**JONES — Punched-Card**

<0 images>

**Punched-Card Humanities: Roberto Busa and IBM in Historical Context**

Jones, S.

It’s the canonical founding story of the digital humanities: In November 1949, the Italian Jesuit scholar Father Roberto Busa came to IBM headquarters at 590 Madison Avenue in New York to meet with the founder and head of the company, Thomas J. Watson Sr. He was seeking technical and financial support for the project of mechanizing the building of a massive lemmatized concordance to the works of St. Thomas Aquinas. Drawing on Father Busa’s own papers, recently accessioned in Milan, as well as IBM archives and other sources, I aim to uncover the historical complexities behind the founding story, the history behind the myth. In this paper, based on a book in progress, I’ll look at a few details of that 1949 meeting between Father Busa and Watson using the metaphor of the exploded-view diagram—like those used by engineers, descended from examples in the notebooks of Leonardo, for example—suspending and zooming in on constituent details of the meeting and the meanings each detail invokes, in order to understand the story as part of a larger history of computing in the 1940s and 1950s, and as part of the prehistory of the digital humanities.

In a paper delivered at DH 2014, Geoffrey Rockwell and Stefan Sinclair briefly consider Father Busa’s work with IBM, along with other examples of ‘the period of technology development around mainframe and personal computer text analysis tools, that has largely been forgotten with the advent of the web’ (2014). They advocate a media-archaeology approach (they cite Zieliniski, 2008; I’d add Parikka, 2012) as a way to begin to ‘understand how differently data entry, output and interaction were thought through’ in the mainframe era (2014). I share those goals. In my case, I’ll also draw on the related approach known as platform studies, examining specific calculating and computing platforms available to Father Busa and his collaborator at IBM, Paul Tasman—including the ‘bootstrapped’ systems they configured for punched-card processing, which soon included early tape-drive stored-program computers. The affordances of these platforms led to the inclusion in Father Busa’s research agenda on the *Index Thomisticus* of similar work on the newly discovered Dead Sea Scrolls in the later 1950s. I’ll look at relationships between the layers in each platform: hardware, software, human agents, history, and culture. My central question is how specific technologies afford and constrain cultural practices, including the academic research agendas of humanities computing and, later, digital humanities.

My point of departure is a seemingly insignificant detail: a poster Father Busa says he carried into the CEO’s office that read, ‘The difficult we do right away; the impossible takes a little longer’. In his telling, the poster becomes a rhetorical device for persuading IBM to collaborate. I show that in the 1940s, at least for Watson, it would also have been closely associated with the SeaBees, the U.S. Navy Construction Brigade, and its famous ‘can-do’ philosophy, and, thus, with the wide context of the postwar period of reconstruction and of the role of technology in the new era, including IBM’s own complicated history in Europe. Just months before the meeting with Father Busa, IBM had created its World Trade Corporation, and, significantly, it was to this international subsidiary and the technical head of it, Paul Tasman, that Father Busa’s project was assigned. The collaboration belongs in that wider historical context, as well as the more specific context of the particular punched-card machines Busa and Tasman went on to use for the initial experiments in machine-generated indexes and general ‘language engineering’, the results of which were first published in Busa (1951). I’ll look at how the punched cards continued to be used for years, even in hybrid (and transatlantic) workflow systems that, at one point, included punching, sorting, and collating the cards in Italy, then bringing them to New York for transfer to and sequential processing on the magnetic tapes of the room-sized IBM 705. (Publicity photos from the era show Father Busa sitting at the control console of the 705, its tall tape drives in the background.)

