

The Digital Humanities Classroom. From the Toolbox to the Mindset?

Florentina Armaselu

Luxembourg Centre for Contemporary and Digital History (C²DH), University of Luxembourg
florentina.armaselu@uni.lu, @florentina_a

1. Introduction

The paper is based on a case study, an introductory course on computational text analysis and text interpretation for students enrolled in the *Bachelor en Cultures Européennes* (BCE) programme at the University of Luxembourg. Given the hybrid character of DH education implying various subject matters and skills as well as didactic and evaluation strategies, the proposal will address, through the lens of the classroom case study, the challenge of DH teaching shifting from a toolbox-based instruction paradigm to a more reflective mindset that considers the new technologies in a broader cultural and pedagogical context.

Studies in digital pedagogy have already pointed out the need for “curriculum change” and shift toward an “inquiry-based” syllabus (Whitby, 2007: 2,3) fostered by the new technologies and focusing less on “things to know” and more on developing “strategies for learning” by stimulating imagination and engaging students in “taking responsibility of their own learning” (pp. 6, 7). By questioning the “what works” paradigm in “currently dominating educational research”, Ross (2017: 1, 2) assumes the relevance of “speculative methods” that may capture the “‘not-yetness’ of technologies” and provide a “conceptual handle for digital education approaches” and for openness to new practices and ideas while allowing “curiosity, critique, doubt, unintended consequences and emergent properties of the technologies in use”. Other studies, focusing on Digital Humanities, have highlighted different aspects to be considered while teaching in this area. Mahony and Pierazzo (2012: 7) argue that DH teaching should be relevant to the student’s study and research interests, and deal not so much with skills, though important, but with “new methodologies and ways of thinking”. Cordell (2019) recommends to cultivate a “mindset for approaching data, exploring it” and understanding “what questions computation might help answer about it”, rather than bringing the students to expertise in any particular computational method. In their article, “Beyond buttonology”, Russel and Hensley (2017) affirm that tutorial-based teaching focusing on tools should be complemented by “critical engagement with digital methodologies” and “humanities sources as data”. Other scholars (Sinclair and Rockwell, 2012; Papadopoulos and Schreiber, 2019) advise practical approaches in the classroom, such as asking students to develop their own research questions based on their own texts or applying problem/project-based learning and enabling students to become producers rather than just consumers of knowledge. The current proposal is intended to contribute to this discussion by turning a course for teaching text analysis in cultural studies into an object for digital pedagogy research.

2. Case study

The course was taught during the winter Semester 2019 – 2020, started in September 2019 and ended in February 2020 with a final examination. It was an optional, 3 ECTS¹ general course in the Arts and Media Studies module of the BCE programme. 11 undergraduate students have completed it, with profiles in different areas such as history and English literary and linguistic studies. The course introduced concepts,

¹ European Credit Transfer and Accumulation System. https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en.

methods, tools and data for computational text analysis and interpretation from six categories: corpus linguistics, parsing, named entity recognition, sentiment analysis, topic modelling and word embedding. Based on an application-oriented approach, it was intended to illustrate how these techniques can be applied in answering or formulating historical, linguistic or literary research questions. For instance, in analysing the evolution of a concept in an art history collection, main topics in the European parliament's news releases, specific vocabularies in a series of transcribed life-history interviews or the sentiment-based plot arc of a novel. At the same time, the course aimed at allowing basic understanding of the theoretical assumptions beneath the "black box" guise of the user interface and assimilation of elementary principles of programming in R and Python. Each session, except for those dedicated to the project, included theory and example presentations followed by hands-on activities. The final assessment consisted in individual projects, the students using the tools/data of their choice from those studied in class.

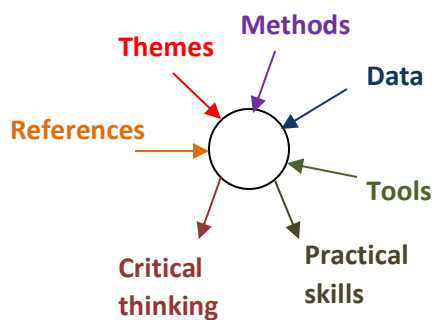


Figure 1. The DH classroom "node" metaphor

Inspired by the *node* or *crossroads* metaphor, the pedagogical approach applied in the classroom supposed the intersection of different areas of enquiry and virtual connections within the course or programme as a whole (Fig. 1). For instance, thematic (culture-oriented), theoretical (referring to text analysis methods), data-related (dataset processing and assessment), practical (hands-on activities with different tools), referential (further reading and links to relevant information to be studied individually).

