of St. Thomas (p. 42).[7] This atomization of text later stimulated Busa's reflection on the methodological possibilities disclosed by automation (Chapter 5).

In Chapter two (« Oracle on 57th street The IBM SSEC Large-scale Calculator, Representations of Computing, and the Role of the Adjacent Possible, 1948–1952 »), the author shifts the focus from the CEO's office to the ground-floor of the IBM headquarters, where the IBM SSEC (Selective Sequence Electronic Calculator) was publicly displayed between 1948 and 1952. The SSEC, which was designed for processing numerical data, was not utilized by Tasman and Busa in their work. However, the proximity of this powerful technology to their own work allows the author to reflect on the relationship between IBM's corporate interests and priorities and the field of Humanities Computing, as well as the social implications of the technologies developed. The choice to publicly display the SSEC was a clear marketing strategy, which also served to highlight the 'human' element of even the most complex machinery. Additionally, IBM had a history of carrying out projects with social utility, even when they were not economically lucrative. Furthermore, Tasman's descriptions of Busa's experiments seem to anticipate the later developments in the fields of Information Retrieval and Natural Language Processing (NLP) (p. 61). IBM had thus a clear interest in the partnership with Busa, both for marketing purposes and for the potential to open new markets. While the collaboration demonstrates that the interests of humanists, institutions, and companies can align, the chapter effectively illustrates the complexity of this

---

6   The concept is borrowed from Steven Johnson, *How We Got To Now: Six Innovations That Made the Modern World*, Riverhead Books, New York 2014.

7   Here, Jones refers to the words of Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism*, University of Illinois Press, Urbana 2011 (Topics in the Digital Humanities).

relationship. The author primarily presents the projects, declarations, and slogans used by IBM to link the machines and the manifestations of human society and intellect. To provide an example, the distance between the priorities of 'humanity scholars' and IBMers is evident in a sentence found on the IBM website, quoted by Jones (p. 75), which states,

> IBMers have been using cultural projects to stretch the boundaries of technologies for generations. In the process, they have made it possible for scholars, museums, libraries, and governments to make their work accessible and newly understandable to people all over the world.[8]

Humanities Computing is presented here as a side effect of IBM's pursuit of technological progress. The final part of the chapter recounts the public dedication of the SSEC in 1948, and through an examination of the speeches given at the event, highlights once again how IBM representatives emphasized the role of humans in the creation and use of the SSEC.

In chapter three (« The Mother of all humanities computing demos The First Public Demo of Busa's and Tasman's Punched-Card Method of 'Literary Data Processing', June 27, 1952 »), the narrative progresses three years forward to 1952, where Busa conducts a public demonstration of the punched-card process for constructing the *Index Thomisticus* at the IBM headquarters. Jones examines how the project developed between 1949 and 1952. While Chapter two centres around technology and its relationship with human society, Chapter three highlights the importance of the human network established by Busa to secure IBM's long-term support. Busa utilized his connections within the Jesuit order, such as Cardinal Francis Spellman, and in the Canadian and North American academic world to send recommendation letters to IBM. The cautious endorsement of the project by scholars like Werner Jaeger (Harvard), who declared their lack of understanding of the technical aspects of the work (p. 87), immediately resonates with the practitioners of DH, used to the difficult task of mediating within different communities. Busa announced the project and described the methodology in two publications, a short announcement in *Speculum* (1950)[9] and the *Varia Specimina Concordantiarum* (1951), [10] which is a 180-page publication that details the methodology and provides sample data. The demonstration brought together a

---

8 <https://www.ibm.com/ibm/history/ibm100/us/en/icons/preservation/> (Accessed in January 2023).

