"Rational Creatures": Using Vector Space Models to Examine Independence in the Novels of Jane Austen, Maria Edgeworth, and Sydney Owenson (1800–1820)

Sara Jane Kerr

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

Maynooth University

Faculty of Arts, Celtic Studies and Philosophy

February 2019

HoD & Supervisor 1:

Prof. S. Schreibman

Supervisor 2:

Dr C. Brunstrom

Contents

List of Figures

List of Tables

Abstract

Sara Jane Kerr

"Rational Creatures": Using Vector Space Models to Examine Independence in the Novels of Jane Austen, Maria Edgeworth, and Sydney Owenson (1800–1820)

Recent trends in digital humanities have led to a proliferation of studies that apply 'distant' reading to textual data. There is an uneasy relationship between the increased use of computational methods and their application to literary studies.

Much of the current literature has focused on the exploration of large corpora. However, the ability to work at this scale is often not within the power (financial or technical) or the interests, of researchers. As these large-scale studies often ignore smaller corpora, few have sought to define a clear theoretical framework within which to study small-scale text collections. In addition, while some research has been carried out on the application of term-document vector space models (topic models and frequency based analysis) to nineteenth century novels, no study exists which applies word-context models (word embeddings and semantic networks) to the novels of Austen, Edgeworth, and Owenson. This study, therefore, seeks to evaluate the use of vector space models when applied to these novels.

This research first defines a theoretical framework - enhanced reading - which combines the use of close and distant reading. Using a corpus of twenty-eight nineteenth century novels as its central focus, this study also demonstrates the practical application of this theoretical approach with the additional aim of providing an insight into the authors' representation of independence at a time of great political and social upheaval in Ireland and the UK.

The use of term-document models was found to be, generally, more useful for gaining an overview of the corpora. However, the findings for word-context models reveal their ability to identify specific textual elements, some of which were not readily identified through close reading, and therefore were useful for exploring texts at both corpus and individual text level.

Acknowledgements

Firstly, I would like to thank my supervisors Professor Susan Schreibman and Dr Conrad Brunstrom for their guidance, advice and support throughout my time at Maynooth University. Their support has enabled me to develop into a more confident researcher, writer and presenter.

This research was made possible by a John and Pat Hume Scholarship, without which I would not have been able to finance my studies.

I would also like to extend my appreciation to Dr Costas Papadopoulos, a valued member of my advisory committee, and Conor Wilkinson, from Graduate Studies, for all their help and support. Developing the ideas which have come to fruition in this thesis would not have been possible without the feedback I have received at conference presentations: thank you to the organisers, and to those who asked questions.

In addition, I would like to thank the members of my Viva Committee: Dr JoAnne Mancini, Dr Conor McCarthy and Dr Derek Greene, for their time, their feedback, and their kind comments on my work.

On a more personal note, I would like to express my sincere thanks and gratitude to my husband Iain, who has not only supported me emotionally and financially while I have completed my PhD, but has also put up with my work materials taking up almost every surface in our house. A special mention goes

out to my team of proofreaders and advisors: my sister Joanna, and my good friend Dr Stafford Murray. Finally, I would like to thank my parents, Martin and Elsie, who have always supported my academic endeavours.

Dedication

For Wan, Mollie and Erin.

Introduction

> My own sex, I hope, will excuse me, if I treat them like rational creatures, instead of flattering their fascinating graces, and viewing them as if they were in a state of perpetual childhood, unable to stand alone
>
> Mary Wollstonecraft, *A Vindication of the Rights of Woman*

Background of the Study

Jane Austen, Maria Edgeworth, and Sydney Owenson (later Lady Morgan) are three roughly contemporary authors for whom independence is significant in differing ways. Their success as writers, coming at the height of a backlash against revolutionary views, especially those voiced by women, is very different today from what it was during the period studied in this thesis. While Austen is vastly popular today, by 1820 her novels were no longer in print. Owenson, who is now relatively little known, was a popular and controversial author, her skills, in the eyes of some of her critics at least (see Chapter 1), surpassing those of the already well established Edgeworth:

> Lady Morgan's novels breathe of all the peculiar, tastes and feelings of her country, softened by the gentleness of her sex. They give us a view of Irish nature, as seen by female eyes ... There is

> in her works all the boldness of outline, with all the delicacy of touch — the quickness of perceiving truth and beauty . . . which may so frequently be observed in the best productions of Irish genius. She differs from Miss Edgeworth, as she has more heart and less judgement; deeper glimpses into the soul and less consistent views of superficial character; more passion, and less prudence; higher power to abstract us from the world, with less of practical wisdom to direct us in it. We turn from the dazzling brilliancy of Lady Morgan's works to repose on the soft green of Miss Austen's sweet and unambitious creations.("On the Female Literature of the Present Age" 1820, pp. 636–637)

The end of the eighteenth century and the start of the nineteenth century was a volatile period socially and politically. The industrial revolution challenged the traditional methods of production: "[a]long a broad frontier, technological changes transformed the way production took place in textiles, power technology, materials, and transport." (Mokyr and Nye 2007, p. 51). These technological developments changed the structure of society with migrants moving from rural villages to large cities for work (Plane, Henrie, and Perry 2005). This period marked an increase in the number of trade unions and demands for better pay and conditions, as well as more overt challenges to the political *status quo* (for example: the Luddite protests in 1811 and 1812, and the Peterloo Massacre in 1819). Improved literacy levels and the technology to produce newspapers and books cheaply and in greater numbers, also caused a dramatic increase in the number of books being published to meet the demand for new novels.

Writing at this turning point in social, economic, and literary history, Jane Austen, Maria Edgeworth, and Sydney Owenson sought to capitalise on the rising popularity of the novel form, and therefore it is not surprising that their works can provide insight into their political ideologies. As Alessa Johns points out, "ideas were not simply in their air, in the manner of the Zeitgeist, but were reproduced discursively and found their way . . . through traceable literary channels" (2014, p. 459). Part of the 'traceable literary channels' mentioned by Johns were the literary salons which provided a venue for learned discussions on art, literature and philosophy. Originating in France in the seventeenth century, the salon became an important feature of London society: "[n]otable among these were the Bluestocking salons associated with the salon hostesses Elizabeth Montagu (1718–1800), Elizabeth Vesey (c. 1715–1791) and Frances Boscawen (1719–1805)" (Prendergast 2011, p. 96). The term bluestocking initially referred to those who attended intellectual gatherings , but underwent "a curious anomaly in which a name used to denote a group of women *and* men became instead a term to describe and to satirize literary and learned women" (Haslett 2010, p. 434). However, the influence of the bluestockings in the development of ideas and literary style during this period cannot be denied. Deborah Heller presents "the Bluestocking exercise of reason in public as *the* decisive feature that defined the 'bluestocking movement' and provided its emancipatory impulse" (2011, p. 159). A central aspect of the movement "was its inclusion of *women* as co-partners in the conspicuously public activity of creating and disseminating opinions on matters ranging from literature to social and political issues" (p. 155). In Ireland too the salon became a space for shared literary and intellectual discussion. Edgeworth held salons at her

home in Edgeworthstown, and both she and Owenson were regular attendees at Lady Moira's salons held at Moira House in Dublin and Castle Forbes in County Longford (Prendergast 2011). This shift in role, from the domestic and private to the social and public, raised questions, still relevant today, of "how we speak in public, why and whose voice fits ... what [do] we mean by the 'voice of authority' and how we've come to construct it." (Beard 2017, p. 45).

Edgeworth and Owenson, as Anglo-Irish and Irish writers respectively, were writing at a time of "an ongoing mediation between borders" (Wohlgemut 1999, p. 645) both physical and ideological, and reflect the divided nature of Ireland at the start of the 1800s. The use of Irish or Anglo-Irish here is, to some extent, for convenience. The terms provide a loose indication of each authors' political leanings, as neither term adequately reflects the reality of the complex nature of nationality during this period. Although we can consider the core difference to be between the legal land ownership of the Anglo-Irish, and the traditional, or moral, rights of the native Irish to their homeland, the hyphenated Anglo-Irish identity was one which blurred the lines between what it was to be English or Irish. Edgeworth of course, living on the family estate in Longford County that took her family name (Edgeworthstown), belonged to the Anglo-Irish social class descended from the members of the Protestant Ascendancy (R. Tracy 1985). However, both authors fit the *OED* definition of Anglo-Irish as being "of English descent but born or resident in Ireland". Owenson is also "of mixed English and Irish parentage", something hinted at by her surname: 'Owenson' being the Anglicised version of her father's original name: 'MacOwen'. Like many of

those who attended Lady Moira's salons, they shared a "great enthusiasm ... for antiquarianism: the study and appreciation of the language, customs and cultures of ancient Ireland" (Prendergast 2011, p. 100). Yet the two authors can be seen to represent the "estrangement between Anglo-Irish and Irish, Protestant and Catholic, legality and ancient traditionary right, landlord and tenant" (R. Tracy 1985, p. 19), made even more complex following the 1798 rebellion and the 1801 Act of Union. Owenson wrote as a champion of the Irish cause, while Edgeworth, although sympathetic to the Irish, wrote from what Terry Eagleton refers to as an "outside vantage point" and "a site of power" (Eagleton 1995, p. 176).

While the political views of Edgeworth and Owenson are often stated directly, in their novels or other writings, Austen's are much more difficult to define. Ashley Tauchert suggests that:

> Her six complete novels offer tantalizing glimpses of the subtle range of female agency within lesser or greater material and proprietal constraints of subjected domestic existence. These narratives work through variations on the resolution of intimate domestic tableaux disordered by the social demands of courtship, as experienced and understood through the consciousness of a young woman between adolescence and marriage. Austen offers plausible narratives of positive female agency, and this is represented through feminocentric narration concerned with the significance of specifically feminine modes of female reason, and its ability to reform and transform the otherwise degraded social context. (2003, p. 150)

However, not all critics would entirely agree with this interpretation (Tobin 1990; Butler 1997; Martin 1998; Downie 2010). The difficulty in definitively assigning a political leaning to Austen's works, along with her economic lexical style (explored further in Chapter 4), may in part be responsible for Austen's lasting cultural popularity. She has, in effect, become a cultural touchstone, a shorthand for the expression of a broad range of ideas, especially those defining Englishness. There is a darker side to Austen's cultural popularity, however: she has been increasingly lauded by the far-right as a writer who embodies their views. Nicole Wright's article 'Alt-Right Jane Austen' identifies "several variations of Alt-Right Austen 1) [a] symbol of sexual purity; 2) [a] standard-bearer of a vanished white traditional culture; and 3) [the] exception that proves the rule of female inferiority" (2017). In the light of this trend, it becomes increasingly relevant to question the true nature of Austen's work. As such, Austen is the central focus within this thesis.

Objectives of the Research

Examining the novels of Jane Austen, Maria Edgeworth and Sydney Owenson published between 1800 and 1820, this study provides an insight into the ideology of the three women, and how that manifested itself differently given the authors' economic, social, political and national differences. Traditional close reading by necessity focuses on the detailed analysis of small sections of text. It must be selective in the examples chosen to support the argument being presented. While it is possible to construct a convincing argument regarding the social and political beliefs of Austen, Edgeworth and Owenson, supported

by extensive quotations from their novels, it is equally possible to construct an opposing political viewpoint using the same novels. Critics disagree, in particular, in regards to where Austen lies on the political spectrum, placing her anywhere from deeply conservative to radical (Butler 1997; Kirkham 1997; Emsley 1999). This interpretive variation is, of course, part of the appeal of literary studies, but is it possible to discern the political position of an author using digital techniques? In searching for insight into the traces of a novelist's political views, we need to look for more subtle patterns within and across the texts. In effect, we are looking for an understanding which goes beyond the individual novel, and therefore techniques which go beyond close reading.

"While the terms 'data' and 'modeling' may be new, many of the activities and intellectual frameworks they entail are familiar and deep-rooted ... In a more specific sense, our models represent the shaping choices we make in representing and analyzing the materials we study" (Flanders and Jannidis 2018, p. 3). In contrast to previous research, which has used digital tools to explore large corpora of nineteenth century literature (Allison et al. 2011; Jockers 2013a; Moretti 2013; Algee-Hewitt et al. 2016) and more specialised corpus-based research on the novels of Austen (Burrows 1987; Burrows 1989; DeForest and E. Johnson 2001; Burrows 2002; Fischer-Starcke 2010), this research capitalises on the "conceptual shift that digital humanities brings" and "the opportunity to formalize and exploit information models" (Flanders and Jannidis 2018, p. 7), examining the effectiveness of applying a combined close reading and computational approach to a mid-sized corpus.

The aim of this thesis is to evaluate the effectiveness of distant reading techniques in the analysis of small to mid-sized literary corpora, and to demonstrate how the applications of the techniques can lead to enhanced readings of the works. Focusing on a corpus of 28 novels from Austen, Edgeworth, and Owenson, this thesis investigates the use of a specific set of distant reading techniques: vector space models. It also seeks to provide a theoretical framework in which both close and distant readings of literary corpora can be combined.

Specifically, this thesis investigates the use of two main models of distant reading: the more established term-document model, which includes frequency analysis and topic modelling, and the less well known word-context model, which includes word embeddings from which semantic networks can be created. These methods are commonly used in computer science, in particular for information retrieval and artificial intelligence, and have been applied to very large corpora of contemporary language. However, previous studies have not dealt with the application of these methods to small and mid-sized corpora of nineteenth century novels. This thesis aims to address that gap.