Expected outcomes consisted of critical thinking and practical skills enabling the students to apply the acquired knowledge to their course project or projects from other

disciplines. These assumptions were tested via an anonymised questionnaire proposed to the students and completed at the end of the course, and included questions such as: (1) role of the text analysis tools used in their projects and in answering the research questions; (2) "added value" determined by the use of this type of analysis as compared with a more "traditional" non-digitally-based study and interpretation of texts; (3) reflections on the innovative character (if any) of the computational text analysis approach and/or its limitations, bias, etc.; (4) evaluation of the course scenario.

3. Discussion

In his model of the public image of a city, Lynch (1997: 47) defines the concept of nodes as "primarily junctions, places of a break in transportation, a crossing or convergence of paths, moments of shift from one structure to another. Or [...] simply concentrations, which gain their importance from being the condensation of some use or physical character, as a street-corner hangout or an enclosed square". This definition conveys a composite image, of foci to and from which one can travel, junctions, moments of shift, crossing of paths and concentrations, which may serve as a metaphor for the digital humanities classroom as a "node". The metaphor encompasses the hybrid character of teaching in the DH implying the concentration of various subject matters, pedagogical and evaluation strategies to be applied in the classroom, as well as the idea of circulation, exchange and connection with other disciplines inherent to the concept of node as a unit within a network. It was assumed that applying this type of hybridity as an underlying principle in the course design might foster reflection on the role and characteristic usage of digital technology in the Humanities, beyond the acquisition of technical skills and knowledge of computational methods and tools.

3.1. Lesson structure

Following this principle, the lessons were conceived as a combination of theoretical basic information about the tools and methods to be studied, together with topics and sets of data from different areas of enquiry (history, literature, philosophy, arts, etc.). These themes and datasets were used as examples and starting points for experimenting during the hands-on activities. The main goal was to provide an overview of different categories of tools and methods for computational text analysis and a thematic ground for interpretation via these tools. Six categories of digital tools (see section 2) were presented and worked with during the course sessions, which included lexical, syntactic and semantic standpoints in analysing texts. The themes were chosen to cover areas of interest for students with different backgrounds enrolled in the BCE program, and to provide a relevant context for applying what was learned in class.

The degree of heterogeneity of the lessons varied. Some dealt with the investigation of certain concepts, such as culturomics (Michel et al., 2011), hermeneutical tool (Rockwell and Sinclair, 2016), close/distant reading (Moretti, 2013; Underwood, 2019), hidden thematic structure (Blei, 2012) or Vonnegut's chalkboard shapes of stories, and with their practical application using tools for n-gram detection and visualisation, Web-based text reading and analysis, textometry exploration, topic modelling or lexicon-based sentiment analysis. Other lessons relied on theoretical subjects from various areas, such as Pennebaker's (2013) linguistic features for "predicting" honesty and deception, Kripke's (1981) theory of naming, Thompson's (2007) study of ideology in modern culture or Hazard's (1961) reflections on the crisis of the European conscience. These theoretical aspects were combined with methods and tools as for instance, keyness-based comparison of two corpora, named entity annotation and query, context-free grammar and dependency parsing, or word similarity computation and representation of meaning through vector semantics.

The datasets proposed for analysis were selected and downloaded from online sources, such as *The Linguist List Texts & Corpora*, *CLARIN Resource Families*, *Project Gutenberg* or researchers' published data, and included a variety of textual collections, from oral history interview transcriptions, parliamentary news, movie summaries and novels, to fake and legit news, children literature and academic articles in art history. The aim was to provide the students with a variety of materials - theory, tools and data references, which could help in formulating their own research questions and the development of their projects.

