

Challenges for Digital Literacy in English Curriculum

Charles Lam and Catherine Wong

Department of English

Hang Seng Management College

{charleslam, catherinewong}@hsmc.edu.hk

Abstract

This paper presents and describes the implementation of digital literacy in English curriculum in the context of Hong Kong. We outline two major challenges of introducing digital pedagogy in humanities classrooms at the undergraduate level and discuss with specific reference to two cases the strategies we have taken to tackle them.

1 Introduction

In this paper, we discuss challenges of teaching digital literacy in the context of English literature and linguistics curriculum. As English majors, students are typically trained with analytic skills that targets literary and linguistic analyses, such as close reading skills in literature and analysing syntactic patterns in theoretical linguistics. Blending digital skills into the training of English majors and humanities students is often considered desirable or even necessary. Meanwhile the task is also challenging for both instructors and students. This paper discusses the challenges and some steps we have taken to tackle them. Some of the challenges can be generalised to different populations (e.g. difference in computer skills), some others are specific to only some populations (e.g. English being a second language to students in Hong Kong). Towards the end, a few use cases and students' projects are discussed to illustrate our outcome.

2 Goals

Situated in the English curriculum, our objectives are to enable students to apply their literary and linguistic knowledge with some level of digital literacy. The primary goal of the English curriculum is to equip students with language and analytic

skills that are highly transferable. English graduates pursue various career paths related to languages and writing, such as teachers and editors. A portion of them find positions in areas, such as banking, marketing and merchandising, often in places where their writing skills can be applied.

To achieve the general goals in the English curriculum, that is to train students with language skills that are highly transferable, we set out to enhance students' awareness in finding and articulating patterns in textual data and incorporating their knowledge in language with skills in information technology.

In our current curriculum, classes in the English major often adopts more traditional approach, which consist of lectures with slideshows and/or handouts. While this approach is not ineffective, students often report in course evaluations that the classes can be dull or boring and they therefore demand for more interactive methods. Drilling on individual topics in a unidirectional manner in lectures gives little opportunity to turn their knowledge into application. We therefore consider the hands-on approach as one that relates directly to their learning motivation and therefore a suitable tool in facilitating students' proficiency in digital literacies.

We should acknowledge that the English curriculum is internally divergent by nature. As a result, the methods and training between literary and linguistic studies are rather different too. While the training of literary studies equips their students with assertiveness and the ability to raise their opinions and make critical statement, they are relatively weak at working backwards and deliberating their cognitive process of reaching their conclusion. Conversely, linguistics-leaning students are more well-trained in quantitative methods of research, yet, they usually stop when they have finished analysing their dataset without any

attempt to make a hypothesis before or a conclusion in the end. Digital humanities create a platform for these students to enhance their subject related skills and transfer them into a cross-disciplinary context.

3 Challenges

3.1 Insufficient IT skills

Typically, students who enrol in English are strong in language and textual skills but weak in computer skills. In addition, students often attribute their lack of computer literacy to their discipline. It is common to hear from students that they are not expected to be ‘good at computer’, because they are language students. For a more concrete example, students are often proficient in using word processing software, but are unfamiliar with spreadsheets. More specifically, basic functions in spreadsheet software (such as sum and average) are considered advanced functions and need to be explicitly taught.

The lack of skills or confidence becomes a self-fulfilling prophecy that they would end up failing to acquire new skills. A related challenge is that the students are inexperienced in reading and learning from error messages, software documentations or forum discussion. We consider it necessary to raise students’ awareness of the learning strategies, often explicitly and through exercises, at earlier stages of the courses.

Despite the recent developments of various platforms, e.g. popularity of smartphones and social media, the user experience as technology consumers does not often spill over to skills or interest in using technology in a productive manner.

3.2 Inadequate foundation of formal systems in humanities

At present, we observe that the constraint encountered when teaching digital literacy in the English curriculum is also due to a lack of training of formal methods in humanities. In the traditional curriculum, there has been little requirement of incorporation and integration of formal and computational methods in humanities scholarship. Besides having insufficient IT skills, these students, at the theoretical level, are unprepared to adopt a scientific point of view in approaching their subject matters and, at the organisation level, have little training in applying quantitative methods in processing and analysing their data. More specifically, they often lack the awareness to generalise

from individual observations. That is to say, instructors need to remind the students often whenever the task requires making generalisation and a generalised solution. This generalisation skill has been discussed also in terms of ‘modeling’ (McCarty 2004) and ‘operationalizing’ (Moretti 2013) in the context of digital humanities, and ‘computational thinking’ (Wing 2006) in the context of higher education in general. A concrete example is that students were asked to separate the Chinese characters in a spreadsheet from their romanisation within the same cells to the next column. Though the Chinese characters are always in front within the individual cells, length of these Chinese varies from two to four. Students often by default separate the cells by taking out the first three characters (typical length for Chinese names) without considering the less common two- or four-character long strings. The challenge was to ensure that the students consider the rarer cases and create a general function in spreadsheet that accommodates to all the different cases.