9 ROBERTO BUSA, « Announcements: Complete Index Verborum of Works of St. Thomas », *Speculum*, 25/3 (1950), p. 424–426.

10 ROBERTO BUSA, *Sancti Thomae Aquinatis Hymnorum Ritualium Varia Specimina Concordantiarum. A First Example of Word Index Automatically Compiled and Printed by IBM Punched Card Machines*, Fratelli Bocca, Milano 1951 (Archivum Philosophicum Aloisianum. A cura della Facoltà di Filosofia dell'Istituto Aloisianum S.J. Serie II, n. 7).

diverse group of people including IBMers, representatives of the Catholic Church, academics, and representatives of scholarly organizations. The demonstration itself concretely showed the workflow leading to the creation of the *Index*. Jones also discusses how Busa always kept an eye on the « adjacent possible » by networking with scholars engaged in similar projects but using different technologies. For example, he maintained an active correspondence with Reverend John W. Ellison, who used the UNIVAC (Universal Automatic Computer) produced by Remington Rand and based on magnetic tapes, to build a concordance of the Revised version of the Bible. The Chapter illustrates the frenetic activity and foresight with which Busa established an infrastructure that brought together diverse profiles in order to provide stability and a long-term perspective to his vision. This effort was crucial for the emergence of humanities computing.

In Chapter four (« Centers of activity The Founding of CAAL, the First Literary Data Processing Center in Gallarate, Italy, 1954–1956 »), the author presents the story of the foundation of the CAAL (Centro per L'Automazione dell'Analisi Letteraria, or Linguistica) in 1956, which was a Humanities Computing center directed by Busa in Gallarate, Italy. The Chapter specifically deals with the establishment of an institutional infrastructure to efficiently carry out humanities computing work. The setting is now firmly in Italy and the focus shifts from the well-established IBM Corporation to the bustling reality of Northern Italian industrialization, where Busa quickly becomes a key connector. The goal of the CAAL was to continue the work of the *Index Thomisticus*, while also training selected and hard-working keypunch operators (p. 119). The activity of the CAAL was closely linked to the Jesuit Aloisianum College, to which Busa was affiliated, the local textile industry, which provided funding and space for the CAAL, and the nearby Euratom (European Atomic Energy Community) in Ispra, with which Busa established a productive exchange of data, expertise and financial support. The author, in particular, describes the process by which Busa connected the automatic-translation project of Georgetown University to the Euratom, while providing the atomic centre with machine-readable Russian texts that dealt with Cold War-related topics (p. 110–112): scholars interested in the complex relationships between academia, governments, and private corporations will read these pages with great interest. Internally, in 1961, the CAAL was divided into two sections: one for scholars and researchers, whose members played a key role in promoting the study of computational linguistics in Italy, and one for technical operators. The interaction between these two groups remains a characteristic of today's DH centres, despite the continuous efforts to create bridging profiles. The chapter finally examines the use of the terms 'Literary' and 'Linguistic' to describe the work carried out at the CAAL and in the joint projects between IBM and Busa. According to Jones, IBM's insistence on the term 'Literary' fit their purpose of 'humanizing' the computing machine, but in reality, Busa mainly targeted

linguistic phenomena « usually at a level well below the semantic » (p. 127). IBM's 'humanizing' ambitions are also evident in the fact that Busa's project was presented at the IBM Pavilion at the Expo 58 (the 1958 Brussels World's Fair at the Atomium in Brussels) in order to reinforce the main narrative of the Expo, which showcased how technology could support a peaceful and human-centred society.

In the final chapter (« Computing philology The Dead Sea Scrolls Project, 'A Quality Leap and New Dimensions', 1957–1959 »), the author examines the Dead Sea Scrolls project. Following the discovery of a large number of scrolls by a shepherd in 1947, the story of their controversial publication began. In 1957, the Vatican Library announced its intention to apply the method developed by Busa to create a concordance of the published volumes (p. 142). The work was carried out between 1957 and 1960, but ultimately did not result in any publications, and may have contributed to a nervous breakdown of Busa (p. 164). Through the examination of the Dead Sea Scrolls project, the author investigates Busa's vision of what he referred to as the 'new philology'. Through this chapter, the reader can appreciate how the ten-year collaboration with IBM increased Busa's interest in perspectives opened for the human mind via the use of machinery: discovering new patterns in language and accessing the roots of authors' thoughts and expressions.[11] Even though the concordance was not ultimately published, the chapter highlights the challenges and lessons still relevant for contemporary DH: the amount of work required and the difficulty of recruiting human lemmatizers with the necessary expertise (such as the ability to work with Hebrew) led to the discontinuation of the project (p. 162).

