Sober Hopes

Designing an Introductory Course for Computational Research in Hebrew / Arabic Texts[1]

Itay Marienberg-Milikovsky, BGU Literary Lab / Department of Hebrew Literature

Ben-Gurion University of the Negev, Beer-Sheva, Israel

[A]

In his introduction to the first edition of *Beginning Theory* (1995), a relatively influential text book in the field of literary theory, Peter Barry asked why we needed a book like this at all, coming as it did a decade or so after the high points in this field, and when other scholars had already published books and articles with titles such as *After Theory* (Thomas Docherty, 1990) or *Post Theory (*Nicolas Tredell, 1993). "As such titles suggest, the 'moment of theory' has probably passed", Barry writes, "so why another 'primer' of theory so late in the day?".[2] Twenty five years later we know that the story of literary theory is far from coming to its end, but Barry's argument seems especially relevant:

> "The simple answer is that after the moment of theory there comes, inevitably, the 'hour' of theory, when it ceases to be the exclusive concern of a dedicated minority and enters the intellectual bloodstream as a taken-for-granted aspect of the curriculum. At this stage the

---

* Email: marienbe@post.bgu.ac.il. Website: https://www.lit-lab.bgu.ac.il.

[1] DH2020 conference, ~~Ottawa~~ (virtual conference, *Humanities Commons*, July 20-24).

[2] Peter Barry, *Beginning Theory: An Introduction to Literary and Cultural Theory* (Manchester University Press, 1995), p. 1.

glamour fades, the charisma is 'routinised', and it becomes the day-to-day business of quite a large number of people to learn or teach (or both) this material."

This answer may be of interest to anyone involved in computational literary studies, not only because it is a unique branch of literary theory (according to of those of us who do see the field as *theory*), one that maintains complex relationships with several other branches, but because the academic dynamics described in it appear appropriate to the present era. We are no longer in the 'moment' of the digital humanities, but in the 'hour'. The field still enjoys a certain aura, its activities are still vibrant and appealing to many, but we did not need the climate crisis or the corona crisis to know that this situation is not irreversible, and to remember that what seems very important in one moment may seem less so in another.

However, what is true of the central and hegemonic centers of academia in Europe and North America, tends to reach peripheral areas somewhat late. In such areas – and below I will use the Israeli case as a test case – the research should in a sense maintain the glamor and charisma, and the advantage of primacy, while at the same time to act as if it is 'day-to-day business'. You cannot treat the field as merely a promising start, with a brave new world waiting around the corner, because you are already aware of the limitations that researchers overseas have pointed out; and on the other hand, you cannot steamroller the limitations, as you still have to build a domain from the ground up.

In fact, despite technologically impressive beginnings, the integration of the digital humanities into Israeli academy has been somewhat sluggish when compared to other countries.[3]

---

[3] Itay Marienberg-Milikowsky, "Beyond digitization? Digital humanities and the case of Hebrew literature," *Digital Scholarship in the Humanities* 34:4 (2019), pp. 908-913.

Perhaps one of the most noticeable delays has been in the production of textbooks and instruction books in Hebrew – books whose very existence may have accelerated the field's development; as of today there are still none. However, it may be that the slow pace has positively affected the formation of a more reflective concept of digital humanities. Learning from the experience of others,[4] and studying the controversies that are still far from being resolved,[5] can lead to a more critical view of the field.

[B]

In this short presentation, I would like to demonstrate how a critical standpoint – or even an ambivalent or skeptical one – influenced the construction of a unique introductory course for computational text research, with an emphasis on literary texts.[6] Thanks to a collaboration with the Open University of Israel (led by my colleague Prof. Ophir Münz-Manor), which undertook the publishing of the first introductory book in Hebrew (forthcoming December 2020), we had a rare opportunity to design a field of knowledge from its inception, in a particular cultural setting.

---

[4] Brett D. Hirsch (ed.), *Digital Humanities Pedagogy: Practices, Principles and Politics* (OpenBook Publishers, 2012).