3.2. Assignments

The course included 14 sessions, 11 for subject presentation and hands-on activities, and 3 dedicated to the work on the individual projects in a lab-like mode. Consultation hours and extra time of access to the course computer room were also provided for experimentation and finalisation of the project reports. Four assignments composed the final grade as follows: (1) *project proposal* as a 1-page term paper that contained the intended topics, research questions and envisaged approach – tools, methods data to be used (30%); (2) *project realisation* which consisted of the actual work on the project using methods, tools and data learned in class, and a set of deliverables, such as input/output data samples, pieces of code, readme files, diagrams, etc. (30%); (3) *project presentation* including a project report, a 3-5 page written essay, and a 10 minute oral presentation during the final examination that took the form of a mini-colloquium (30%); (4) *active participation* in the course as a whole (10%). The fifth assignment, ungraded, consisted of the students' involvement in providing responses to a feedback questionnaire. Different types of documents to be submitted for the assignments were chosen (overview of the general idea, raw data and result files, essay, presentation, questionnaire answers) in order to document the process at different stages of the project and to enable reflection and a diversity of forms of expression as a learning experience.

The proposed projects varied in terms of topics and selected tools. The majority of students (6) opted for the use of Voyant, some of them (3) in combination with AntConc for methodological comparison. GATE was chosen by a smaller number (3), while MALLET (1) and RStudio and sentimentr (1) were selected by a single student, each. No project used the other remaining tools studied in class (Google Books Ngram Viewer, TSM, Stanford parser and word2vec via Python IDLE), although some of them were mentioned as possible alternatives in the initial project proposals. From the point of view of themes and datasets, the projects dealt with various subject matters at different scales, from the analysis of a small number of documents, e.g. particular US and European Commission presidential speeches, national and international treaties and conventions, a novel, to the compilation and study of larger corpora containing speech transcripts of candidates to the 2020 US presidential race, extracts from the UK Hansard parliamentary reports, customer reviews from the Facebook page of a commercial clothing company or comments scraped from a pets' dedicated discussion board. Regarding the research questions and the proposed approach, the students were interested in aspects such as comparing or drawing the evolution in time of different types of discourse (speeches, treaties, parliamentary debates) via corpus linguistics methods, applying named entity recognition techniques to assess the capacity of the tool to detect unfamiliar names of persons and places in a novel, using topic modelling to identify the main arguments in the discourse of the top 5 Democrat frontrunners in the US presidential elections, computing sentiment scores for customers reviews and evaluating the overall accuracy of the method, or creating a classification of pet-related subjects (categories, habits, needs) as reflected in the discussion forum of a pet owners community.

Although the variety of project ideas, themes, analysed data and overall results was generally rewarding, certain aspects, as discussed below and in the following section, seem to require further attention and possibly adjustments in a second iteration of the teaching experiment. Some projects showed initiative in collecting and preparing the sets of data, creativity in combining different features, tool tuning, labelling, visualising and interpreting the results, as well as awareness of the benefits and limitations related to the data size and format, tool, methodology and the applied approach itself. Other projects, though demonstrating a relevant amount of effort in testing, producing raw results and interpreting, were less effective in communicating, through the final report and presentation, how the tools and methods were actually used to get the results and support the proposed interpretation. Additional categories included projects that focused more on the description of technical details of the analysis and less on the text interpretation itself or that proposed interpretations mainly based on features that didn't go beyond lists of frequency counts, word clouds and word trends inside documents, applied to a limited amount of data. Correlating these observations with the students' feedback helped in evaluating the overall experience of the course.

3.3. Feedback

The feedback was provided via two channels, the official evaluation carried out at the Faculty level, after the end of the teaching period, and the responses to the questionnaire² proposed as a fifth assignment at the end of the course, after the final examination. While ten (out of eleven) students completed the official evaluation, only eight filled in the assignment questionnaire as well. No profile information was available in the official report. The proposed course questionnaire comprised a section for the description of the respondent profile, an anonymization code and a formal agreement for the use of the collected data for research and publication purposes. The group that filled in the course questionnaire included 2 female and 6 male students, age range 18 - 34, enrolled in the BCE programme with a main background

² The questionnaire and students' responses can be consulted as supplementary material published with the paper: Excel file *BCE-EU-105_02_CourseEvaluationQuestionnaire*.

in history (4) and English studies (4). No previous knowledge of computational text analysis tools was reported. Six and respectively two students selected the answer 1 and 2 on the Likert scale 1 to 5 (*Not at all* to *Expert*) for the self-evaluation of their general knowledge of digital tools and methods. One student indicated *Microsoft Excel* as a tool already worked with before its use in class for the course.