These students, especially when in the literature classroom, fail to integrate the philosophy of science with tackling humanistic topics in the subject - they want evidence-based, clinical reasoning and habits of synthesis in reaching their conclusions and their ‘observations’ are usually impressionistic rather than objective. Conversely, these students, even when they have made good observations, they are reluctant to use statistical methods in deriving a pattern for these unclassified observations. They are not familiar with the use of machine to treat language and various forms of communications in terms of language processing. Their tendency to distrust the idea of digitalisation of language prevents them from constructing a large dataset and expanding the scope of their study. In fact, these students sometimes fail to see the importance of formalisation of the treatment of their textual materials and are unable to grasp the correlation between pattern recognition, coding of textual materials and computer simulation.

4 Implementations

4.1 Tailor-made version of corpus linguistics/ NLP

In order to achieve our aforementioned didactic goals, we need to bridge the vast discrepancy by boosting students’ IT skills from very basic level. One implementation focuses on a corpus linguistics course. While students enrolled are typically more interested in linguistics, the course also caters to literary studies. While the course covers

topics about corpus linguistics, from basic concepts like type-token ratio or collocation to existing corpora like COCA (Corpus of Contemporary American English) and F-LOB (The Freiburg-LOB Corpus of British English), the day-to-day lessons stress on its hands-on approach to let students gain first-hand experience in working with language raw data. Assuming no prior knowledge in computer or statistics, the lessons began with the aim to familiarising students with lower level operations (data pre-processing, spreadsheet), while gradually including more challenging tasks that mimic real world problems, such as reference resolution or sentiment analysis. For example, the corpus tool AntConc (Anthony, 2014) is covered in weeks 5-7, after some basics concepts are covered and a few weeks of practising in spreadsheet. Spreadsheet was introduced before a dedicated corpus software in order to tackle students' lack of IT skills or the lack of confidence, as mentioned above. In the final assessment, students are required to give a group presentation on a research topic, using the technology appropriate for their own topic. For local relevance in Hong Kong, the use of parallel corpora and Cantonese/Chinese corpora are also discussed in the course. These various topics all serve the purpose to bridge the traditional linguistics contents to practical applications of language technologies.

4.2 Reverse engineering from research questions

Another implementation to illustrate our pedagogical model involves several final-year capstone projects. Authors of these projects have a strong interest in literature and at the same time the linguistic aptitude to pursue a research in English studies. Projects are administered in the form of independent studies while students receive one-on-one supervision and support from their advisors. Students are motivated to choose a topic of interest across but not necessarily be a part of, the curriculum. Topics range from postcolonial stylistics to historical linguistics and to interdisciplinary ones. As independent studies, students are expected to be well-versed in search techniques and incorporating technology into their preliminary library research and data collection. While most projects use literary texts as the subject matter of their studies (e.g. racism in Harper Lee's *To Kill a Mockingbird*, anxiety and mood fluctuation of the narrator in Edgar Allan Poe's short stories), the research methodology and strategies tend to go beyond the traditional literary spectrum and are

usually an integration of the skills and knowledge students acquired in their linguistics training (e.g. the investigation of stylistic and social stratification within Lee's fictional world, the measuring of sentence length and syntactic complexity of Poe's narrative).

To overcome the methodological challenges mentioned in the previous section, a reverse engineering approach is used. Students are encouraged to conduct a close thematic analysis on a small segment of extracted qualitative material (e.g. an excerpt from Poe's 'Tell Tale Heart' which best illustrates the extreme emotions of the mentally unstable narrator) without the use of any computer or other digital tools. At this stage, thematic coding of the text segments is done implicitly:

- (1) Ha! would a madman have been so wise as this, And then, when my head was well in the room, I undid the lantern cautiously--oh, so cautiously --cautiously (for the hinges creaked) --I undid it just so much that a single thin ray fell upon the vulture eye. (sentence 32, word count: 50)

As shown in (1), students are required to make simple observations on the sentence length in Poe's short story and identify narrator's mood changes in relation to the use of very long or short sentences. This step is done repeatedly until an annotation framework of thematic pattern is formed (as shown in tables 1 and 2). This approach allows students to transfer their traditional literary skills to a formalisation of data without triggering their technophobia.