[5] Nan Z. Da, "The Computational Case against Computational Literary Studies," *Critical Inquiry* 45:3 (Spring 2019), pp. 601-639; Fotis Jannidis, "On the perceived complexity of literature," *Journal of Cultural Analytics* (June 17, 2019).

[6] Compare: Stéfan Sinclair and Geoffrey Rockwell, "Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies," in: Brett D. Hirsch (ed.), *Digital Humanities Pedagogy: Practices, Principles and Politics* (OpenBook Publishers, 2012), pp. 241-263; Tanya Clement, "Text Analysis, Data Mining, and Visualizations in Literary Scholarship," in: Kenneth M. Price and Ray Siemens (eds.), *Literary Studies in the Digital Age*, Modern Language Association of America, 2013; Brandon T. Locke, "Digital Humanities Pedagogy as Essential Liberal Education: A Framework for Curriculum Development," *DHQ* 11:3 (2017).

At first, we perceived our mission as breaking barriers and not creating too many new ones, and so we have invested much effort in trying to build a bridge between two different academic worlds, "Two Cultures", as C. P. Snow called it.[7] However, while composing the coursebook and editing its accompanying materials, we found ourselves assailed by doubts: digital humanities conferences during these years seemed increasingly focused on sophisticated technology; and from our perspective, as 'traditional' literary scholars, who regarded themselves in the first place as *mediators*, the field was becoming more and more a branch of computer science, and not an inter-faculty bridge. Notwithstanding its limitations, Moretti's example, which we saw as a bold integration between sharp literary theory and informed use of empirical data,[8] seemed to us to be almost a thing of the past.

[C]

In the following paragraphs I will seek to describe the guiding principles that influenced the final organization of the course, whose purpose was to address these issues in a variety of different ways.

A. **Accessibility.** The guiding principle in constructing the course was to make it accessible – both technologically and conceptually. If we see digital humanities as a

---

[7] C. P. Snow. *The Two Cultures and the Scientific Revolution* (Cambridge University Press, 1961).

[8] Franco Moretti, *Distant Reading* (Verso Books, 2013); Moretti et al., *Canon/Archive: Studies in Quantitative Formalism* (n+1 Foundation, 2017).

scale that stretches between computer science (and math) and humanities, we must admit that different types of activities in the field need to be distinguished. Many studies are at the computer science end of the scale - for example, those conducted under the umbrella of data science, whose data, quite by chance, is 'humanistic'. Many other studies, probably from the core of the digital humanities, and by the major players operating in this arena, lie somewhere in the middle, reflecting dual training (institutionally or privately). Although a common perception (which [usually] I do not disagree with) sees such groups of studies as *the* ideal, I have chosen to focus on placing the course at the other end of the scale – those students and future researchers, who will not become digital humanists, but will be the ones to incorporate a 'soft' computational approach into the mainstream of their traditional fields, in a balanced way (a task which in my humble opinion has not yet been accomplished, except in a few exceptional cases).[9] This principle was applied from the outset with the purpose of maximizing the target audience of the course.

B. **Careful selection of tools.** As part of our emphasis on accessibility, the tools chosen were all ready-made and easy-to-use, requiring no prior knowledge. In this regard it should be noted that although such tools exist and are well known, only a few studies are based on them exclusively. This is usually justified – researchers need a tool more tailored to their needs – but it gives a false impression to the novice researcher, as if

---

[9] This is somewhat similar to gender studies: despite the tremendous contribution of dedicated departments and curricula dedicated to the field, its great achievement may have been that it developed in such a way that gender aspects became an integral part of the world of many researchers who did not necessarily specialize in the field.

the ready-made tools were a kind of toy for beginners, and not something that can or should be built on beyond that. It gives a false impression, first of all, because some of these tools are extremely sophisticated, and offer diverse, fruitful, and profound analysis options. And second, because if the purpose of computational tools in the humanities is, amongst other things, to produce a *defamiliarization* of phenomena that have hitherto been examined in other ways,[10] then even the simplest tool, whatever its limitations, may serve this purpose, and be capable of driving a complex process of *thought*, which is the most desired achievement of our course.