Responses in the positive range were provided regarding the significance of the role played by text analysis tools in answering the research questions formulated in the projects. Two answers were placed in the middle (3 points), four and respectively two answers on the right side (4 and 5 points) on the 1 to 5 scale (*Not at all significant* to *Essential*). All the respondents agreed that these tools allowed them to discover something new or formulate new questions for the studied dataset, besides their initial assumptions from the project proposal. Six out of eight also considered that we can speak of an "added value" determined by the use of this type of analysis as compared with a more "traditional" non-digitally-based study and interpretation of texts. Asked to provide details about this "added value", the respondents mentioned the possibility to process "large quantities of text in a short time", provide a "quick overview [of the] main topics of a big text", enable "objectivity and orientation for textual analysis" or allow the student to use these tools for "other classes and make [his/her] research easier".

The questionnaire included as well a section for the evaluation of the course scenario (Fig. 2). The overall assessment ranged from four answers (one/three) in the *not interesting* and neutral area to four (three/one) in the *interesting* and *very interesting* area. The theoretical materials were considered *appropriate* and *very appropriate* by six students (three/three), while two (one/one) rated them neutrally or as *not appropriate*. Five students (three/two) *agreed* or *fully agreed* that the hands-on activity provided enough background for the development of the projects, while three (two/one) answered either neutrally or *didn't agree*. Regarding the course assignments, one student evaluated them as *very difficult*, three neutrally and four (two/two) as *moderately* or *not at all difficult*. Although the number of respondents was relatively small, the answers provided a rough idea of how the pedagogical approach was perceived by the students.

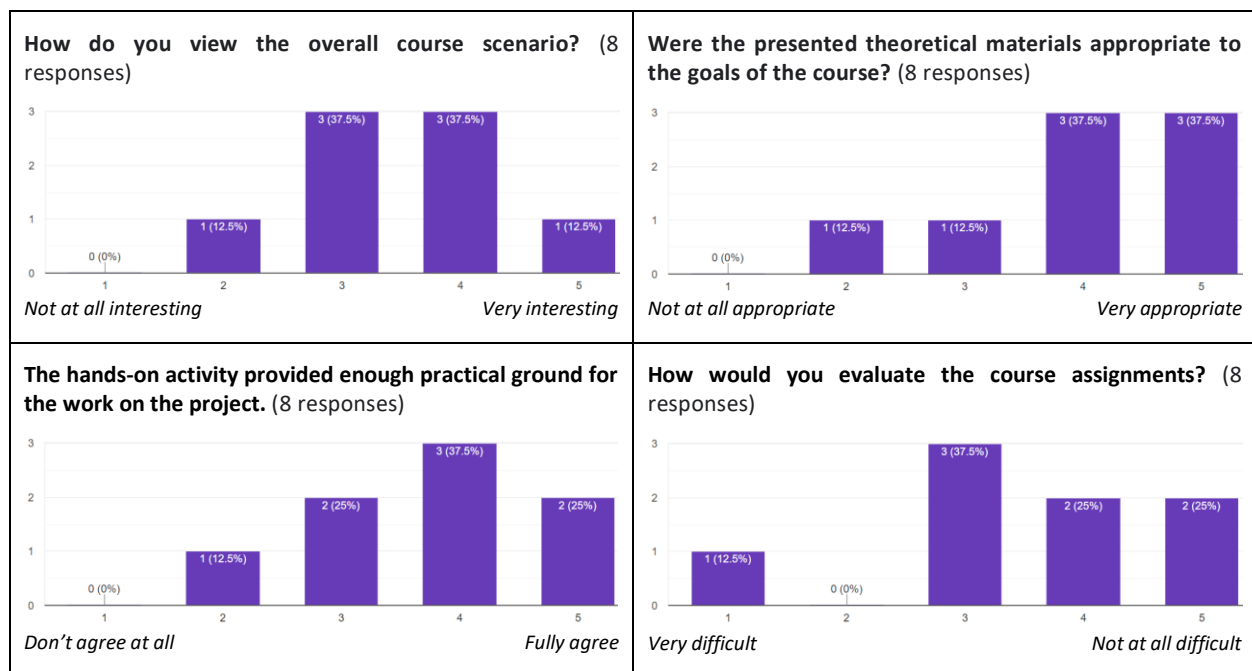


Figure 2. Evaluation of the course scenario (course questionnaire)