Structure

This study is divided into five main chapters. Chapter 1 considers the literary and historical context of the novels, and includes a critical analysis of the representation of independence. Chapter 2 presents a discussion of the challenges and controversies currently surrounding distant reading and proposes 'enhanced reading' as a combined approach, presenting it as a methodological

framework for exploring the small to mid-sized corpus. Chapter 3 outlines the specific methods and processes used in this thesis. Presented in Chapter 4, is an exploration of the texts at a macro-level, using frequency analysis and topic modelling. Chapter 5 examines the use of word embedding models and semantic networks to explore the theme of independence.

A Note on Sydney Owenson

Sydney Owenson is the only one of the focus authors to marry and therefore while her early novels were published as Sydney Owenson (or S.O,), after her marriage in 1812 her later novels were published under the name Lady Morgan. Within the academic literature there is no set convention as to which name is used. Rather than using 'Morgan (née Owenson)' when referring to her later novels, this thesis uses Owenson throughout.

1 Independent Women

> The extremists are afraid of books and pens, the power of education frightens them. They are afraid of women. The power of the voice of women frightens them
>
> Malala Yousafzai, 'Speech to the United Nations' (2013).

To examine the interrelationship between the contexts and meanings of texts, and the methods used to extract these, this chapter contains two main sections. It will first examine the contemporary reception of Austen, Edgeworth and Owenson, before moving on to consider the recent literature which explores the authors' representations of independence. *The Historical Thesaurus of English* provides the following definitions for 'independence' in use during the period covered in this study (University of Glasgow 2015):

- Wealth: Sufficient means
- Lack of subjection
- Lack of submission to rightful authority

For 'independent' the definitions are:

- Not submissive
- Not influenced by others
- Self-sufficient: not dependent on another for a living

Drawing on these definitions, independence here is used in its broadest sense, encompassing financial, social, national and personal freedoms.

An understanding of the social, cultural and political landscape within which the novels were written is essential for an analysis of the texts, whether computational techniques are being used or not. A text which may appear to be relatively innocuous to a modern reader may have been viewed very differently when it was originally published. This section will explore women's reading and women's writing, before moving on to consider the contemporary reception of Edgeworth, Owenson and Austen's novels between the years 1800–1820; it is not designed to be an exhaustive survey of the early critical reception of the novels, for this see Belanger 1998, Belanger 2007, and Mazzeno 2011.

1.1 Contemporary Reception

1.1.1 Women Readers, Women Writers

The end of the eighteenth and the start of the nineteenth century marked a broad shift in literacy levels, in particular amongst the middle and working classes, "printed and hand-written commercial documents had become ubiquitous, and an ability to read was essential to the success of even small-scale" businesses (O'Ciosain 1997, p. 27). This need for literacy in businesses dealings was exacerbated by the Restriction Period (1797–1821) during which the Bank of England stopped "paying out gold" and "issued £1 and £2 notes" (Bank of England n.d.). Numerous provincial banks also produced their own

notes, causing challenges for the public if the banks failed and necessitating careful reading to ensure a note was valid currency, something Jane Austen would have been more than aware of following the collapse of her brother Henry's bank (Caplan 1998).[1]

The end of the long eighteenth century also marked a dramatic shift in the number of publications being produced.[2] The cost of books gradually reduced. A greater focus on education for all classes and improving literacy rates meant that an ever growing population of readers wanted material. The importance of education for all had its roots in the radical writings of the late 1700s, such as Thomas Paine's *The Rights of Man* and Mary Wollstonecraft's *A Vindication of the Rights of Woman.* Less controversial educationalists, such as the Edgeworths, produced pedagogical guides to aid those educating their own children, and societies such as the National Society for Promoting the Education of the Poor (1811) were founded. There was also an increased political interest in education, for example Samuel Whitbread's (1807) Bill which, unsuccessfully, proposed that all children under 14 should receive two years of education.[3] Improved female literacy meant that books designed for consumption by a female audience were particularly in demand. Circulating libraries

[1] Although the £1 note was the smallest denomination issued by the Bank of England, smaller denominations were also circulating (Boyle and Geary 2003). In Ireland, the "shortage of metal during the wars led to a reliance on paper currency for sums as small as three shillings" (O'Ciosain 1997, p. 27).

[2] There is no formally agreed start and end dates for the long eighteenth century in academia, start dates range between 1660–1688 while end dates run from 1815–1840. The dates for the long eighteenth century used here are 1680–1830.

[3] It wasn't until 1870 that the first Education Act was passed creating compulsory education for children between five and 13.

offered improved access to books for those in the middle classes and the working classes who could afford them.[4] No longer was access to reading material the sole preserve of the land-owning gentleman in his library. Increased mechanisation, serialisation and improved access opened the world of books to a much broader audience than ever before. It was, as Armstrong notes: "the new domestic woman rather than her counterpart, the new economic man, who first encroached upon aristocratic culture and seized authority from it" (Armstrong 1987, p. 59).

Readers and the Role of the Circulating Library

While it was highly desirable for women to be able to read in order to teach their children, the type of text being read, and how and when it was read, became the focus of moralists. Reviewers, the vast majority of them male, suggested the types of books to be read by women: sermons, moral tales, and educational texts were all approved, novels and texts which dealt with the male public sphere were not: "Young ladies ought to be particularly careful in choosing the subjects of their reading; they have much leisure for reading, and on that account books have a great share in their meditations, and strongly influence their characters" ("An Essay on Reading , Addressed to the Ladies" 1810, p. 407). However, many commentators regarded reading and learning as an addition to a woman's "personal charms" for the opposite sex rather than

[4]The term 'middle class' is a somewhat contentious one, James Downie suggests the use of 'pseudo-gentry' as it allows a greater accuracy in the definition of social class (Downie 2006). However, the term was used during this period (for example by Mary Wollstonecraft) and, although it has become loaded with socio-economic and political meaning which may mask the sense in which it was used during this period, this is the term which will be used here with this caveat in mind (Wollstonecraft 2008).

for their own benefit (p. 407). Reading was to be done after a day's useful work had been completed with the purpose of education and moral improvement rather than entertainment. Magazines and sermons raised considerable concern regarding novel reading and the lower classes. The fear was, perhaps, that the traditional patriarchal hegemony would be called into question and challenged by those who had previously accepted their lowly position in society, male and female alike.[5]

The greatest outrage, however, was kept for novel readers. While the educated man could read novels with impunity, should he so choose, they were considered dangerous for the delicate mind of the woman, in particular the young. This was part of "an anti-novel discourse that repeatedly claimed the debilitating effects of novel-reading on youthful minds" (Runge 2009, p. 279). Polemics against reading warned of the dangers to the moral and physical well being of the female reader. In effect, the dangers of novel reading were conflated into the dangers of loose morals and revolutionary intent (Vogrinčič 2008). It is perhaps not surprising that, during a period of revolutions and almost constant war, the desire to maintain the status quo was often at the heart of criticism against women and other disempowered groups. There was fear that the influence of revolutionary thought from France and America would spread among the "lower and middling orders of the people" ("Miss Edgeworth's Popular Tales" 1804, p. 330). As an anonymous reviewer in 1814 stated:

> The influence of a novel upon the generality of its readers has

[5] A letter to The Times from 6th June 1815 declared: "IT IS NOT BONAPARTE THAT AT PRESENT FORMS THE DANGER OF EUROPE ... IT IS THE NEW OPINIONS" (1815, p. 705).

> been much undervalued; it was once considered as a relaxation only for the minds of the studious, and as a momentary resource for empty volatility, or thoughtless indolence . . . So much has this power been acknowledged of late, that it is now the fashion to interweave history, morality, and religion into the text of a novel, and to the reader what was intended only as a refuge for the indolent a vehicle of instruction and a means of improvement. All these circumstances call for a strict examination of the principles of those works which have so powerful, though imperceptible, an influence on the public mind, lest popularity should be mistaken for truth. ("Patronage by Miss Edgeworth" 1814, pp. 159–160)

The Belfast Monthly Magazine, severely criticised novel reading for women in particular:

> The first regulation of a lady's library ought to be the exclusion of novels and romances . . . The generality of novels are as contemptible in a literary point of view, as hurtful in a moral one; the ideas they contain are mostly far fetched, the expressions ridiculous, and the language such as can excite nothing but contempt. Novels occasion a dreadful loss of time, they spoil the taste of their readers, they unbend the mind in the most unfavourable manner, and too often hurt the morals. ("An Essay on Reading , Addressed to the Ladies" 1810, p. 407)

This view of novels reveals the attitude of the wider, patriarchal society, that

women's morals were vulnerable to attack from all sources and that the foremost purpose of a woman's education was to create a pleasurable companion for her husband. As such, those who wrote novels were viewed as influencing their readers, and their own morals, in particular if they were female, were open for comment.

Amongst those who were criticised were the 'bluestockings'. The term 'bluestocking' which was "a term developed in the mid-eighteenth century to denote, initially, a social circle interested in intellectual improvement and the exchange of ideas through salon-style conversation and correspondence", had, by the early nineteenth century, become pejorative (Johns 2014, p. 13).[6] Satirical pictures, for example Thomas Rowlandson's 1815 print 'Breaking up of the blue stocking club', presented the intellectual woman as a hideous and boorish character who neglected her womanly duties, thus linking her to ideas which were anti-domestic and revolutionary (V. Jones 1990). The bluestockings blurred the lines between the public and private worlds, as Deborah Heller notes, "[s]ituated at the intersection of *female*, *intellectual*, and *public*, the notion of bluestocking, if not a lightning rod for spite and misogynist rancor, was at least a magnet for conflict around the issue of gender propriety" (2011, p. 155).

Austen, Edgeworth and Owenson found themselves writing in the midst of "the contest over the cultural meanings of the word" 'bluestocking' (Haslett 2010, p. 434), and therefore had to position themselves very carefully to avoid open hostility. Although the early bluestockings had been celebrated for their

[6]Although the term 'bluestocking' was introduced to England in the mid-eighteenth century, the term 'bas bleu' had been used in France to denote literary women since the 1500s.

intelligence and learning, by the early nineteenth century this had changed; too much education and a woman could face the "stigma of the learned lady, whose reputation was bad throughout the period" (Pearson 1999, p. 15).[7] This conflict is presented by Edgeworth in *Patronage* as Georgiana Falconer claims to be intimidated by Caroline and Rosamond Percy's education, saying: "they are all so wise, and so learned, so blue, such a deep blue" (1814, p. 6). Bluestockings, like Hester Chapone, attempted to distance themselves from excessive learning by expressing horror at the prospect of a learned, unfeminine lady: "The danger of pedantry and presumption in a woman — of her exciting envy in one sex and jealousy in the other — of her exchanging the graces of imagination for the severity and preciseness of a scholar, would be, I own, sufficient to frighten me from the ambition of seeing my girl remarkable for learning" (Chapone 1990, pp. 105–106).

Austen too explores this debate in *Pride and Prejudice*:

> 'Miss Eliza Bennet,' said Miss Bingley, 'despises cards. She is a great reader and has no pleasure in anything else.'
>
> 'I deserve neither such praise nor such censure,' cried Elizabeth; 'I am *not* a great reader, and I have pleasure in many things' (1985, p. 83).

Caroline Bingley's attack on Elizabeth is a far greater criticism than it appears to be to a modern reader. The emphasis on anti-domesticity and independent thought encompassed by the phrase 'great reader' is highly critical and Elizabeth's response places her within the traditionally accepted role of reading for

[7] The period of time referred to by Jacqueline Pearson is 1750–1835.

pleasure. The choice of the word 'censure' indicates an awareness of the true meaning behind Caroline's comment. As Jacqueline Pearson notes, "when accused of being a 'great reader', a heroine generally protests...only foolish or malicious characters make such accusations" (1999, p. 15). We view Caroline as both. However, Austen is clear to show that while the exchange conforms with the traditional view of women's reading and education, she and her hero Mr Darcy do not agree with Caroline's criticism.

Bluestocking salons, like those of Lady Moira discussed in the Introduction, were almost exclusively an upper class phenomenon, designed to provide a venue for the discussion of art, literature and philosophy. In contrast, circulating libraries were far more egalitarian and were instrumental in increasing access to popular literature, especially for those who could not afford to purchase books. The business of a circulating library could be very lucrative: "a library proprietor could purchase a few copies of a popular novel, but loan out each volume to individual patrons, thus maximizing the profitability of the book" (Mandal 2015, p. 19). Lee Erickson explains:

> By 1800 there were estimated to be a thousand circulating libraries in Britain. The small ones primarily stocked fiction and were often supplied by William Lane, who specialized in novel publishing at his Minerva Press, which he established in 1790 ... At its peak in the first decade of the nineteenth century the Minerva Press brought out 30 percent of all new novels ... In the 1810s and 1820s, Henry Colburn similarly moved from owning a circulating library and the weekly *Literary Gazette* to becoming London's largest and most colorful publisher of novels. (2009, p. 226).

The increase in popularity and demand for books, novels in particular, led to a rapidly growing number of authors, many of them women. The quality of many of these productions is hard to ascertain as only a small proportion have survived even in digital format. "Barbara Benedict maintains that circulating library novels were largely formulaic, calculated to encourage quick, cursory reading and to engross readers at the end of each volume" (Wyett 2015, p. 267). However Jennie Batchelor positions this as a "representational trend that . . . figured the turn of the nineteenth-century literary marketplace as a degradingly feminized arena, prostituted by genius's 'commodified counterfeits': namely, the illegitimate labours of the female drudges" (2010, p. 160). This was, as Batchelor suggests, part of a broader message that women's work held less value than that of men. This message also sought to restrict and trivialise the types of employment deemed suitable for women.