Word count	Sentence length	Frequency
1-3	Short	17
4-13	Moderately short	86
14-23	Moderate	38
>23	Long	25
	Total:	166

Table 1: Frequency of sentences of different lengths in 'The Tell-Tale Heart'

Emotions	Frequency
Anxiety	61
Confidence	35
Fear	29
Irritation	8
Anger	4
Certainty	3
Desperation	3
Joy	1

Table 2: Frequency of emotions being experienced by the narrator in each sentence

At a later stage of the project, digital tools and statistical methods, such as spreadsheet, data processing tool and corpus tool are introduced to optimise the research outcomes:

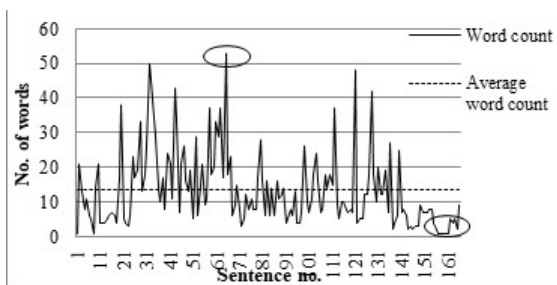


Figure 1: Sentence length in “The Tell-Tale Heart”

With a formal model derived from preliminary results acquired through qualitative analysis, students can then use machine and computer applications to measure, classify and analysis the stylistic elements observed in an objective and systematic manner. Ultimately, students can apply computer simulation to detect stylistic and thematic patterns in larger datasets such as to use text mining to uncover the hidden correlation between sentence complexity and the narrator’s mood fluctuation across all short stories of Poe.

Without compromising their critical judging and aesthetic appreciation on literary texts, the quantitative approaches to literary studies, strengthen literature students’ analytical skills and encourage them to evidence-based reasoning through adopting a numerical and scientific way in identifying, classifying and analysing texts. At the same time, the data-driven approach guides

linguistics students into considering more than a few constructed examples. In other words, students are expected to adhere to scientific methods, which includes observation, formulating hypotheses and testing the hypotheses with appropriate methods.

5 Outcome

5.1 Quantifying stylistic observations

The introduction of digital skills and formal systems in the training of humanities students in many ways complement the teaching and learning of traditional humanities subject contents. It enables students to apply traditional literary methods of close reading and deep analysis at the microscopic level to macroscopic genre studies of a larger dataset. We observe that through this training, students become more readily aware of the interrelationships between quantity, frequency, intensity of words and themes, patterns and styles of works. Literature students, who are used to informal methods of observations, are more willing to statistically organise their observations and to find connections and associations between these text segments. On the other hand, linguistics students, who have an inclination towards quantitative methods, are readier to probe into technical details and study the thematic networks portrayed by the statistics.

5.2 Identifying patterns in texts

In a typical theoretical linguistics curriculum, undergraduate students would encounter problems like identifying a specific morpheme from a given dataset containing inflected forms of the same lemma. An important part of language education is to enable students to independently identify patterns and make generalisations in chaotic language data. As mentioned above, students in English major tend to be weak in quantitative methods. We therefore approach problems by asking students to first formulate their hypotheses based on their knowledge in linguistics, and develop their arguments with the help of attested data.

In the corpus linguistics course, students have submitted projects on various topics that illustrate insights in linguistics with the support of quantitative data. These projects include ‘change and use of “-phobia” in COCA’, ‘semantic change in “get” and “like” in English’ and ‘acceptance of English-Cantonese code-mixing among Hong Kong Cantonese speakers’. In the projects, students have employed tools that they have learnt in

the course, such as AntConc, which they reported to have no prior knowledge about at the beginning of the course.

6 Conclusion

In this paper, we have discussed challenges of teaching digital literacies in curricula focused on languages. Students are often inexperienced in learning new skills related to computer and technology. We argue that it is desirable to introduce the technical skills in specific contexts, so that students are not lost in the technical details, such as syntax of commands or use of particular software.

The discrepancy between the skills of incoming students in language majors and minimal requirement to carry out NLP tasks can be intimidating and even prohibitive in some cases. We believe the teaching strategies described in this paper serve as more palatable first steps to encourage students to be more involved in learning about NLP and digital humanities in general.

Reference

- Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Davies, Mark. (2008-). The Corpus of Contemporary American English (COCA): 520 million words, 1990-present. Available online at <http://corpus.byu.edu/coca/>.
- McCarty, W. (2004). Modeling: a study in words and meanings. *A companion to digital humanities*, 42, 254. <http://www.digitalhumanities.org/companion/>
- Moretti, F. (2013). Operationalizing’: or, the Function of Measurement in Modern Literary Theory. *Literary Lab Pamphlet*, 6.
- Poe, E.A. (1903). *The Works of Edgar Allan Poe, The Raven Edition, Volume 2*. New York: P. F. Collier and Son.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.