Further details were obtained via open questions asking the students to enumerate some of the strong and weak points in the course approach, provide suggestions for improvement and general comments about the appropriateness (or not) of the course to the BCE programme, its potential (or not) from a pedagogical perspective, its usefulness (or not) as related to their other projects, etc. Among the strong points were mentioned the possibility to see “a lot of tools that can be used”, the combination of “practical work together with theoretical work [that] enabled [them] to learn better”, the property of the course of being “practical, descriptive and cover[ing] wide areas of interest”. The weak points listed aspects such as the difficulty in following the course “without prior knowledge of informatics”, “explanations were good but can be improved”, “too much in too little time, too little time to practice during the semester”. Additional comments referred to the relevance of the course for the BCE program, especially from a “research point of view” and to the fact of teaching students to “objectively look at all data and not believe something blindly”. It was suggested as well that the course should “focus on less tools but explain them more in depth”. Similar remarks were collected from the official evaluation. On the one hand, the students assessed that the course helped them to “gain knowledge in programming tools”, “get insight into the possibilities of computer science in the humanities” and appreciated that the “students should be trained more in digital tools” and the “introduction to the different programs” can be later used for “research in text interpretations”. On the other hand, it was considered that “at some points it was hard to follow the lecturer as [not being a] computer programmer and [feeling] a bit lost with all the information” or it was suggested to focus on “less content but longer explanation of the most important” items and to “spend more time on a program”.

While the general idea of providing an overview of a variety of tools and methods for computational text analysis and the combination of theoretical and practical approaches seems to have been positively perceived by the students, the projects evaluation and the collected feedback suggest that the number of programs studied in class and the time allocated to some of them should most probably be adjusted. Apart the first 2 introductory sessions on the general topic and online data collections and the 3 lab-mode sessions for the work on the projects, 9 sessions were dedicated to the study of 9 tools and the corresponding theoretical context and datasets used in the hands-on activities. Since some of the classes, especially those dealing with more complex software or R and Python programming imply a higher degree of difficulty, it is planned to allocate more than a session to these lessons and slightly decrease the total number of proposed programs during the next iteration of the course for the winter semester 2020-2021.

4. Conclusion and future work

The paper describes a digital pedagogy and evaluation setting applied to a case study combining different approaches - theoretical, application-oriented and project-based, and underpinned by a view of the DH classroom metaphorically defined as a “node”. The metaphor encompasses the hybrid character of teaching in the DH implying the use of various categories of study materials, themes, methodologies and skills, as well as the idea of exchange and connection with other disciplines. It was assumed that this type of hybrid approach applied in the course design together with the coverage of a variety of tools to be studied in class might foster a reflective mindset that enables the students to contextualise the digital technologies within different Humanities areas, beyond the acquisition of technical skills and knowledge of computational methods and tools usually referred to as the toolbox-based education paradigm.

The case study involved a small number of students and a single iteration of the teaching experiment. Although the findings stand yet for a limited outcome and require further investigation, the evaluation of the students’ projects and the analysis of their feedback seem to suggest that a balance between the toolbox- and the mindset-based pedagogical perspective would allow a proper framing of the DH classroom. Namely, as a physical and symbolic space of training, experimenting and reflection for

independent learners able to formulate questions and answers, explore new paths and make connections, and understand technology both as a tool and an asset for continuous questioning and discovery.