Austen, herself a subscriber to circulating libraries, comments on the freedom a subscription could offer to someone without the benefit of a well stocked library at home in *Mansfield Park*. Fanny Price returns to her family home in Portsmouth with the sum of £10: "wealth is luxurious and daring, and some of hers found its way to a circulating library. She became a subscriber; amazed at being anything in propria personal [sic] amazed at her own doings in every way; to be a renter, a chuser of books!" (1994b, p. 404). As Peter Knox-Shaw explains, the access to the library and the freedom of choice it represents propels Fanny from being a mere dependent, "causing her to feel for the first time that she is a person in her own right" (2009, p. 196). It is notable that, while living at Mansfield Park, Fanny does not appear to have free access to its library. Instead, her choice of reading materials is selected

for her by Edmund with the aim of improving her mind.

The question of who subscribed to circulating libraries, and to what extent they were a democratising influence, is partly an economic one. Annika Bautz, citing the high subscription fee ("between two and five guineas") of the Minerva Library, comments that "while new fiction may have reached some middle-class readers through libraries, novels are unlikely to have reached far below middle class" (2007, p. 1414). However this is unlikely to have been the norm. Christopher Skelton-Foord notes that "[w]e know that in the 1780s, Baker's [in Southampton] were allowing readers to rent an unlimited number of volumes for just 3*s.* the season, or 10*s.* 6*d.* the year; and by 1823, E. Skelton and Co. were providing a new family subscription rate of 10*s.* per month" (1998, p. 350). There may even have been libraries which specifically targeted their pricing to the working classes as part of their business plan, for example "William Harrod's circulating library in Stamford, which had a keen eye on the developing 'lower orders' market" (Skelton-Foord 1998, p. 353).

Examining the numerous criticisms of circulating libraries and novel reading reveals issues of gender: "[t]he fortunes of circulating libraries and novels were inextricably linked', writes Anthony Mandal, 'often to the detriment of both: the former decried as sites of inappropriate class and gender mixing, of improper flirtations and illicit seductions; the latter vilified as purveying bad morals and unrealistic romances that would corrupt their young (female) readers" (2015, p. 19). There has been a pervasive assumption that the readership of novels at this time was mostly female. This is an idea that Jan Fergus criticises, highlighting that this has been largely based on the accounts of reviewers and moralists, rather than an exploration of the book reading public.

Her research highlights the numbers of men, and schoolboys who made up the buying and subscribing public, and "reveals women of all classes extensively read works traditionally considered 'feminine"', but also "that women also read heavily in 'masculine' genres, including periodicals and philosophy" (Fergus 2007, p. 61).

With women at the heart of the domestic sphere, and "the patriarchal household that represents the nation in microcosm" (Wolf 2004, p. 281), any challenge to the home was perceived as a challenge to the security of the country. This conflation of family and country is outlined in Edmund Burke's (1790) *Reflections on the Revolution in France* where he wrote of: "binding up the constitution of our country with our dearest domestic ties, adopting our fundamental laws into the boson of our family affections, keeping inseparable and cherishing with the warmth of all their combined and mutually reflected charities our state, our hearths, our sepulchers, and our altars" (1951, p. 32). The protection of English domesticity is pitched against the threat of the, often Catholic, anti-domestic 'other'. Female morality was at the heart of this threat; the perceived threat from the licentiousness of post-Revolution France with its calls for greater freedoms for women, challenged the patriarchal norm.[8] The role of the family was viewed by some religious commentators as a structure at risk from immoral ideas, as argued, for example, in the anonymous pamphlet *The Evils of Adultery and Prostitution; with an Inquiry into the Causes of their Present Alarming Increase, and Some Means Recommended*

[8]Defining Britain in contrast to France was a common part of setting Britain against the Catholic 'other'. Pre-Revolutionary France was also criticised, with some cause, for licentiousness as documented by Robert Darnton, for example, in *The Forbidden Best-Sellers of Pre-Revolutionary France* (1996).

for Checking their Progress (1792).

The rise of the cult of domesticity, the emphasis placed on a feminine ideal which highly valued decorative and domestic skills, and the criticism of educated women, underpinned a more visceral fear. If women of the middle and upper classes were no longer prepared to accept their domestic life, the traditional role of the family with its patriarchal head was threatened and the rights of those in power, supported by the patriarchal structure of the family and primogeniture, could be questioned by the masses without political and personal power. Sensibility, and its link to anti-domesticity, was also perceived as a threat to domestic stability. In her article on the the feminine voice, which focuses on Shelley's *Frankenstein*, Rhonda Batchelor states:

> Through the overview that Shelley offers, we can see how nineteenth-century woman, as a representative of the stilled domestic, was refused a public social voice and allowed only a private one. All of these novels [*The Wrongs of Woman*, *Frankenstein*, and *The Scottish Chiefs*] tacitly acknowledge patriarchal society's perception of its own need to silence the feminine voice from which it had derived the means for moral selfhood ... Without their voices and energy, disguised and muted as they may be, woman might have remained the male-dependent, dead weight at the heart of the stabilized domestic. (1994, p. 368).

Batchelor argues that the ideal nineteenth century woman was one whose place was in the home, and whose voice was silent in deference to the men she was dependent on, what Batchelor refers to as the 'stilled domestic'.

As Ruth Bloch explains in her 1978 essay 'Untangling the Roots of Modern Sex Roles: A Survey of Four Centuries of Change': "[w]hile men generally dominated the highly fluid and competitive economic and political spheres, women increasingly gained ascendancy in the sacred, moral, and emotional spheres of life: religious benevolence, sentimental fiction, and the family" (2003, p. 49). This concept of separate spheres became codified during the nineteenth century through conduct books, magazine articles and sermons, with an increasing focus on women's role in the home, an ideology eventually becoming known as 'True Womanhood'. The insidious hold of this concept still influences attitudes towards gender roles today. It is perhaps unsurprising, as we shall see later in this chapter, that many of the criticisms faced by women writers during this period centre on the concept of their place within the home.

Yet, to a certain extent these representations of women, as defined by the separate spheres ideology, were themselves literary constructions, supported by conservative commentators and conduct books (Heller 2011). Amanda Vickery, in her review of Leonore Davidoff and Catherine Hall's (1987) *Family Fortunes: Men and Women of the English Middle Classes, 1780–1850*, argues that, "rather than conclude from positive female testimony that women were not necessarily imprisoned in a rigidly defined private sphere, the dominant interpretation simply sees the private sphere in a better light" (Vickery 1993, p. 386). She further suggests that Davidoff and Hall's argument that "gender played a crucial role in the structuring of an emergent, provincial, middle-class culture" is based on a series of "problematic assumptions" (p.

394) that middle-class women's lives were restricted largely to the home. Susan Steinbach explains that the separate spheres ideology: "was not a rigid set of rules internalized as natural and adhered to unquestioningly. Rather, separate spheres were in the process of being constructed, rife with internal contradictions, and frequently challenged (both overtly and covertly)" (2012, p. 830).

The "discourse of domesticity" (Klein 1995, p. 101) may have been part of a broader moral panic about the "rise of luxurious corruption" and that this may be realistically attributed to "the persistence of *male* anxieties" (Vickery 1993, p. 407), in part in response to the societal changes caused by the Industrial Revolution. However, it can certainly be argued that the targets of those seeking to define women within a domestic setting were, initially at least, those in the middle classes of society. As Ruth Bloch notes the: "familiar evolution of the 'woman's sphere' must start with the decline of economic production in the home and the transformation of domestic work into an activity with no commercial value" (2003, pp. 49–50). As pay and working conditions improved even working class women began leaving work outside the home when they had families as the idealised version of 'True Womanhood' gained hold.

It is clear that, although the sweeping generalisations occasioned by the 'separate spheres theory' seem to be inaccurate, or at the very least overstated, the desire for female domesticity and the rejection of anti-domesticity, with its links to revolutionary thought, were a feature of early nineteenth century criticism of women writers and the educated woman, as we will see in Section 1.1.2 below. While women's actions and influence were restricted to the

home they could be denied influence in the public sphere, and calls for greater gender equality could largely be ignored. It is notable that this somewhat contentious negotiation of gender roles, and a parallel exploration of identity and nationhood, is a feature of many of the novels written by Austen, Edgeworth and Owenson. Edgeworth in particular was a proponent of domesticity, viewed as: "a mild reformist, a writer whose support of rational domesticity and the improvement of female education did not challenge dominant social structures" (Weiss 2007, p. 442).

Writers and Contemporary Publishing

The late 1790s marked the start of a rapid increase in the number of novels being published in the United Kingdom. Between 1800 and 1820 a total of 1502 novels were published, almost 50% written by women authors, defined by Garside and Schöwerling as 'Named', 'Identified' and 'Implied' (2000, p. 73).

Garside and Schöwerling's bibliography adopts:

> a more rigorous definition of the 'novel' than in earlier volumes ... This is partly because of the huge upturn in the number of publications over this period ... but also because the forms of the prose novel became more readily identifiable. [It] includes what contemporaries thought of as novels, incorporating works categorized as 'novels' in contemporary periodical reviews and under 'novels' headings in circulating library catalogues ... Collections of tales (including some mixed genre compilations) are included. (2000, p. 4)

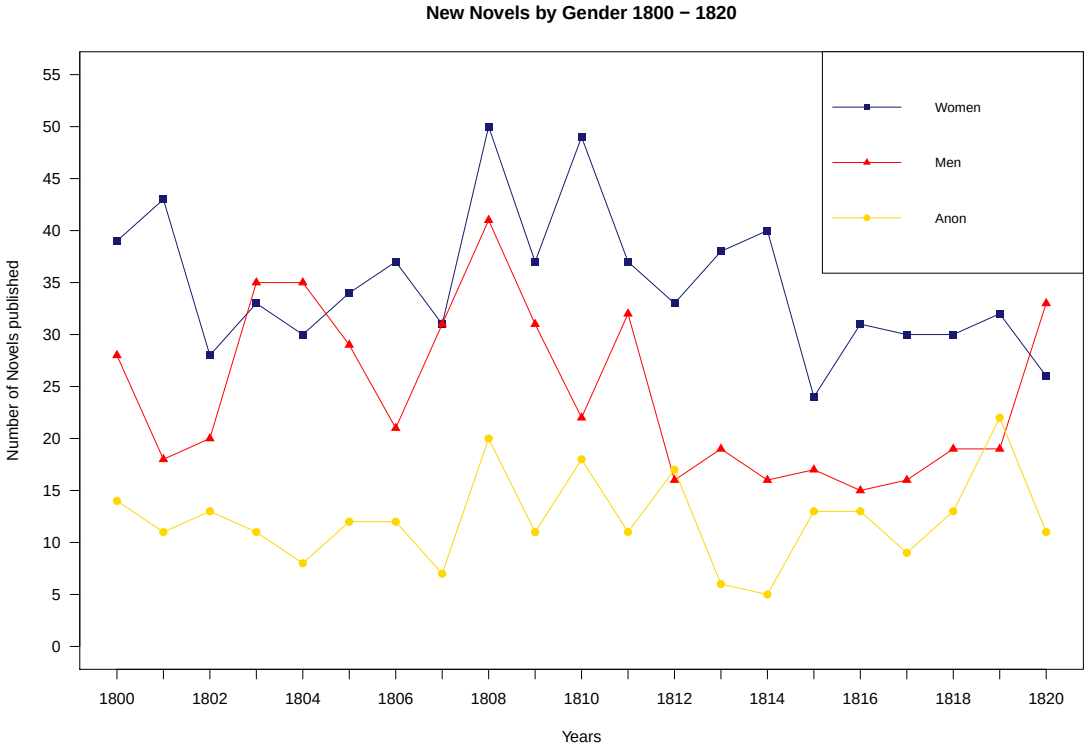


FIGURE 1.1: Novels By Gender 1800-1820. (data from Garside and Schöwerling 2000, p. 73)

The 1810s in particular marked a high point for women novelists and it was during this period that Austen's novels were published (see Figure 1.1). During the same period there was a rapidly increasing international interest with 179 American editions of British novels published and 158 editions translated into French and German (Garside and Schöwerling 2000, pp. 98-101). This positions the novels written by Austen, Edgeworth and Owenson within a broader international context of publishing at the start of the nineteenth century.

It is clear that there was a social and political agenda which underpinned

women's reading during this period; this also logically attaches to books written by women, especially novels. At this time writing, as well as reading, was a political act. Kathryn Kirkpatrick goes further, indicating that the "distinction between the administrative practices of the state and the cultural processes of the nation ... position writing as part of the ongoing activity of (trans)forming and articulating national culture and, hence, national identity" (2008, p. 5). This idea of literature creating, defining and refining national identity is evident in the works of Owenson. *The Wild Irish Girl*, *O'Donnel*, and *Florence Macarthy* in particular discuss what it was to be Irish, using Owenson's antiquarian knowledge to place her notion of Irishness within a rich historical context.

The domination of female writers in the early part of the nineteenth century marked a shift in how women were represented: "narratives which seemed to be concerned solely with matters of courtship and marriage in fact seized the authority to say what was female" (Armstrong 1987, p. 5). The threat felt by some of the reviewers may have been prompted, according to Nancy Armstrong, by a shift where:

> In place of the intricate status system that has long dominated British thinking, these authors began to represent an individual's value in terms of his, but more often in terms of *her*, essential qualities of mind. Literature devoted to producing the domestic woman thus appeared to ignore the political world run by men. Of the female alone did it presume to say that neither birth nor the accoutrements of title and status accurately represented the individual (1987, p. 4).

The influence of the Act of Union in 1801 altered the publication landscape in Ireland, something we can see reflected in the publication history of Owenson who, while initially choosing a Dublin publisher for *St Clair* was quick to secure a London publisher when she could.[9] While Dublin publishing houses were still influential at the start of the century this was not to continue: "The Union is traditionally held to have affected the Dublin printing trade adversely ... The extension of the copyright law to Ireland meant the end of the reprint trade ... The early nineteenth century was therefore a period of slack demand and low wages within the printing trade as a whole" (O'Ciosain 1997, p. 56). As the vast majority of novels printed in this period came from "London publishing concerns" in effect what this meant was that for a novel to have the best likelihood of success, access was needed to a London publisher and the London book trade. A writer also needed to ensure that the novel itself was appealing to an English audience (Garside and Schöwerling 2000, p. 76).

