

Sophister Module Description Template 2021-22

Full Name: Distant Reading

Short Name: Distant Reading

Lecturer Name and Email Address: Mark Faulkner (faulknem@tcd.ie)

ECTS Weighting: 10

Semester Taught: HT

Year: SS

Module Content:

Much literary scholarship (and, one might argue, your degree) rests on the close reading of a limited range of works that are in some senses canonical. Judging what is special about these works usually rests on the readers' impressions: that a specific word choice is particularly significant, that alliteration is frequent, a simile unusual, that a work echoes some other or is adopting a particular register. This module, by contrast, examines what kinds of literary scholarship are possible when we get our hands on datasets of eight million books and half a trillion words, see literary works primary as material or linguistic objects and open ourselves to quantitative as well as qualitative approaches to that data. Thus rather than merely stating our impression we think an author uses long sentences, we might work out how to compute her average sentence length relative to her contemporaries; rather than avowing the use of a particular word is striking, we might investigate whether its employment in the context the author has used it is actually unparalleled. The module draws inspiration from a number of directions, principally Franco Moretti's Distant Reading and the linguistic subfield of Corpus Stylistics, but its basic goal is to illustrate the range of techniques that are available to bring an empirical backbone to literary history. Each week focuses on a published computational or corpus-linguistic study of a particular literary work ranging in date from the seventeenth century to the present day, and on understanding, critiquing and developing the methodologies used.

Learning Outcomes:

By the end of this module, students should be able to:

- Describe some of the major distant reading and corpus stylistic approaches
- Evaluate what these approaches can (and cannot) tell us about literary-historical questions
- Use a range of techniques (e. g. regex, basic NLP techniques) to interrogate corpora
- Design and implement a distant reading study of a text or texts of their choice

Learning Aims:

- To introduce students to 'big data' approaches to literary history
- To show students how computational, quantitative techniques can complement traditional qualitative, human insights into literary texts
- To encourage students to reflect on what constitutes evidence in literary studies

Assessment Details:

- Number of Components: 2
- Name/Type of Component(s): series of four critiques of articles discussed on course; distant reading study of a text or texts of the author's choice.
- Word Count of Component(s): 4 x 500 words; 3,000 words
- Percentage Value of Component(s): 30%; 70%

Preliminary Reading List:

- Franco Moretti, *Distant Reading* (London, 2013)
- Dan McIntyre and Brian Walker, *Corpus Stylistics: Theory and Practice* (Edinburgh, 2019)
- Mark Faulkner, 'Corpus Philology, Big Dating and Bottom-Up Periodisation', in *Dark Archives: the medieval unread and unreadable* ed. Stephen Pink and Anthony Lappin (Oxford, forthcoming)

Please note:

- Curricular information is subject to change.
- Information is displayed only for guidance purposes, relates to the current academic year only and is subject to change.