References

- Anthony, L. (2019). *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp>.
- Blei, D.M. (2012). "Topic Modeling and Digital Humanities", *Journal of Digital Humanities*, Vol. 2, No. 1 Winter.
- Cordell, R. (2019). "Teaching Humanistic Data Analysis", talk delivered on 2 May 2019 as part of Gale's Digital Humanities Day at the British Library. <https://ryancordell.org/research/teachingHDA/>. GATE: <https://gate.ac.uk/>.
- Google Books Ngram Viewer. <https://books.google.com/ngrams>.
- Hazard, P. (1961). *La crise de la conscience européenne, 1680-1715*, Fayard.
- Heiden, S., Magué, J-P., Pincemin, B. (2010). *TXM : « Une plateforme logicielle open-source pour la textométrie – conception et développement »*. In Sergio Bolasco, Isabella Chiari, Luca Giuliano (Ed.), *Proc. of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010* (Vol. 2, p. 1021-1032). Edizioni Universitarie di Lettere Economia Diritto, Roma, Italy. <https://halshs.archives-ouvertes.fr/halshs-00549779/fr/>. TXM Website: <http://textometrie.ens-lyon.fr>.
- Kripke, S. (1972, 1981). *Naming and Necessity*, Blackwell Publishing.
- Lynch, K. (1997). *The Image of the City*, The MIT Press, Cambridge, Massachusetts, and London, England, 1960, 1997.
- Mahony, S. and Pierazzo, E. (2012). "Teaching Skills or Teaching Methodology?" In Hirsch, B. D. (Ed.), *Digital Humanities Pedagogy: Practices, Principles and Politics*. Open Book Publishers. Retrieved from <http://books.openedition.org/obp/1639>.
- MALLET. <http://mallet.cs.umass.edu/>.
- Michel, J-B. et al. (2011). "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science*, 14 Jan 2011, Vol. 331, Issue 6014, pp. 176-182.
- Moretti, F. (2013). *Distant reading*, Verso, London, New York.
- Papadopoulos, C. Schreibman, S. (2019). "Problem/Project Based Learning in the DH Classroom: Situated Learning, Empowerment, and Knowledge Creation", DH Benelux Conference, 11 – 13 September 2019, University of Liège, Belgium. http://2019.dhbenelux.org/wp-content/uploads/sites/13/2019/08/DH_Benelux_2019_paper_50.pdf.
- Python and Python IDLE (3.7.4). <https://www.python.org/downloads/release/python-374/>.
- Rehurek, R. (2009-2019). *Word2vec embeddings*. <https://radimrehurek.com/gensim/models/word2vec.html>.
- Rinker, T. W. (2019). *sentimentr: Dictionary based sentiment analysis that considers valence shifters*, 2019. <https://github.com/trinker/sentimentr>.
- Rockwell, G. Sinclair, S. (2016). *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, The MIT Press.
- Ross, J. (2017). "Speculative method in digital education research", *Learning, Media and Technology*, 42:2, 214-229, DOI: [10.1080/17439884.2016.1160927](https://doi.org/10.1080/17439884.2016.1160927).
- RStudio. <https://rstudio.com/>.
- Russell, J.E. and Hensley, M.K. (2017). „Beyond buttonology. Digital humanities, digital pedagogy, and the ACRL Framework”, *College & Research Libraries News*, Vol. 78, No 11. <https://crln.acrl.org/index.php/crlnews/article/view/16833/18427>.

- Sinclair, S., Rockwell, G. (2012). "Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies". In Hirsch, B. D. (Ed.), *Digital Humanities Pedagogy: Practices, Principles and Politics*. Open Book Publishers. Retrieved from <http://books.openedition.org/obp/1644>.
- Stanford parser*. <https://nlp.stanford.edu/software/lex-parser.shtml>.
- Thompson, J.B. (1992, 1994, 1996, 2006, 2007). *Ideology and Modern Culture: Critical Social Theory in the Era of Mass Communication*, Polity Press, Malden, US.
- Underwood, T. (2019). *Distant Horizons. Digital Evidence and Literary Change*, The University of Chicago Press.
- Vonnegut, K. On the Shapes of Stories. <https://www.youtube.com/watch?v=oP3c1h8v2ZQ>.
- Voyant*. <https://voyant-tools.org/>.
- Whitby, G. (2007). "Pedagogies of the 21st Century", ACEL 2007 International Conference, Sydney, Australia. <http://cctionline.org/wp-content/uploads/2014/08/pedagogy-for-the-21st-century.pdf>.