Finding a publisher for a novel could prove particularly challenging for women writers as they were often discouraged, although not actively prevented, from conducting business in their own right, relying instead on the men in their family to act for them. Maria Edgeworth relied upon her father and his contacts with the bookseller Joseph Johnson in London to publish *Castle*

[9]In a letter to Mrs Featherstone written in October 1802 Owenson writes "My novel is publishing this month back, in Dublin, and will be out early next month" (Morgan and Hepworth Dixon 1863a, p. 223). Writing to Miss M. Featherstone in June 1803 she reveals: "that I have got a price far beyond my most sanguine wishes for *St. Clair*. Mr. Harding, of Pall Mall, says it will be done in a very superior style, and will be certainly at Archer's in three weeks" (p. 237).

Rackrent (Garside and Schöwerling 2000), while Austen's father, unsuccessfully approached publisher Thomas Cadell in 1797 (Morris 1982) with *First Impressions*, an early draft of *Pride and Prejudice*. This was not the only difficulty Austen faced while trying to find a publisher. In 1803, through her brother Henry, Austen sold an early draft of *Northanger Abbey*, then known as *Susan*, to Richard Crosby and Sons for £10 with a promise of early publication: "It had in fact been announced, in *Flowers of Literature* Vol.1, for 1801 and 1802, published in 1803, as being 'In the Press', but publication had not taken place" (Gilson 1982, pp. 82–83). Six years later, Austen sent a letter to Crosby and Sons under the pseudonym Mrs Ashton Dennis (allowing her to sign the letter "I am Gentlemen, MAD" (Gilson 1982, p. 83)) which attempted to prompt them to publish the novel, asked whether the manuscript had been lost, and threatened to find another publisher if she heard nothing from them.[10] Crosby responded, informing her that they were not obliged to publish but would prevent any attempt by another publisher to do so, and that she could buy back the manuscript for the price originally paid for it (Gilson 1982). Her brother eventually bought back the manuscript in 1816 but it remained unpublished until after Austen's death.

Sydney Owenson, working as a governess at the time she wrote her first novel and having no business-minded male relative to act for her, undertook the work of finding a publisher herself. Owenson's socio-economic challenges certainly gave her a greater necessity for becoming a successful author. Unlike

[10]There has been some discussion over the true authorship of this letter, under the suggestion of the Austen-Leighs that it was not in Austen's hand. Perhaps this rather blunt and forceful business letter did not fit with the accepted modest image of Jane Austen? Arthur Axelrad's close examination of the original letter, however, confirms that it was most likely written by Austen herself (1994).

Austen and Edgeworth, who had their families to support them, Owenson took on the role of supporting her father and sister. Although young and female, she clearly had a talent for business and self promotion, as well as a solid understanding of the market for her work, successfully publishing music, poetry and drama, as well as her novels.

Owenson wrote directly to the publisher Richard Phillips with a view to publishing her second novel *The Novice of St. Dominick*. Phillips' reply, although hesitant to name a price without seeing the manuscript, indicated that he was clearly interested based on the initial success of *St. Clair* "of whose merit I entertain no doubt, since it is demonstrated arithmetically by the number that has been sold" (Morgan and Hepworth Dixon 1863a, p. 250). Owenson ignored Phillips' suggestion that she send the manuscript to him, deciding instead to personally deliver it, travelling alone from Ireland to do so. She must have made a good impression on Phillips as not only did she successfully sell the manuscript, but he wrote to her with literary advice (Morgan and Hepworth Dixon 1863a, pp. 254–255). For the publication of *The Wild Irish Girl* she successfully created a bidding war between rival publishers Phillips and Johnson, who published Maria Edgeworth's novels. The flurry of letters from Phillips and Johnson in April and May 1806 demonstrate Owenson's business sense and determination to garner the best price for her work, finally settling on £300 and £50 for each subsequent edition (Morgan and Hepworth Dixon 1863a, pp. 268–275).

Ironically, it was the novel, which was viewed as such a frivolous genre, that provided some women, like Owenson, with a means for their own independence. The unmarried woman, often viewed as a burden on her family or

society, could find a way to support herself in a manner which was on the edge of respectability but lacked the subservient nature of the other traditional roles deemed suitable for genteel women such as governesses and ladies' companions.[11] A contemporary source noted in 1800:

> The penalties and discouragements attending the profession of an author fall upon women with a double weight; to the curiosity of the idle and the envy of the malicious their sex affords a peculiar incitement: arraigned, not merely as writers, but as *women*, their characters, their conduct, even their personal endowments become the subjects of severe inquisition. In detecting their errors and exposing their foibles. malignant ingenuity is active and unwearied (*Public Characters, Vol. IV* qtd. in Adburgham 1972, p. 7).

Edgeworth was already an established author by 1810 having received £300 for *Belinda* in 1801. Her success as an author was emphasised when she received £1050 for the second series of *Tales of Fashionable Life* in 1812 and £2100 for *Patronage* two years later in 1814 (Garside and Schöwerling 2000, p. 45). Owenson too proved that a woman could earn an independent fortune from writing; when she married Sir Charles Morgan, in 1812, "she had saved about five thousand pounds, the proceeds of her writings" (Morgan and Hepworth Dixon 1863b, p. 4). Publishing with Henry Colburn she received £550 for *O'Donnel* in 1814 and £1200 for *Florence Macarthy* in 1818, in addition to

[11] "Authorship might have provided the 'literate middle-class' with 'a substitute for the declining home industries which had once enabled the housewife to contribute to the support of the family', as Jane Spencer has observed, but to be a truly respectable writer, as Janet Todd has argued, an author was obliged either to obscure her financial motivations to publish, often by emphasizing her desire to tend to her readers' moral improvement, or to present her economic situation as so grave that her recourse to such an improper and immodest employment was unavoidable." (J. Batchelor 2010, p. 14)

the amounts secured for her poetry, music, non-fiction, and new editions of her earlier novels (Garside and Schöwerling 2000). In contrast, Austen made relatively little from her writing, an estimated £1600, half of which was paid to her family after her death. Austen even suffered losses: a second edition of *Mansfield Park*, which she published at her own cost, sold poorly resulting in "an initial loss of £182.8.3" (Gilson 1982, p. 60).

Success meant sales, and while both Edgeworth and Owenson's novels had gone through multiple editions by 1820, only Austen's *Pride and Prejudice* made it as far as a third edition. By 1820 Austen's novels were no longer in print. Without the interest of her brother Henry, who acted as her literary agent, and publisher Richard Bentley who bought the copyright to the novels, Austen may have joined many of the other novelists of this period in obscurity. Henry Austen, who had written a brief biographical notice to accompany the posthumous publication of *Northanger Abbey* and *Persuasion*, extended and amended it to accompany Bentley's 'Standard Novels' edition of *Sense and Sensibility* published in 1833. Adding to the earlier notice, Bentley includes "extracts from a critical journal of the highest reputation" highlighting that "Unlike that of many writers, Miss Austen's fame has grown fastest since she died: there was no éclat about her first appearance: the public took time to make up its mind; and she, not having staked her hopes of happiness on success or failure, was content to wait for the decision of her claims" (Gilson 1997, p. 16). Bentley's own note also served to emphasise Austen's skills as an author and placed her in the company of her more successful contemporaries:

> Miss Austen is the founder of a school of novelists; and her followers are not confined to her own sex, but comprise in their number some

> male writers of considerable merit. The authoress of 'Sense and Sensibility' had for her contemporaries several female novelists, whose works attained instant popularity – Madame D'Arblay, Miss Edgeworth, Mrs Opie, Miss Porter, and others (Gilson 1997, p. 16).

1.1.2 The Reviewers

The beginning of the nineteenth century marked a shift in the declared aims of the many review magazines in circulation. Laura Runge explains that during the eighteenth century:

> Female novelists saw their works reviewed by comprehensive journals that strove for both objectivity and universal coverage. The comprehensive project obliged reviewers to criticize novels by women however their own literary tastes or gendered ideologies might operate; so despite cultural imperatives for gallant condescension to women or assumptions about the catastrophic effects of novel reading, reviewers were bound to assess novels by women ... As a result of this confluence of publishing trends, we have multiple first-hand accounts of the reception and critical treatment of hundreds of early books by women. This would not be the case when the nineteenth-century Reviews adopted a highly selective format which virtually erased the presence of female novelists. (2009, pp. 276–277)

To be a female novelist in the early nineteenth century was not only putting work onto a public platform for criticism but also, to some extent (as seen

above), placing one's morals up for scrutiny. The reviews published at this time often represented very conservative political, moral or religious views. The reviewers often saw themselves "as guardians of the intellectual health of our readers" ("Florence Macarthy by Lady Morgan" 1819, p. 483). A 'victim' of the reviewers of the *Anti-Jacobin Review* was Mary Wollstonecraft, prompted by the publication of her husband's biography of her life (although by 1803, when the review was written, she had already been dead for six years), the critic damned her saying "Miss Mary lived as a courtezan ... and as a courtezan she is to be rated" ("Defence of Mary Wollstonecraft Godwin" 1803, p. 186). The conflation of Wollstonecraft's personal reputation with her work was enough to make her books and ideas unacceptable in polite society. This was the threat that female novelists faced if they chose to publish under their own name. The largely male attackers sheltered behind a cowardly facade of anonymity and the pretence that their comments were for the moral good of the nation rather than silencing ideas which did not agree with their own. In turn, the choice of anonymity added force to the critics' views, as Henry Whitfield states in the Preface of his 1804 novel, *A Picture From Life: or. The History of Emma Tankerville and Sir Henry Moreton*: "do we not see names to every thing estimable?" (1804, p. xv).

The reviewers not only positioned themselves as arbiters of taste and decency, but also as authority figures who could teach the wayward female writer. The shift in focus of the reviews were therefore not merely in response to the increased number of novels and texts being published by women, but also the challenge they potentially posed to those in power. Eighteenth century gender codes favoured professional male authors over "female or aristocratic

amateurism" (Runge 2009, p. 277):

> [The eighteenth century] reviews testify to a set of critical alternatives shaped by gender. Women, unlike men, are ostensibly entitled to gallant protection or critical leniency, and so, in keeping with the objectives of criticism, reviewers construct a matrix of feminine literary values that allows them to praise the woman in the work. The questions of character, plotting, style or intellect can be avoided altogether if the gallant critic can at least compliment the female writer on achieving a representation of the respected categories of womanhood. Sadly, this critical strategy survives the shift to the new format of reviews in the early nineteenth century. (Runge 2009, p. 279)

The approach of the nineteenth century review meant that, therefore, many of the reviews available seem to focus on the extremes — those 'approved' as safe for a largely female readership, and those worthy of censure. The reviewers anonymously voiced their opinions of the novels, and advice, both for the reader and the author, which ranged from the relatively benign to the vitriolic. Within this broad scope of reviews lie those of Austen, Edgeworth, and Owenson.

Austen's anonymity during her lifetime, and the relatively uncontroversial nature of her novels, meant that her work was less likely to be reviewed than the previously known Edgeworth, or the controversial Owenson. The *British Critic*, in its1812 review of Austen's *Sense and Sensibility*, having praised her characters for being "happily delineated", goes on to say "we will, however,

detain our female friends no longer than to assure them that they may peruse these volumes not only with satisfaction but with real benefit, for they may learn from them, if they please, many sober and salutary maxims for the conduct of life" ("Sense and Sensibility by A Lady" 1812, p. 527).

Assigned to the category of novels which were moral, uncontroversial and educational, Austen's works attracted little attention from the reviewers until a critical essay on *Emma*, written by Walter Scott (although published anonymously) appeared in *The Quarterly Review* in 1816. As Peter Sabor notes in his article about the significance of this review: "In taking notice of an obscure female novelist, and in commissioning a review from the greatest man of letters of his age, Murray was tacitly acknowledging the particular significance of *Emma*: only a very few of the hundreds of contemporary novels would ever be so favoured" (Sabor 1991). Scott used the essay to draw attention to Austen's previous novels (*Sense and Sensibility* and *Pride and Prejudice*, but not the less popular *Mansfield Park*) as well as indicating the superiority of Austen's works when compared to "the ephemeral productions which supply the regular demand of watering-places and circulating libraries" ("Art. IX. Emma; a Novel. By the Author of Sense and Sensibility, Pride and Prejudice, etc." 1816, p. 189). In addition, Scott's essay, commissioned by John Murray the publisher of *The Quarterly Review* as well as Austen and Scott's novels, also acknowledges Austen's almost unique position in novel writing at the time, proposing that her novels belonged to a new "class of fictions which has arisen almost in our own times, and which draws the characters and incidents introduced more immediately from the current of ordinary life than was permitted

by the former rules of the novel" (p. 189).[12] Scott informs his readers that *Emma*, although having "even less story than either of the preceding novels" (p. 195), has "subjects . . . [which] are finished up to nature, and with a precision which delights the reader" (p. 197). He praises her "quiet yet comic dialogue" (p. 199) but criticises "the minute detail" of her "[c]haracters of folly or simplicity" which "is apt to become tiresome in fiction" (p. 200). This review was critical in securing the future value placed upon Austen's novels in the late nineteenth century and beyond.