Throughout the five chapters, Jones, consistently examines the role of women during the decade under examination, from the feminine symbolism assigned to IBM machinery in media representations to the figure of the keypunching supervisor in Gallarate, Livia Canestraro (p. 124). In this regard, the author partially draws on the work of Julianne Nyhan and Melissa Terras[12] to present a picture of a biased reality, where women, despite being essential elements of the technical workflow, were not given proper recognition or were hardly ever assigned supervisory roles, generally reserved exclusively for men.

As the readers progress through the book, they are taken on a journey through a diverse array of images, texts, voices, and places. Additional images can also be

---

[11] ROBERTO BUSA, « L'analisi linguistica nell'evoluzione mondiale dei mezzi d'informazione », in SERGIO MORANDO (ed.), *Almanacco Letterario Bompiani*, Bompiani, Milano 1962, p. 103–107. The translation by Philip Barras is available in NYHAN, PASSAROTTI (eds.), *One Origin of Digital Humanities*, p. 75–86.

[12] See MELISSA TERRAS, JULIANNE NYHAN, « Chapter 6. Father Busa's Female Punch Card Operatives », in MATTHEW K. GOLD, LAUREN F. KLEIN (eds.), *Debates in the Digital Humanities*, University of Minnesota Press, Minneapolis – London 2016, p. 60–65.

found online.[13] Each idea presented is grounded in a documented piece of information or in the analysis of a specific event. The author effectively debunks the myth surrounding Busa and presents a dynamic image of him as an integral part of an evolving and expanding international network. Anyone familiar with the challenges of DH will be struck by the contemporary relevance of Busa's reflections and strategic decisions: he was able to secure funding, balance academic, societal, and private interests, take advantage of evolving technology, establish an efficient network of support, and ultimately adapt the scholarly methodology to the possibilities opened up by the technical implementation of the workflow. Beyond the striking modernity of Busa's approach, the points of attention raised by the author resonate with readers who are witnessing the evolution of the DH community in the last few years. The tension that partially arises between the legacy of Computational Humanities and the Digital Humanities community, as well as the challenging effort to bridge NLP with literary and content analysis of texts, are noteworthy aspects to consider.

As to the first aspect, in July 2019 Folgert Karsdorp, a researcher at the Meertens Institute of the Royal Netherlands Academy of Arts and Sciences (Amsterdam, the Netherlands), tweeted « I'm thinking about developing a workshop/event/journal/community for computational research in the humanities that doesn't exclude people with profound computational skills and knowledge. Who's in? ». Given the number of positive responses, the community was started, in the form of an online forum [14] and an annual conference (Computational Humanities Research) now in its third edition.[15] An interesting post (2019) by Leah Henrickson,[16] Lecturer in Digital Media at the University of Leeds, explores the reasons behind Karsdorp's initiative and highlights the need for a space where the relationship between computer science and 'humanities mining' can be further developed. Discussions about the transition from Humanities Computing to DH have been extensive, not to mention the difficulties in defining the field of DH itself. Leah Henrickson argues that the emergence of the Computational Humanities community is not simply a terminological problem, but rather a return to the original spirit of the Humanities Computing community: the multiple souls (and origins) of the DH domain have as a consequence the need for carving out a dedicated space for the interested scholars.[17] Jones' book is

---

[13] <https://priestandpunchedcards.tumblr.com/> (Accessed in January 2023).

[14] <https://discourse.computational-humanities-research.org/> (Accessed in January 2023).

[15] <https://2022.computational-humanities-research.org/> (Accessed in January 2023).

[16] <https://www.3ammagazine.com/3am/humanities-computing-digital-humanities-and-computational-humanities-whats-in-a-name/> (Accessed in January 2023).