I’ll also consider one example of an ‘adjacent possible’ platform (Johnson, 2014, 156), a road not taken, as it were: the SSEC, the first publicly recognized large-scale calculator (and about the status of which as a stored-program *computer* there has been a history of debate). Father Busa would have had to walk past this big machine on his way to the fateful meeting, a kind of pre-mainframe, operating out in public, on display in the custom-designed showroom on the street level of 590 Madison Avenue, on the corner of 57th Street, and he would perhaps have seen the plaque mounted there with Watson’s words, dedicating it to science, education, and government, and the exploration of ‘the consequences of man’s thought to the outermost reaches of time, space, and physical conditions’. The machine was in newsreels, a *Vogue* magazine spread, newspapers (one cross-marketing ad with Shell Oil included a cartoon depicting it as a colossal female ‘Oracle on 57th Street’), and was even used in an early cold-war Hollywood film noir, which cast IBM engineers and operators as extras, all during the years Father Busa was demonstrating his system for linguistic analysis using the available punched-card machines. The example of the SSEC likely led to Busa’s interest within a few years in using the 705 (in a line that replaced the SSEC in IBM’s public showroom starting in 1952, just when Busa and Tasman were demonstrating their techniques at IBM). Despite the movies and other representations of the popular imagination, fueled by publicity campaigns, the ‘questions’ posed to the ‘oracles’ that were early 1950s computers were, of course, in the form of mathematical calculations, and the data to be processed was for the most part numbers, not natural language. But according to Tasman, the development of Information Retrieval at IBM, including the standard KWIC (Key Word In Context) protocol, grew out of Father Busa’s early experiments (1968).

The whole campaign around the SSEC, in terms of its mode of publicity and its computing aims, bears a striking resemblance to the most recent IBM campaign this past season (Fall 2014), at a different showroom, down at Astor Place in New York, with another personified machine, but this time one that *does* answer natural language questions by recourse to very deep data to which it applies artificial intelligence in the form of cognitive computing. Watson, named for the founder and first CEO, is in part descended from Father Busa’s project, started in that CEO’s office, and on the kind of research it represented, its processing of language instead of numbers, and eventually (and perhaps at first accidentally), its focus on viewing a text-base in the way a database would later be viewed—as a store of information to be mined, analyzed, retrieved effectively, and even rearranged algorithmically in order to reveal patterns or answer new questions. At mid-century, the SSEC may have been (literally) the poster-child for these aspirations for computing, but it was the ‘humanistic’ work with punched-card machinery, like the experiments of Father Busa and Paul Tasman, trying at first just to sort words more efficiently, only later learning they could engage in more sophisticated text analytics, that marked a significant swerve that would eventually lead into these larger aspirations—not to mention to what came to be known as the digital humanities. It’s in that more complicated sense that the prehistory of the digital humanities, at least one prehistory, is punched-card humanities.

**References**

**Bogost, I. and Montfort, N.** (n.d.). Platform Studies. http://platformstudies.com.

**Busa, R.** (1951). *Varia Specimina Concordiatarum: A First Example of Word Index Automatically Compiled and Printed by IBM Punched Card Machines.* Fratelli Bocca, Milan.

**Busa, R.** (1980). The Annals of Humanities Computing: The *Index Thomisticus. Computers and the Humanities,* **14**(2): 83–90.

**Johnson, S.** (2014). *How We Got to Now: Six Innovations That Made the Modern World.* Riverhead Books, New York.

**Parikka, J.** (2012). *What Is Media Archaeology?* Polity, Cambridge, UK.

**Rockwell, G. and Sinclair, S.** (2014). Past Analytical: Towards an Archaeology of Text Analysis Tools. Paper delivered at *DH 2014*, Lausanne, http://dharchive.org/ paper/DH2014/Paper-778.xml.

**Tasman, P. (**1957). Literary Data Processing. *IBM Journal of Research and Development,* (**1)**3: 249–56.

**Tasman, P.** (1958). *Indexing the Dead Sea Scrolls, by Electronic Literary Data Processing Methods.* IBM World Trade Corporation, New York.

**Tasman, P.** (1968).Oral History. IBM Archives, Somers, NY, accessed 2014.

**Zieliniski, S.** (2008). *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means.* MIT Press, Cambridge, MA.