Edgeworth was already an accomplished and respected writer when she began writing novels. When *Belinda* was published in 1801, reviewers noted the shift from educational treatises, stating that they had been "called to notice a production of a different nature, but which may in reality be considered as designed to answer purposes somewhat similar, since the author offers it as 'a Moral Tale"' ("Belinda, By Maria Edgeworth." 1802, p. 368). This positioning of the novel as a morality tale was designed to mitigate the negative perception of books of this type, and, certainly for the critic at the *Monthly Review*, may have backfired. The critic acknowledges "respect for her talents" and calls her work the "production of no common pen" but goes on to say that although the novel starts well it fades into "tameness and insipidity", disappointed that the female duellist Lady Delacour has been "reformed" (p. 368).

Edgeworth's popularity drew notice from reviewers and public alike, both seemingly willing to overlook any perceived faults in her writing due to their

[12]This defence of the novel was not entirely lacking in self-interest as Scott had published *Waverley* and *Guy Mannering* anonymously in the previous two years.

educational and entertaining content; for example the 1804 review of her *Popular Tales*: "The design of these tales is excellent, and their tendency so truly laudable as to make amends for many faults of execution" ("Miss Edgeworth's Popular Tales" 1804, p. 329). The critic goes on: "the execution is extremely unequal. Many of the incidents are childish, and several of the stories unmeaning and improbable; yet they all inculcate an unexceptionable and practical morality" (p. 331). She elicited particular praise for targeting a group of readers who had commonly been overlooked, "that great multitude who are neither high-born nor high-bred" (p. 329), and, although "we cannot help observing that the execution is extremely unequal ... they [the tales] all inculcate an unexceptionable and practical morality ... There is nothing tawdry or sophisticated about them" (p. 331).

The publication of *Patronage* received a much less favourable review from *The British Critic and Quarterly Theological Review*. The reviewers admit that "[w]e can hardly recollect a novel which has attracted the attention of the public in so strong a degree" ("Patronage by Miss Edgeworth" 1814, p. 159), however felt that this attention placed a grave responsibility upon Edgeworth, and upon the reviewers themselves — that it was "the duty of those who claim any influence over the public mind to ascertain the grounds upon which its reputation is founded, and strictly to examine the probable consequences of its favourable reception on the taste or the morals of the age ... it is the office of criticism to expose the errors, and to detect the fallacy of such representations" (p. 159).

Edgeworth made the error of daring to set part of her novel beyond the perceived female sphere of the domestic and into the male, public sphere of law,

politics and the Church. This led the reviewers to patronise and chastise her in equal measure. Her understanding of diplomacy could only have come from "some ape of his superiors", and her descriptions of political characters were "absurd" and "raise an incredulous disgust" ("Patronage by Miss Edgeworth" 1814, p. 164). Rather more menacingly they "advise her, as she regards her own reputation, not to libel our English Church" (p. 168). This phrase concludes a series of comments regarding her morality ("To the morality of Miss Edgeworth we can raise no objection" (p. 160)) and her private life ("With the private lives of those whose works are before us we have not the slightest knowledge" (p. 168)). The message is clear: Edgeworth should only describe the machinations of the female sphere lest she be regarded as a bluestocking or a radical to the detriment of her reputation. Ultimately, in perhaps the greatest insult to the author, the reviewers conclude: "If we shall be thought to be severe upon those parts ... it is to be remembered, that it is not upon our ingenious and lively authoress that our censures rest so heavily, as upon that Father" (p. 173).[13] After all her work and success as an author, her work is essentially attributed to her father.

This review, however, was not left unanswered, although Edgeworth makes no direct reference to it or its anonymous authors. In her 'Preface to the Third Edition' of *Patronage*, dated 1st August 1815, she starts by confirming the popularity of her novel ("The public has called for a third *impression* of this book") before addressing aspects of the text which were "objected to as impossible or improbable" but in fact "were true" (M. Edgeworth 1815, p. v).

[13]This is perhaps in response to Richard Lovell Edgeworth's note "To the Reader" in *Patronage* where he "give[s] the sanction which she requires" (M. Edgeworth 1814).

However, it is for the criticism in the review, and the suggestion of her father's influence, that she reserves her most detailed response, taking full ownership of the novel and its ideas, and diplomatically yet irrefutably suggesting that the reviewers did not understand the purpose of a novel:

> It has been supposed that some parts of PATRONAGE were not written by Miss Edgeworth. This is not fact: the whole of these volumes were written by her, the opinions they contain are her own, and she is answerable for all the faults which may be found in them ... To fulfil the main purpose of her story it was essential only to show how some lawyers and physicians may be pushed forward for a time, without much knowledge either of law or medicine; or how, on the contrary, others may, independently of patronage, advance themselves permanently by their own merit. If this principal object of the fiction be accomplished, the author's ignorance on professional subjects is of little consequence to the moral or interest of the tale (pp. vi–vii).[14]

As with Edgeworth's novels, Owenson's works were very popular with the public and early reviews were relatively positive: "It has been said ... that the Novice of St.Dominic [sic], a former work of Miss Owenson, was the last book that amused the hours of illness of the late Mr Pitt" and "tho' we cannot speak of it [*The Wild Irish Girl*] in the first type of panegyric, is yet in many parts capable of exciting considerable interest, and may well amuse a leisure

[14]The attribution of her work may have been a sore point for Edgeworth; articles appeared in 1809 and 1812 refuting authorship of texts which had been published under her name, volume 4 of the third edition of *Patronage* includes a full list of her works (*Miss Edgeworth's Works* 1809; R. L. Edgeworth 1812).

hour" ("The Wild Irish Girl, By Miss Owenson" 1806, pp. 327-328).

However, Owenson, in her public response in 1804 to John Wilson Croker's anonymously published *Familiar Epistles, to Frederick J–S Esq, On the Present State of the Irish Stage* (which attacked many of her father's friends), made an influential, political enemy. In a series of letters to *The Freeman's Journal* in 1806 and 1807, Croker attacked Owenson and her writing: "her merits have been over-rated", he wrote under the pseudonym M.T., "and her arguments over-praised ... I accuse Miss OWENSON of having written bad novels, and worse poetry ... I accuse her of attempting to vitiate mankind — of attempting to undermine morality by sophistry" (qtd. in Connolly 2000, p. 98). This led to *The Wild Irish Girl* becoming something of a 'media event' and made novels "a matter to be discussed in the public press" (Connolly 2000, p. 103).

This war of words may have provided Owenson with more support than Croker anticipated; one particular letter writer, identifying himself only as the 'Son of Ireland' suggested that Croker's motive was jealousy and that "the puny pretender to wit is prompt to undervalue the talent that can detect his insufficiency" (qtd. in Connolly 2000, p. 104). Yet there are suggestions that Croker was not just attacking Owenson's work and political stance. He was also attacking her as a woman who sought financial independence and admittance to the higher ranks of society. "Croker makes it clear that had Owenson not been in search of commercial gain — as in her countertype, the independently wealthy and ultimately frivolous author ... her reputation might have remained her 'private property"' (Connolly 2000, p. 112).

The personal attacks, by Croker in *The Quarterly Review* and by other reviewers, often targeted Owenson's sex and challenged her right and authority to address the public. In its review of *Ida of Athens*, a book in which Owenson herself was disappointed calling it a "bad book" (Morgan and Hepworth Dixon 1863a, p. 352), the *Belfast Monthly Magazine* suggested she "relinquish the profession of author altogether, or subject her imagination to the discipline of meditation and reflection, which alone can qualify her for the honourable but arduous task of instructing her sex" ("Woman : Or Ida of Athens (Continued) by Miss Owenson" 1809, p. 224). Croker's review echoed this sentiment suggesting that if she "practise a little self denial, and gather a few precepts of humility . . . she might then hope to prove, not indeed a good writer of novels, but a useful friend, a faithful wife, a tender mother, and a respectable and happy mistress of a family" ("Woman: or , Ida of Athens" 1809, p. 52). The reviewers railed against a woman who lacked "a disposition to defer to the judgement and feelings of others" ("Florence Macarthy by Lady Morgan" 1819, p. 483) and failed to keep "within the prescribed limits of legitimate namby-pamby" (1819, p. 487).

As a Tory politician, Croker, may also have found Owenson's support of Whig politicians and the cause of Catholic emancipation a threat to the Tory party and his own emerging political career. Indeed, her political support of the Irish was memorably addressed in *The Wild Irish Girl* controversy in 1807: "her 'Wild Irish Girl' . . . is to be stripped and scourged for presuming to inculcate the moral of benevolence and extinction of sectarious differences" (qtd. in Connolly 2000, p. 104). Even favourable notices for her novels sought to mitigate their political nature: "It may perhaps be supposed that this work

[*O'Donnel*] will in some measure bear upon the present interesting question of Catholic Emancipation, but we understand that the political allusions are not stronger on this Work than in the 'Wild Irish Girl"' (*Thursday 17th February 1814* 1814, p. 4). Yet, it seems difficult to equate twenty years of increasingly vitriolic reviews solely to political differences.[15] Thomas Tracy, in his paper 'The Mild Irish Girl: Domesticating the National Tale', notes that "Croker equates the political and social vision she expresses with socio-sexual impropriety, and attributes her egalitarian 'promiscuity' to a root cause: Gaelic Irish anti-domesticity" (T. Tracy 2004, p. 82).

Croker seems to have relished the vicious attack. Owenson was not his only target (he also attacked Mary Shelly's *Frankenstein*, and the works of Hugo, and Dumas) but he seemed to have a particular animosity towards her. As is the case with modern 'trolls', the attacks frequently went beyond the content of her novels and other writings and attacking Owenson herself, almost always from the safety of anonymity. However, it was Croker's review of Owenson's travel book *France*, to which he dedicated 23 pages of the *Quarterly Review*, where the true depth of his dislike for the author was revealed. Croker's review calls the book a "worthless publication. . . composed in the most confused manner, and written in the worst style - if it be not an abuse of language, to call that a style". He goes on to say that "Our charges . . . fall readily under the heads of — bad taste — bombast and nonsense — blunders — ignorance of the

[15] Owenson's *Italy*, published in 1821, as well as her husband, was savaged in the *Quarterly Review*: "Notwithstanding the obstetric skill of Sir Charles Morgan (who we believe is a male midwife), this book dropt all but stillborn from the press" (qtd. in Adburgham 1972, p. 255)

French language and manners — general ignorance — Jacobinism — Falsehood — licentiousness, and Impiety" ("France, by Lady Morgan" 1817, p. 264). He describes her as "equally ignorant, on all subjects connected with the arts" and states that "Lady Morgan hardly knows, surprizing as such ignorance must appear, the difference between sculpture and architecture ... never lets pass the double opportunity of shewing her ignorance and irreverence for sacred things" (p. 284). The review closes by saying that he "shall not ... offend our readers by any further exposure of the wickedness and folly of her book" and he hopes that the review will "prevent in some degree, the circulation of trash which under the name of a Lady author might otherwise find its way into the hands of young persons of both sexes, for whose perusal it is utterly, on the score both of morals and politics, unfit" (p. 286).

Although Owenson's *Memoir* hints that Croker's dislike may have been due to a romantic rejection, one suspects that it was a combination of Owenson's politics, social climbing, refusal to be silenced, and her continuing success which caused much of Croker's animosity.

1.2 Independence

Introduction

Jane Austen's novels have a long tradition of being viewed as a reflection of the domestic focus of her own life, dating back to the preface to *Northanger Abbey* and *Persuasion* written by her brother Henry, and the memoir written by her nephew James Edward Austen-Leigh (Kirkham 1997; C. L. Johnson

2012). Deborah Kaplan refers to Austen living and writing within a world dominated by an "ideology of domesticity" (1994), while Janet Todd refers to the generic classification of Jane Austen's writings as "domestic fiction" (2005, p. 7); however, as Amanda Vickery notes, this is a simplification: "[t]he public/private dichotomy may, therefore, serve as a loose description of a very long-standing difference between the lives of women and men" (1993, p. 411). Austen's chosen focus, use of irony and lack of overtly political comment make her ideological views particularly challenging to interpret.

In her book, *General Consent in Jane Austen: A Study of Dialogism*, Barbard Seeber cautions that "Austen's focus on the domestic does not make her novels apolitical; for it is precisely the private matters that were the site of the ideological battles of the time" (2000, p. 5). While some commentators criticise Austen for her interest in the minutiae of rank, fortune and the marriage market, it is because wealth and marriage had such an impact on the independence or dependence of an individual that these areas are of such interest.

The shift in popularity of our three authors, with Austen's increasing, and Edgeworth, and Owenson's decreasing, was in part due to their chosen subject matter. Austen's focus on the seemingly uncontroversial and non-political marriage plot, contrasted with Edgeworth and Owenson's which often addressed "the question of social hierarchy and its legitimacy" which "concerned the relationships between three groups: the aristocracy and gentry; the larger tenants, particularly those of them who had lost gentry status in the seventeenth century; and the smaller tenantry and peasantry" (O'Ciosain 1997,

p. 170). Unlike Austen, Edgeworth and Owenson, although enjoying something of a resurgence of interest in recent years, have failed to capture the popular and critical imagination to the same extent. Sean Ryder provides a possible reason: "The qualities singled out by Leavis [in *The Great Tradition*] ... were noticeably absent in Irish fiction of the same century, dealing as it did with quite a different set of contexts and demands from those of Jane Austen ... For these reasons, the bulk of nineteenth-century Irish writing remained unexamined by critics, or was judged merely as historical background to the twentieth century" (Ryder 2005, pp. 120–1).

As discussed at the start of this chapter, the concept of independence includes: financial freedom, freedom of personal thought and expression, sovereign nationhood, as well as a sense of rebellion ("Lack of submission to rightful authority") (University of Glasgow 2015). While explorations of Austen's interpretation of independence have focused largely on the domestic and social, those exploring Edgeworth and Owenson have tended to focus on the political and national aspects of independence. Between them, Austen, Edgeworth and Owenson allow us to explore a broad range of difference, the gender difference that links the three authors, but also differences between them, through nationality, socio-economic position, and status as a writer. Each of these differences allow us to further understand the three authors as well as the time in which they lived. This section will examine three inextricably linked socio-economic elements of independence in the novels: education, wealth and work, and marriage and sexual conduct, before moving on to consider the concept of national independence.