[17] See also Frederik Elwert's blogpost: <https://belter.hypotheses.org/64>, who retraces the dynamics behind the 'sudden' emergence of Computational Humanities (Accessed in January 2023).

certainly a celebration of this hands-on, data-processing-oriented spirit. Jones also notes Busa's interest in the « sheer materiality » (p. 93) of the activity of the scholar who is constantly working with the machinery and physically moving packages of punched cards. This can be seen as a parallel to contemporary Computational Humanities scholars who are constantly experimenting with new ways to create, manage and analyze humanities-related data. Even the networking aspect of Busa's enterprise is mirrored in the relevant efforts by the Computational Humanities scholars to create an engaging community, both online and offline. Jones appears thus to anticipate the need for some members of the DH community to bring the idea of 'computing' back to the centre of the research agenda. However, it is important to note that while Humanities Computing was heavily textual-based, mostly for technical reasons, the field of Computational Humanities now encompasses a broad range of disciplines, as it is visible in the programs of the conferences.

As per the second aspect, throughout the book, Jones delves into the nuanced relationship between the literary and linguistic aspects of Busa's work. The primary focus of the joint venture between IBM and Busa was the processing of linguistic data, but there is a persistent literary aspect that can be seen, for instance, in the early expansion of the CAAL and in Busa's desire to use the concordance to gain a deeper understanding of Thomas' language and philosophy. It is clear that Busa's primary interest was not solely in studying linguistic properties, but rather in utilizing linguistic data to gain a more comprehensive understanding of the text and the author's intentions. However, the process of creating a computing process and then completing the index acquired a central importance in Busa's trajectory. The tension between the effort required for linguistic processing, on the one hand, and the exploitation of data for interpretive work, on the other hand, as described in Jones' book, is similar to the ongoing relationship between NLP, Computational Linguistics and DH. Scholars such as Nan Z. Da [18] have argued that Computational Literary Studies have significant methodological flaws (due to the data and the statistical tools used) which can lead to uninteresting or incorrect results. The claim has sparked significant debate within the field.[19] When it comes to NLP, Barbara McGillivray, Thierry Poibeau and Pablo Ruiz Fabo [20] state that both the NLP and DH community place a central emphasis on texts, but their differing priorities – NLP focusing on advancing

---

[18] NAN Z. DA, « The Computational Case against Computational Literary Studies», *Critical Inquiry*, 45/3 (2019), p. 601–639.

[19] Some of them are listed here: <https://demo.hedgedoc.org/s/rJ_YoK_cH> (Accessed in January 2023).

[20] BARBARA MCGILLIVRAY, THIERRY POIBEAU, PABLO RUIZ FABO, « Digital Humanities and Natural Language Processing: Je t'aime… Moi Non Plus », *Digital Humanities Quarterly* 14/2 (2020), <http://www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html> (Accessed in January 2023).

techniques, DH on investigating humanities phenomena – often result in a lack of communication and missed opportunities for mutual enrichment. However, the authors highlight the potential benefits that can come from better interaction between the two communities, with NLP techniques having the potential to greatly impact DH studies, and DH corpora and questions providing challenging opportunities for the NLP community. Simon Hengchen, Nina Tahmasebi, Dominik Schlechtweg, and Haim Dubossarsky also note similar difficulties in the specific area of semantic change.[21] Furthermore, the current explosion of public interest in language models and their applications, such as ChatGPT,[22] raises the question of the transferability of these techniques to literary studies. [23] Jones' book illustrates that the connection between technical data processing and text interpretation was, even at the beginning of humanities computing, a critical issue. Overall, the book serves as a reminder of the importance of considering both technical and interpretive aspects when engaging with DH studies.

In conclusion, the book, despite focusing on events from 75 years ago, provides a refreshing and highly relevant read: I warmly recommend it to scholars within the DH constellation and to those wishing to better seize the historical roots of the current debates.

---

[21] SIMON HENGCHEN, NINA TAHMASEBI, DOMINIK SCHLECHTWEG, HAIM DUBOSSARSKY. « Challenges for Computational Lexical Semantic Change », in NINA TAHMASEBI, LARS BORIN, ADAM JATOWT, YANG XU, SIMON HENGCHEN (eds.), *Computational Approaches to Semantic Change*, Language Science Press, Berlin 2021 (Language Variation 6), p. 341–372.

[22] It is worth mentioning, given the topic of this review, that GPT was used by the author to improve the academic English of the text.

[23] See Ted's Underwood discussion <https://tedunderwood.com/2019/07/15/do-humanists-need-bert/> (Accessed in January 2023).