C. **Domain focus.** One of the typical problems with the training and the image of digital humanities is their perception as a comprehensive, cross-disciplinary solution for various needs. This perception causes them to be presented too generally in the early stages of academic training (…and by academic institutions, which are looking for us to cover more or less anything related to 'humanities' and 'computers'). Consequently, the field has not been sufficiently tested by a specific discipline which would seek to place it in a particular context. I have found that when the field is defined more specifically, while one loses some of the more general application of the digital humanities, one gains a concrete context that allows one to give the material a profound critical perspective. To this end, I chose to write about the digital humanities from the perspective of my own narrow field – Hebrew literature – and thus enjoy both the natural connection to a long-standing disciplinary tradition, as well as the known

---

[10] See, for example: Adam Hammond, *Literature in the Digital Age: An Introduction* (Cambridge University Press, 2016), p. 89.

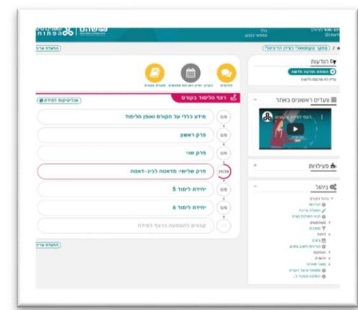challenges that any literary study poses to computational research (e.g., poetic use of language).

D. **From digital to computational.** While developing the course, and in particular from the particular literary context chosen for it, I realized how problematic the term 'digital humanities' is. Instead of focusing on what can be done *with* texts, or on the quantitative operationalizing of qualitative questions, the term 'digital' narrows the discussion to aspects of representation. Therefore, the name of the course leaves the term 'digital' only as a time mark ("*Textual Research in the Digital Age: Theory and Practice*"), and emphasizes the *research* aspect.[11] In addition, this understanding has led to a relative reduction in engagement with topics such as TEI and other formats of digital representation, without denying their importance to research.

E. **Linguistic sensitivity.** Given the peripheral languages (from a non-Western perspective) of our students – languages for which many tools are unsuited – I had to not only choose tools carefully, but also point out their limitations when it comes to analyzing texts in Semitic languages. Thus, despite the accelerated development of NLP applications in Hebrew and Arabic, some such applications still suffer from many disadvantages: they are not accessible enough, their use requires programming abilities,

---

[11] In this context, it should be noted that the movement from 'digital humanities' to 'computational humanities' is gaining momentum in various research forums, but the conceptual change reflected in it is currently seen as a development that actually strengthens the technology-and-mathematics-dependent foundations of the field, and may not attract newcomers from the core fields of the traditional humanities.

and some have so far been focused on relatively poetically neutral texts, such as newspapers and Twitter tweets. But as part of our critical approach, these shortcomings are not presented in the course solely as unfortunate technological *gaps*, but as a fact that sharpens students' awareness of the uniqueness of Semitic languages, and the challenges involved in their research. Beyond that, pointing out such technological gaps may give the impression that computational research of texts in these languages is not yet possible – and our aim is to encourage, even now, the application of alternative, albeit partial, solutions.

F. **Distant learning framework.** The Open University's proven capabilities in developing excellent distance learning aids have also contributed to making the course particularly accessible to many audiences, by creating a groundbreaking interactive learning environment. The course consists of several elements that are combined with one another: a printed book that serves as the main axis of the course; reading selected articles translated into Hebrew for the first time; and tutorials for practical experience in all the tools discussed in the course. All of these are integrated into a single virtual organizer framework, which allows for easy transition between the various parts, and,