1.2.1 Education

As noted earlier, in Section 1.1.1, the end of the eighteenth century and the start of the nineteenth marked a greater interest in the topic of education, and in particular the education of women. "[T]he writings and intellectual discourse generated by the Enlightenment made it one of the most influential forces shaping western society. It set the agenda for scientific, political and social thought for the eighteenth and nineteenth centuries" (Meaney, O'Dowd, and Whelan 2013, p. 1). Mary Wollstonecraft's *A Vindication of the Rights of Woman*, and other writers who could be categorised as early feminists, criticised the limitations of the traditional roles assigned to women, arguing the need to recognise them as equal partners both inside and outside the home (Wollstonecraft 2008). Catherine Macaulay challenged what she called "the absurd notion, that the education of females should be of an opposite kind to that of males" (1990, p. 112). Published during the early nineteenth century, many of Austen, Edgeworth and Owenson's novels reflect these concerns. Education of an appropriate nature underpins all three authors' world views - to improve and flourish in society, whether at a national, social or personal level, requires an appropriate education, and this is something that we often see lacking in the characters presented in the novels. Maria Edgeworth was at the forefront of this discourse on education, publishing a number of books and essays on the subject with her father and step-mother (*Practical Education* (1798) and *Professional Education* (1809)), and advocating the importance of science education for girls in *Letters for Literary Ladies* (1795).

This shift in thinking was reflected in much of the literature of the time and

represented not just a call for the education of women, but a redefinition of women beyond the world of men:

> In place of the intricate status systems that had long dominated British thinking, these authors began to represent an individual's value in terms of his, but more often in terms of *her*, essential qualities of mind. Literature devoted to producing the domestic woman thus appeared to ignore the political world run by men. Of the female alone did it presume to say that neither birth nor the accoutrements of title and status accurately represent the individual. (Armstrong 1987, p. 4)

Although an accepted area for female involvement, the educational theories and debates of the period were highly politicised, representing "critiques of the conduct-book construction of femininity" (V. Jones 1990, p. 98), and "the key medium through which women can be empowered in the public sphere" (Ross 2014, p. 378). Women's education could be contentious, and discussions surrounding what girls *should* be taught, or even *whether* they should be educated, had been a popular topic for much of the previous century.[16] Austen, Edgeworth and Owenson's novels frequently focus on the importance of education for women, either directly or indirectly — there is a clear message, that preparing women for nothing more than being decorative wives was a dangerous practice. At the core of many of these arguments was the idea

[16]Gerardine Meaney, Mary O'Dowd and Bernadette Whelan point out that only translations of Classical texts were deemed suitable for a female audience (2013). This may have been to limit women's access to the language of the professions (medicine and law) or, perhaps because some of the texts were of an erotic or obscene nature. For example, Catullus' poem Carmen 16 was considered so obscene that it was not published in translation until the late twentieth century (Winter 1973).

that women should be viewed as equal to men: "That learning belongs not to the female character, and that the female mind is not capable of a degree of improvement equal to that of the other sex, are narrow and unphilosophical prejudices" (Knox 1990, p. 109).

Maria Edgeworth's early writings in the 1790s, with her father and stepmother, were firmly positioned within this discourse, engaging with the ideas of more radical writers like Wollstonecraft (Topliss 1981, p. 15). Her first published work, *Letters for Literary Ladies* (1795), treads a careful line between the "patriarchal norms that locate female subjectivity exclusively within the private sphere" and creating "a public agency for women writers by using its masculinist discourse as a strategic deployment" (Narain 1998, p. 283). This combination of conservative radicalism runs through many of her other works, demonstrating "her opposition to the ideology of separate spheres and her support for female education without the 'unwanted radical overturning of the social order' that the marriage of Glorvina and Horatio [in Owenson's *The Wild Irish Girl*] represents" (Ross 2014, p. 386).

Jane Austen indicates her criticism of the type of education received by many women in her novels, for example through her description of the female inhabitants of *Mansfield Park*. The Bertram sisters and their cousin receive a very different education, even within the same house. Through brief comments as to the content of this education, it is evident that Maria and Julia are receiving what was viewed as a fashionable education of the type recommended by Hester Chapone in 1773:

> With regard to accomplishments, the chief of these is a competent

> share of reading, well chosen and properly regulated; and of this I shall speak more largely hereafter. Dancing and the knowledge of the French tongue are now so universal that they cannot be dispensed with in the education of a gentlewoman; and indeed they both are useful as well as ornamental; the first, by forming and strengthening the body, and improving the carriage; the second, by opening a large field of entertainment and improvement for the mind. I believe there are more agreeable books of female literature in French than in any other language; and, as they are not less commonly talked of than English books, you must often feel mortified in company, if you are too ignorant to read them. Italian would be easily learnt after French, and, if you have leisure and opportunity, may be worth your gaining, though in your station in life it is by no means necessary. To write a free and legible hand, and to understand common arithmetic, are indispensable requisites. As to music and drawing I would only wish you to follow as Genius leads: ...it is of great consequence to have the power of filling up agreeably those intervals of time, which too often hang heavily on the hands of a woman, if her lot be cast in a retired situation. (1990, pp. 104-5).[17]

[17]It is possible to infer that some of Austen's attitudes, and indeed those of her family, towards this type of education came from the periodical magazine *The Loiterer* created by her brothers James and Henry while at Oxford. In a mocking criticism of the aims of fashionable education they write: "As soon as she can understand what is said to her, let her know that she is to look forwards to matrimony, as the sole end of existence, and the sole means of happiness; and that the older, the richer and the foolisher her Husband is, the more enviable will be her situation...By minutely pursuing this system of Prudence, She will reap praise and pleasure in every station of life; She will be an accomplished Coquette, and a Successful Gamester; she will be an unfeeling Daughter, a Childless Wife, and a tearless Widow" (James Austen and H. Austen 1789)

Austen is clear to highlight the deficiencies of such an education.[18]

The three authors present a clear argument for women to receive a broad education to enable them to achieve a greater level of independence, echoing Priscilla Wakefield's 1798 *Reflections on the Present Condition of the Female Sex; with Suggestions for its Improvement*:

> The necessity of directing the attention of females to some certain occupation is not so apparent, because custom has rendered them dependant upon their fathers and husbands for support; but as some of every class experience the loss of those relations, without inheriting an adequate resource, there would be great propriety in preparing each of them, by an education of energy and useful attainments, to meet such disasters, and to be able, under such circumstances, to procure an independence for herself. There is scarcely a more helpless object, in the wide circle of misery which the vicissitudes of civilized society display, than a woman genteelly educated, whether single or married, who is deprived, by any unfortunate accident, of the protection and support of male relations

[18]The curriculum suggested by Chapone had become the target of satire by at least as early as 1848. This is evident in William Thackeray's presentation of Amelia Sedley's six years of education at a fashionable school in *Vanity Fair*:

> In music, in dancing, in orthography, in every variety of embroidery and needlework, she will be found to have realized her friends' *fondest wishes*. In geography there is still much to be desired; and a careful and undeviating use of the backboard, for four hours daily during the next three years, is recommended as necessary to the acquirement of that dignified *deportment and carriage*, so requisite for every young lady of *fashion*.
>
> In the principles of religion and morality, Miss Sedley will be found worthy of an establishment which has been honoured by the presence of *The Great Lexicographer*, and the patronage of the admirable Mrs. Chapone.(1848, p. 3)

.

(1990, p. 123).

Austen also demonstrates her willingness to engage with this perspective on education, challenging which feminine attributes should be valued in society. In *Pride and Prejudice*, Nancy Armstrong notes that "[w]hile excelling on none of the traditionally feminine qualities ... Elizabeth surpasses them on an entirely different plane. Her particular assets are the traditionally masculine qualities of rational intelligence, honesty, self-possession, and especially a command of the language" (1987, p. 50).

The criticisms which focus on the gaps in female education, are relatively obvious. However, less obviously, we are also presented with men who have had the benefit of education at expensive schools and universities and yet appear to lack a clear sense of responsibility. It is perhaps not surprising that much of this criticism is aimed towards the hereditary male landowner, "the miseducation of women ... revealing the frequent miseducation of upper-class men" (McAllister 2004, p. 105). As part of this critique, the authors also challenge the concept of authority, by presenting the schooling of high status men by lower status women, for example Horatio by Glorvina in *The Wild Irish Girl*, or Darcy by Elizabeth in *Pride and Prejudice*, reinforcing the definition of women's worth beyond "birth [and] the accoutrements of title and status" (Armstrong 1987, p. 4). Patrick Fessenbecker demonstrates that this is an indication of power within the relationship: "[f]or Darcy, this is an educational realization: he learned his 'lesson' entirely because the one woman in his life who could teach him, the woman whom he recognized as a superior or a 'woman worthy of being pleased,' taught the lesson" (2011, p. 759).

1.2.2 Wealth and Work

While the novels of Austen, Edgeworth and Owenson are traditionally viewed as having romantic marriage plots, a closer look reveals that concerns over money, inheritance and employment are a clear focus within the texts. This is indicative of the societal changes which had led to the increase in the reading public, including "larger numbers of readers from the middle class, especially women, reflect[ing] a change in the distribution of power and wealth from an aristocratic and landed minority to those whose interests lay in a mercantile economy" (Botting 2005, p. 31). The authors may not all directly criticise the society in which they live. However the criticism is there. We see a large proportion of society, mostly but not entirely women, who are dependent upon the whims of a few, defined and controlled by "what patriarchal society values — wealth, position and influence" (Russell 2010, p. 481).

Independence in the novels is linked to wealth and property. Dependence, something that affects both male and female characters, is generally linked to a lack of these. Some of the worst excesses of the 'wicked' characters are prompted by a lack of full independence. Willoughby's treatment of Marianne, in *Sense and Sensibility*, is in part prompted by his lack of independent wealth, and his dependence on a wealthy female relative. In *Patronage*, Buckhurst Falconer's choice of profession — the clergy — is made partly out of financial necessity. Both men's subsequent marriages are for money and are unhappy ones.

Economic independence, or its lack, has been a recurring and popular theme in Austen, Edgeworth and Owenson's novels, and is without doubt both a

highly politicised and a gendered topic. The status of women and its link to money "in a society in which one's value, indeed, one's very definition depended on class position — or, more visibly, on the rewards of money", has encouraged critics to explore the broader social and political context of wealth and position (Poovey 1984, p. 109). Richard Handler and Daniel Segal state that "By focusing on the ongoing negotiation of status, the novels suggest the ability of fictional narratives to illuminate political and social debates by exploring rather than ignoring cultural ambiguity and diversity" (1985, p. 691). A society populated by women who "had been 'trivialized', totally lacking in power and independence" and who were " the financial responsibility of the male with whom [they] lived", has provided fertile ground for scholars (M. A. Schofield and Macheski 1986, p. 187).

A woman's independence could make her a target for men seeking money. In *Sense and Sensibility* Colonel Brandon's first love becomes a commodity due to her wealth, and her inability to leave the marriage with any income leads to her downfall. A lack of income could also place a woman in a very unpleasant position, facing the prospect of working as a governess like Charlotte O'Halloran, in *O'Donnel*, or Jane Fairfax, in *Emma*.[19]

[19]The lack of options for women, and the possible alternatives, was highlighted by Dr Dodd in a sermon to the governors of the Magdalen Society:

> so scanty are the means of subsistence allowed the female sex; so few the occupations which they can pursue, and those so much engrossed by our sex: so small are the profits arising from their labours, and so difficult often the power of obtaining employment, especially for those of doubtful character; and frequently, so utter their unskilfulness in any branches of their common industry, from a mistaken neglect of their parents in their education ... So scanty are the means of subsistence, arising from these and the like causes, that, it is but too well known, many virtuous and decent young women, left desolate with poor unfriended children, have been compelled to the horrid necessity (and we want not to be told, what numbers in this great city lie in

Money and property leads to security and power, the ability to make choices and the freedom to express opinions. Austen and Edgeworth in particular emphasise the importance of the responsibilities of wealth and property: "The heirs of these families are commended for good estate management and concern for their tenants and the poor, and condemned for extravagance and materialism ... The heirs to landed estates have a job to do in the present in Austen - and are judged harshly if, like Sir Walter Elliot in *Persuasion*, they fail to do it" (Lamont 2003, p. 665).

In the novels of Austen, Edgeworth and Owenson, we get hints of the lives of those who do not inhabit the relative comfort of their main characters. We see glimpses of the 'poor' who rely upon the good will of the rich to visit and provide some comfort (for example in *Emma*). We see those who have fallen from polite society like the abandoned farmer's daughter Kate Robinson (in *Patronage*). In the novels, the need for someone with independence to act on behalf of those without is a repeated theme. In *Persuasion*, Mrs Smith's lack of money or male relatives to act for her leaves her poor and alone. In *O'Donnel* McRory highlights that it is not merely a lack of money that causes difficulties: "for surely it makes all the diffir if a man have a *wote* or have not a wote; that's when he gets into a *scrimmage*: what compensation did I ever get for my poor brother ... who was kilt in a ruction, because I'd no gentleman to back me, having ourselves neither wote nor interest, and being *Romans* to boot" (Owenson 1814a, p. 106). These minor characters hint at a world beyond the big houses and parlours so often associated with

> wait to improve, and turn to their own advantage that necessity) of procuring bread by prostitution! (1990, p. 89).

the novels. Far from reinforcing the belief that the upper classes had an hereditary right to power and property ownership, Austen, Edgeworth and Owenson present a world in which new money and professions seem to be a much more positive driving force behind the success of the country than the influence of the traditional aristocracy.

Since Marilyn Butler's seminal 1975 work *Jane Austen and the War of Ideas*, scholarship has focused intently upon the socio-economic context in which novels were written (1997). Judith Lowder Newton highlights "The privilege that belonged to middle-class men ... access to work that paid, access to inheritance and preference, and access to the independence, the personal power, that belonged to being prosperous and male"(1978, p. 27). She suggests that Austen' s decision to challenge these ideas is an act of rebellion and subversion, an idea echoed by Sandra Gilbert and Susan Guber in their groundbreaking feminist study of women's literature: *The Madwoman in the Attic* (2000). Gilbert and Guber view women writers as "enclosed in the architecture of ... male-dominated society" and as being part of "a distinctively female literary tradition", being within, but separated from the broader, male literary tradition (p. xi). In the 1987 introduction of the new edition of her book, Butler agrees that "the femino-centric novel ... might function as, at the very least, a system within a system" and that "the Austen woman's submissiveness now reads poignantly as the sign of externally imposed constraint", although she questions whether Austen was indeed subversive (1997, p. xxxii).

The work of Mary Poovey, Mary Ann Schofield and Cecilia Macheski comments on the act of being a woman writer during the eighteenth and nineteenth centuries (Poovey 1984; M. A. Schofield and Macheski 1986). They highlight

that freedom of expression and the ability to publish required a certain level of autonomy, or at least support. While some women novelists certainly felt free to express their ideas, others relied upon the relative safety of anonymity. Schofield and Macheski also question whether these writers were "free . . . to develop economic independence"(1986, p. 1). Austen certainly made only a limited amount of money from her novels, enabling her to supplement the meagre income she, her sister and her mother survived on after the death of her father, but not sufficient to support herself as an independent woman.

1.2.3 Marriage and Sexual Conduct

The explicit link between marriage and money in Austen, Edgeworth, and Owenson's novels enables the reader to see not only the necessity of having 'sufficient means' but also the potentially negative impact of marrying solely for money. With marriage at the heart of continuity for the ruling classes in the eighteenth and nineteenth century, Austen's representation of marriage and her intentions for presenting the institution as she does has caused some debate. Beth Tobin presents her as "an apologist for the landed classes . . . defending the 'paternal system of government' from attacks . . . that challenge the hereditary right of the gentry and aristocracy to the exclusive monopoly of the land" (1990, p. 229), a view supported by James Downie (2010). In contrast, Butler claims Austen depicts "the existing marriages of the previous generation, and the heroine's struggles before marriage, in such unflattering terms that the genre could be said to hold the entire institution

to scrutiny" (1997, p. xxxii). Lisa Hopkins suggests that, in fact, Austen's novels represent a move from the old to the new:

> This view of marriage [of the Bertrams at the start of *Mansfield Park*] as a market to which the woman brings the ready cash and the husband the steady income, derived presumably from land, certainly sits well with views of gender relations and money management that would suit a very traditional gentry pattern in Jane Austen's day; but interestingly, by telling us so firmly and so early in the narrative that this happened thirty years ago, she can also be seen as signalling to the old-fashioned nature of this particular kind of transaction and, equally, a shift in values . . . it is . . . another pointer that money has to some extent been superseded as an index of marital eligibility by the less tangible markers of status and class. (1994, p. 77)

Sarah Emsley goes further still in her discussion of the nature of marriage in the eighteenth and nineteenth centuries and the impact of Hardwicke's Marriage Act of 1753, by stating that Austen's representation of equality in marriage "offers a unique and radical vision of marriage" (1999, p. 498).[20]

The importance of marriage in maintaining the social and political status quo

[20]Hardwicke's Marriage Act of 1753 changed the law which had previously allowed marriages to "take place anywhere provided they were conducted before an ordained clergyman of the Church of England". Instead "all marriage ceremonies must be conducted by a minister in a parish church or chapel of the Church of England to be legally binding . . . [and n]o marriage of a person under the age of 21 was valid without the consent of parents of guardians" (*Relationships* n.d.). This law had a particular impact on clandestine marriages, and meant that an eloping couple would need to go to Scotland, where marriage without parental consent was still legal. This explains the Bennet family's (in *Pride and Prejudice*) consternation that Lydia and Wickham had remained in London rather than travelling to Scotland.

became an increasingly fraught political topic. The control of women and the potential seditious effects of their education and reading habits, on the importance of procreation within the bounds of marriage for the sake of the nation, positioned women as a potential threat which needed to be controlled. Although marriage could offer some form of independence for women it could become little more than a prison, as marriage meant surrendering one's 'self', in a legal as well as in a religious sense, to one's husband, a conflict which Austen, Edgeworth and Owenson explore in the novels. In effect, an unmarried woman was viewed almost as the property of those responsible for her, a point Austen parodies in the opening of *Pride and Prejudice* by applying it to single men of fortune; it was also a point Sydney Owenson seems to have felt in Lady Abercorn's organisation of her marriage to Dr Charles Morgan (Morgan and Hepworth Dixon 1863b).

The concept of equality between the sexes was, of course, contentious, as it offered a direct challenge to the traditions of patriarchal society and the role of women within the domestic sphere. If women were equal, and if marriage was a partnership rather than a hierarchy, the arguments which prevented women taking part in a broader public and political life became redundant.

The important function of marriage within the structures of society, and in the transference of land and power, emphasised the importance of sexual conduct, in particular the stark differences between the treatment of women and men who transgressed societal norms. Catherine Macaulay, writing in 1790, suggests possible reasons for this difference:

> I shall intimate, that the great difference now beheld in the external consequences which follow the deviations from chastity in the two sexes, did in all probability arise from women having been considered as the mere property of the men; and, on this account had no right to dispose of their own persons: that policy adopted this difference, when the plea of property had been given up; and it was still preserved in society from the unruly licentiousness of the men, who, finding no obstacles in the delicacy of the other sex, continue to set at defiance both divine and moral law, and by mutual support and general opinion to use their natural freedom with impunity (1990, p. 115).

Although Macaulay comments that the status of women as property was something relegated to the past, in fact the legal situation until the Married Woman's Property Act of 1870 meant that married women were not considered to have a legal entity beyond that of their husband.[21] They were classed as 'feme covert'.[22] Any money made by a women, from any source, automatically became the property of her husband.The legal divorce proceedings, with which the sexual conduct of married women was regulated, positioned the married woman as property of her husband and therefore any seduction or act of adultery was in fact an act against the husband's property. A husband could sue for damages by claiming 'criminal conversation' against his wife's

[21]The Married Woman's Property Act of 1870, later replaced and extended by the Married Woman's Property Act 1882, allowed a wife to keep any wages or property earned or inherited as her own.

[22]In contrast a 'feme sole' had the right to own property and be considered a person in her own right in the eyes of the law.

lover.[23] If a divorce was granted the blame was generally attached to the woman involved.

In *Mansfield Park*, Mrs Rushworth's affair, subsequent disgrace and divorce is presented through the eyes of a number of characters allowing Austen to explore a range of attitudes towards a married woman's sexual misconduct. Edmund and Fanny feel the full weight of the moral impact of this transgression; as the more religious and pious characters their horror is clearly outlined. This reaction is in clear contrast with that of Mary Crawford who suggests a marriage and the gradual reintroduction of Maria into society, although even she accepts that Maria would not be welcome in all society. The final breach between Edmund and Mary is her failure to comply with the moral expectations of society, particularly important to Edmund in his role as a member of the clergy. Sir Thomas's responses are more complex, reflecting his different roles within the family and within society at large. He makes arrangements to support his daughter for life, and is highly critical of his own actions which, inadvertently, led to her fall. However, as a respected member of society and someone with responsibility for his two unmarried nieces, he cannot allow her to reside in the same area, nor does he admit her into the family. Austen uses this to emphasise that while Maria's actions have lasting consequences for her, the same cannot be said for Henry Crawford. Austen presents her own position, highlighting the inequality of social reaction to the couple: "That

[23]Until the Matrimonial Causes Act of 1857 a full divorce allowing remarriage was only available through a Private Act of Parliament, and therefore was the preserve of the rich. Although the Act moved divorce from the ecclesiastical courts to the civil courts, allowed legal separation, and abolished adultery as a criminal offence, it still favoured men as while adultery was sufficient cause for a man to divorce his wife this was not sufficient for a woman to divorce her husband.

punishment, the public punishment of disgrace, should in a just measure attend his share of the offence is, we know, not one of the barriers which society gives to virtue. In this world the penalty is less equal than could be wished" (1994b, p. 474).

The taint of divorce did not just affect the women directly involved. It also affected their children, specifically daughters, even if there was no doubt as to the legitimacy of the child. A tradition of viewing the sins of the parents as influencing those of the child was still commonly held. In *Patronage*, Edgeworth's comments, through the character of Mrs Percy, are rather critical of the daughter of a divorcee. However the focus is on whether the child remained with the divorced mother.

> She said that she thought the important point to be considered was, what the education of the daughter had been; on this a prudent man would form his opinion, not on the mere accident of her birth ... If such had been the case, Mrs. Percy declared she thought it would be imprudent and wrong to marry the daughter. But if the daughter had been separated in early childhood from the mother, had never been exposed to the influence of her example, had, on the contrary, been educated carefully in strict moral and religious principles, it would be cruel, because unnecessary, to object to an alliance with such a woman. The objection would appear inconsistent, as well as unjust, if made by, those who professed to believe in the unlimited power of education. (1814, pp. 132–133).

Edgeworth uses the chapter to outline a number of views regarding Miss Hauton, the daughter of the divorcee Lady Anne Hauton. Godfrey is sympathetic towards her, helped in part by her good looks and pleasant character. However the overall sentiment is clear: the prior bad deed of the mother affects her daughter, in particular if she has remained in her mother's custody. Edgeworth emphasises the importance of the education of the daughter and the influences she has - revealing her somewhat naive belief in the roles of environment and education in the construction of a person's moral code. Edgeworth returns to the character of Miss Hauton later in the novel where she is under threat of divorce herself - the conclusion seems clear that even though she was manipulated into a marriage with a man she didn't love, the bad example of her childhood led to her fall.

In *Patronage* Edgeworth also presents an encounter between Caroline and Rosamond Percy, and Kate Robinson, a young girl who has been seduced and abandoned. Kate is living with her child outside society in a run down cottage, having been rejected by her lover and her family: "Cast off by all, shamed, and forgotten, and broken-hearted, and lost as much as if she was in her grave. And better she was in her grave than as she is" (1814, p. 353). She is reduced to abject poverty with only an old woman, Dorothy, for company. From Dorothy the Percy sisters and the reader learn that Buckhurst Falconer is the young girl's seducer and that he has failed to send any means of support. Buckhurst's actions are criticised by Caroline and Rosamond, but their mother and Edgeworth take the opportunity to highlight the challenges of openly criticising the seducer:

> If, like your sister Caroline, young ladies would show that they

> really do not prefer such men, it would do essential service. And observe, my dear Rosamond, this can be done by every young woman with perfect delicacy: but I do not see how she can, with propriety or good effect, do more. It is a subject ladies cannot well discuss; a subject upon which the manners and customs of the world are so much at variance with religion and morality, that entering upon the discussion would lead to greater difficulties than you are aware of. It is, therefore, best for our sex to show their disapprobation of vice, and to prove their sense of virtue and religion by their conduct, rather than to proclaim it to the world in words. Had Caroline in her letter expressed her indignation in the most severe terms that the English language could supply, she would only have exposed herself to the ridicule of Mr. Buckhurst Falconer's fashionable companions, as a prating, preaching prude, without doing the least good to him, or to any one living. (1814, pp. 363–364)

Edgeworth's understanding of the double standards in play is clear, although Kate dies penitent yet still an outcast. Edgeworth leaves the final word to Dorothy: "A bank note of ten pounds was received by Dorothy soon afterwards for the use of the child, and deep regret was expressed by the father for the death of its mother. But, as Dorothy said, 'that came too late to be of any good to her."'(1814, p. 373).

Austen, however, appears rather more sympathetic to her 'fallen' female characters. In her early novel, *Sense and Sensibility*, we see this through the unfortunate character of Eliza, Colonel Brandon's sister-in-law and first love.

Austen presents Eliza in a sympathetic manner, emphasising the manipulation of her guardian and the cruelty of the man she is forced to marry. Although she is seduced and 'sinks' we feel pity rather than disgust, and she is redeemed to a certain extent through her death, a common trope which runs through literature of this period and the Victorian period.[24] Eliza's daughter in turn falls for the seducer Willoughby, but, perhaps to show the passage of time or to indicate her sympathy for the young girl's situation, Austen chooses not to visit the same fate upon her. The criticism instead is for her seducer, especially in his final encounter with Elinor.

Perhaps the most emphatic challenge to Edgeworth's view is revealed in *Emma* through the character of Harriet Smith, "[t]he natural daughter of somebody" (1994a, p. 23). Austen uses Emma to demonstrate a less critical and more accepting attitude to a poor and illegitimate girl, although tempered somewhat by Emma's romantic ideals. Mr Elton and Mr Knightley are used to reveal societal attitudes towards women of Harriet's status. Austen does not allow these male characters' attitudes to stand. Mr Elton is shown to be a cruel snob who marries for money. However, more importantly, it is Mr Knightley who is shown to have reconsidered, suggesting that not only will Harriet make Robert Martin an excellent wife, but that Emma "would have chosen for him [Mr Elton] better than he has chosen for himself" (1994a, p. 240). The suggestion that an illegitimate woman could be viewed as an appropriate wife for a clergyman emphasises Austen's belief that a person's worth is not defined

[24]There is a combination of moralising and pity for the woman as is evident in the final section of *Oliver Twist* "I do believe that the shade of that poor girl often hovers about that solemn nook — ay, though it *is* a church, and she was weak and erring." (Dickens 1839, p. 315)

by their background.