most importantly, their critical framing. First, the entire book appears as a scanned file embedded in the course website, for the benefit of students who prefer to read from a computer; At the same time, the printed book includes QR codes that allow quick access to digital resources. In this way, the transition between theory and action becomes less a slogan than a *real* and realizable principle. And secondly, with the help of the general plot line of the course, organized through the virtual framework, there is no activity that is not accompanied by questions – from simple informative questions, to more open-ended questions for thought and discussion, sometimes, also by means of various enrichment materials, such as videos of lectures, and more. In this way, everything is *contextualized*, and there is a balance between intellectual openness and pedagogical coherence.
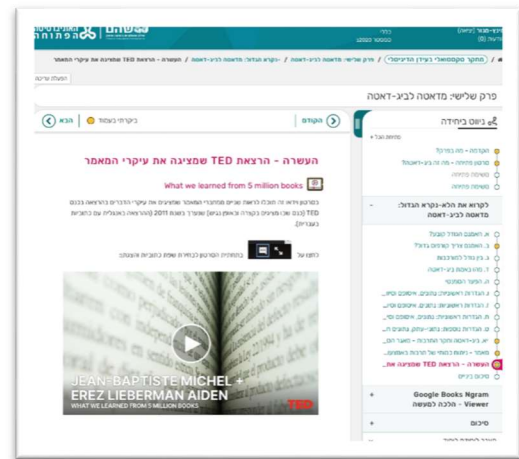
G. **Theory and practice.** As is already quite clear from all the previous sections, the premise of the course is that theory and practice must be intertwined, and balance each other out. Thus, each tool is 'wrapped' in a broad theoretical framework that highlights its contribution as well as its limitations. And at the same time, any theoretical argument is put to a practical test, doing the exact same thing but the other way around. In this sense, engaging in a specific field – literary studies – allows for a particularly concrete focus in a defined theoretical context, while at the same time, inviting the harnessing of practical tools for specific, and not too general, issues. In this way we

fully realize the necessary need that Barry pointed to in describing the transition from the 'moment' to the 'hour': from the glamour of exciting tools and the new theoretical horizons, to the hard, Sisyphean work, that results from crossing between them.

* * *

At the end of an interview Franco Moretti gave a few years ago to the *Los Angeles Review of Books*, he pointed out that all questions during the interview were asked in the future tense.[12] "There wasn't a single question that asked, "Has the digital humanities done anything?" Leave aside what it can do in the future; has it done anything? And I find this fascinating. Somehow digital humanities has managed to secure for itself this endless infancy, in which, it is always a future promise." In the field of computational research of Hebrew literature, we have almost no other choice: the little that has been done so far is simply not enough for us to be able to say – 'see what achievements we have already achieved'. On the other hand, writing the introductory book has also taught us that there is more than one way to talk about the future. Eventually, even though this is a pioneering course within Israeli academia, we decided not to hide our doubts but to incorporate them as a questioning voice throughout the book; in fact, here and there in the book, we have even alluded to the possibility that the impact of computational research on traditional research will *not* be revolutionary at all. Time will tell if this was a pedagogically correct step or not; we believe there was no other truthful choice.

---

[12] Melissa Dinsman, "The Digital in the Humanities: An Interview with Franco Moretti", *Los Angeles Review of Books* (March 2, 2016), https://lareviewofbooks.org/article/the-digital-in-the-humanities-an-interview-with-franco-moretti/.

Appendix: The course outline

Textual Research in the Digital Age: Theory and Practice

Part I: Thinking through Numbers

1. Prologue: Toward an Accessible Computational Literary Studies

2. Are There Computers in this Class? A Very Short Conceptual-Historical Introduction

3. Reading the Great Un-Read: From Data to Big-Data

4. Operationalizing: From Theory to Numbers, From Numbers to Meanings

Part II: Modelling with Computers

5. Digitizing Literature

6. Modelling: From Data Representation to Computational Research

7. Analyzing: Automatization and Calculation in the Process of Interpretation

8. Annotating: The Reader's Response

9. Mapping and Networking

10. Epilogue