This double standard is highlighted by Wollstonecraft's choice of quotation, from Matthew Prior's 'Hans Carvell', regarding women and fate:

> If weak women go astray,
> The stars are more in fault than they. (Wollstonecraft 2008, p. 102)

a quotation also used by Edgeworth in *Belinda* and by Owenson in her *Memoirs*. This raises the point that, while society believed women to be mentally weaker than men and that therefore they should not be educated as rational beings, this was contradicted by the treatment of the fallen woman, as blame attached to her rather than the man involved.

1.2.4 National Independence

A core aspect of independence, hinted at through the meaning 'lack of subjugation to rightful authority' (*The Historical Thesaurus of English* 2018), crosses the boundaries between the domestic and the national. The concept of nationality creates a sense of belonging to a homogenous group, and therefore a sense of identity and place within the larger world. Linda Colley explains that British identity was created by "war — recurrent, protracted, and increasingly demanding ... regular conflict with the Other in the shape of Catholic France" (Colley 1992, p. 322). In effect, the creation of a British and Irish identity has, at its root, a link by which each is defined by its difference from the other, and as such was problematic when incorporating Ireland

into the United Kingdom without a considerable restructuring of the national identity to include Catholics as citizens.

An "understanding of the nation and national identity as self-contained, cohesive, clearly-delineated spaces has become increasingly difficult to sustain ... the notion of national identity as an ongoing process of negotiating identity and difference that takes place both within and beyond the borders of single nations ... practices of rewriting and remapping cultural and political landscapes demand an awareness of the constant displacement, erasure, and redefinition of what constitutes the global, a nation, a national culture, an ethnic group, or a community" (Pfeiffer and Garcia-Moreno 1996, pp. vii–viii). We see evidence of this 'rewriting and remapping' of landscapes, both metaphorical and actual, in the novels of Edgeworth and Owenson, indicating the "confusion and misunderstanding" inherent in defining Irish culture (Connolly 2003, p. 2).

The continuing threat of rebellion, and the enactment of the Act of Union, in January 1801, which formally removed Ireland's ability to self-govern and therefore removed its political independence, led to a pressing desire to define Ireland and Irishness in order to avoid being culturally subsumed. The negotiation and renegotiation of national identity is evident in the work of many writers during this period (McCaw 2004). It is not surprising then, that this period marks calls for an Irish national literature and discussions regarding an Irish canon. Joseph O'Carroll (1877) highlighted the link between canon creation and the definition of nationhood and independence, saying: "every free nation with any claim to be considered civilized makes the possession of an independent literature one of its first objects" (qtd. in Kelleher and Murphy

1997, p. 17). Although Justin McCarthy (1847) warns against an attempt to combine literature with "any great social grievance arising from legislature", in the end it is next to impossible to separate the two (qtd. in Kelleher and Murphy 1997, p. 16). "Among the first hesitant steps towards a national oeuvre was Maria Edgeworth's *Castle Rackrent* ... soon afterwards followed by Miss Owenson's *St Clair*" (Loeber, Stouthamer-Loeber, and Leerssen 2004, p. 13).

With the freedom of the individual being tied closely to freedom of the state, it is no surprise that Wollstonecraft's *A Vindication of the Rights of Woman* was written in the first flush of optimism following the French Revolution. It is equally no surprise that, following the execution of King Louis XVI and Marie-Antoinette and the Reign of Terror, that such ideas became intrinsically linked to those of bloody revolution. The role of women's writing within this, and in particular sentimental novels, was decried in periodicals of the time like *The Anti-Jacobin Review*. Nicola Watson proposes that "the interconnection between the 'dissolute' discourse of the sentimental novel and the adoption of 'vicious refinement' — a term which here implies revolution in the state as well as indiscipline in the family — displays in miniature the intimate relation between sentimental fiction and radical politics" (1994, p. 1).

National identity was perceived as linked to the role of women within society, on a broader scale. Britain became viewed as masculine and Ireland as feminine and it was within that context that we see the rise of the National Tale as a genre: "The national tale deals with national themes, but its discourse is not 'nationalist' — at least not in the sense the nineteenth century gave to the word. The genre as a whole seems to locate itself not within

the boundaries of a nation, but in the site of a 'hyphenated culture', and, at the same time, to question existing hierarchical relationships between nations and cultures" (Lacaita 2001, p. 148). Esther Wohlgemut, in her study of 'Maria Edgeworth and the Question of National Identity', emphasises Edgeworth's importance saying of Edgeworth that "through her interrogation she reinterpreted both cosmopolitan and national definitions of belonging so as to reconstitute 'Anglo-Irish' less as a category than as an ongoing mediation between borders" (1999, p. 645).

Sydney Owenson, in fact, was the first writer to use the subtitle 'A National Tale' for her novel *The Wild Irish Girl* in 1806 (although Edgeworth had subtitled *Castle Rackrent* as 'An Hibernian Tale' in 1800), leading a popular trend in novels written between 1808 and into the 1840s. Despite this, because of Owenson's "reputation for Whig 'radicalism' " Sir Walter Scott makes no mention of her or her influence in his "carefully constructed genealogy of the Waverley novels" (Garside and Schöwerling 2000, p. 61).[25] In *The Wild Irish Girl*, Owenson's "most famous" novel and "the one to which scholars have been consistently drawn", Owenson creates a first person narrator, Horatio, who narrates the majority of events through a series of letters (Nersessian 2011, p. 340). Beyond the surface narrative of a flawed hero overcoming his prejudiced view of Ireland, we have a broader and more complex text. The narrator reveals his growing knowledge of, and sympathies towards, Ireland through its embodiment in the character Glorvina. As Ina Ferris explains:

[25]Katie Trumpener discusses Scott's 'constructed genealogy' of the Waverley novels, and explores the novels and stylistic developments which actually influenced his work in her article 'National Character, Nationalist Plots: National Tale and Historical Novel in the Age of Waverley, 1806–1830' (1993).

'[a]s a worldly and impure genre that sets out to do some-thing with words, the national tale makes central to its whole project the often obscured, performative notion of representation itself" (Ferris 1996, p. 288). In the case of *The Wild Irish Girl* the representation of Ireland revealed by the narrator is expanded and referenced through the author's extensive footnotes. In this manner, Owenson literally marginalises herself, removing her own comments to the footnotes of the text. However, by doing so she subversively positions herself as an expert authority on Ireland and history (a traditionally male role, as lamented by Jane Austen in *Persuasion*).[26]

"Sydney Owenson's solution to the Irish problem can be considered as modern, because on the one hand she champions the cause of Irish identity, and on the other hand she advocates the reconciliation between the English and the native Irish through marriage" (Zeender 2008, p. 15). This is something we also see in some of Edgeworth's novels, for example *The Absentee*, although her relationship to the Irish and the national tale was more complex. In the preface to *Castle Rackrent* (published in 1800 with the subtitle 'An Hibernian Tale'), Edgeworth reveals a very different attitude towards cultural identity to that championed by Owenson: "Nations as well as individuals gradually lose attachment to their identity, and the present generation is amused rather than offended by the ridicule that is thrown upon their ancestors" (2008b, p. 5). However, it is worth noting that Edgeworth is writing as a member of the Anglo-Irish community and the identity being lost is not her own.

Any discussion of nation and nationalism in the novels initially appears to

[26]In Austen's *Persuasion* Anne Elliot says: "no reference to examples in books. Men have had every advantage of us in telling their own story. Education has been theirs in so much higher a degree; the pen has been in their hands" (1994d, p. 235).

exclude Jane Austen, writing as she was far from the borders of England. However a closer reading reveals evidence of a nation at war and under threat. The marriage plots of *Pride and Prejudice* are set against the movements of the Army in Meryton, Brighton and 'the North', suggesting the scope and national significance of the threats faced. *Mansfield Park* uses the Price family to underline the ongoing risk of naval action, and the less obvious threats to the Empire, through slave revolts, are represented by Sir Thomas Bertram's visit to Antigua. Austen's last completed novel, *Persuasion*, has "the characters ... repeatedly refer to the peace of their present times, scrupulously marked as running from late summer 1814 to 1815 ... The historical orientation of the novel — at once displayed and concealed — tells us that the supposed peace informing it will have been a false peace, a knowledge that is oddly suspended in the considerable gap between the characters' perceptions of their historical moment and the readers' " (Rohrbach 2004, p. 748).

More subtly, as perhaps the most significant feature of the English nation, we have the English language itself. In his article, ' "Manoeuvring" in Jane Austen', Brian Southam notes the significance of the word 'manoeuvring' when used by Austen in *Mansfield Park* and *Emma* to signify trickery: "[i]t was seen as a technical loan-word, appearing in English books, but still carrying the nationality of its origin and its associations with warfare" (2004, pp. 466–7). He goes on to point out that "this national aspect is underlined in Maria Edgeworth's half-apology, very early in *Manoeuvring*, for having to call Mrs Beaumont a *'manoeuvrer'*" (p. 468).

Henry Tilney, in *Northanger Abbey*, presents a definition of the English nation and values, challenging the 'foreign' influence of gothic novels:

> Remember the country and the age in which we live. Remember that we are English: that we are Christians ... Does our education prepare us for such atrocities? Do our laws connive at them? Could they be perpetrated without being known in a country like this, where social and literary intercourse is on such a footing, where every man is surrounded by a neighbourhood of voluntary spies, and where roads and newspapers lay every thing open? (Jane Austen 1994c, p. 182)

Henry's challenge to Catherine's gothic imaginings emphasises the 'normality' of England, and yet also presents a slightly disturbing image of constant surveillance and hyper-vigillance at a time of national and international unrest.

Brian Southam places Austen firmly within the genre popularised by Owenson and Edgeworth by proposing that *Emma* is in fact a satirical take on the national tale, with "Englishness ... as a central theme" (2008, p. 189). Austen's chosen focus is England rather than "the so-called 'Celtic fringe' of Scotland, Ireland, and Wales", the centre rather than the periphery (p. 189). The Englishness Austen writes about is reinforced by literature and Christianity, at a time when the concept of a United Kingdom was tenuous. Instead of the exotic 'other' we are presented with the every day: "English verdure, English culture, English comfort" (Jane Austen 1994a, p. 272), the 'English' manners of Mr Knightley contrasted with the 'French' manners of Frank Churchill.

Claire Connolly states that the "desire to seek out new perspectives on Ireland too often retains - or rather is itself retained and restricted by - a narrow

understanding of the subject, the terms set by political and economic rather than social or cultural history" (2003, p. 2). The inclusion of an exploration of the social and cultural aspects of Edgeworth and Owenson's novels, in addition to the traditional 'political and economic' focus, and a consideration of, and comparison with, the political, economic and national in the novels of Austen will therefore add greater depth to the analysis of all three novelists.

2 Enhanced Reading

> Elinor agreed to it all, for she did not think he deserved the compliment of rational opposition
>
> Jane Austen, *Sense and Sensibility*

The rapid development of computer technology, including improvements in storage capacity and processing power over the past twenty years and more, has enabled the creation of large scale digitisation projects. One example is the Google Books scanning project, which allows access to large corpora, some far larger than a single scholar could read in a lifetime. This has necessitated the development of new tools and methods of computer-aided reading of texts. However, the shift in reading from print to digital, and the impact this has on reading and comprehension, has caused deep concern within education (Bauerlein 2005; National Endowment for the Arts 2005; Du 2009; Hayles 2010; Goodwyn 2012). It is not surprising, therefore, that many of these concerns have found their way into literary studies prompting "the need to re-examine what it is we, as literary critics, do" (Anderson 2013, p. 126), and leading to a reframing of reading. Although, as Matthew Kirschenbaum explains "[r]eading is not so much 'at risk' as in the process of being remade, both technologically and socially" (2007).

Adam Hammond, Julian Brooke, and Graeme Hirst refer to the "significant

barriers to true collaborative work between literary and computational researchers", linking this to C. P. Snow's 'two cultures' which is described as "the rift — perceived and generally lamented — between scientific and humanities intellectual cultures" (2013, p. 1). They go on to suggest that "[m]any scholars describing the recent rise of the Digital Humanities ... have argued that it effects a reconciliation of the two alienated spheres, bringing scientific methodology to bear on problems within the humanities" (Hammond, Brooke, and Hirst 2013, p. 1). However, what instead appears to have happened is that the rift has been replicated within literary studies. As Andrew Goldstone points out "[t]he history of literary studies is primarily remembered as a narrative of conflicting ideas" (2014, p. 359). To some extent this is an understandable reaction to the 'otherness' of digital techniques which are so alien to traditional English departments.

The following sections consider 'close' and 'distant' reading, and propose a theoretical framework which may provide a means to satisfactorily combine the two.

2.1 'Close' and 'Distant'

Close reading is something of a paradoxical term, at once understood by scholars and acknowledged as being at the heart of literary studies (DuBois 2003; Jockers 2013a; B. H. Smith 2016), while also, to some extent, seeming to defy a specific definition (DuBois 2003; Hayles 2010; B. H. Smith 2016). The "seemingly nonspecific concept of close reading" (DuBois 2003, p. 3) is difficult to position as a methodology as "the term *methodology* suggests something

more coherent, circumscribed, and specifically research focused" (B. H. Smith 2016, p. 57). Barbara Herrnstein Smith notes that "[f]or all the shifts they reflect, every one of the approaches . . . — from the New Criticism, and indeed from the old historical philology, through deconstruction, to ideology critique — involved reading individual texts closely" (2016, p. 58). She adds that "the term *close reading* refers not only to an activity with regard to texts but also to a type of text itself: a technically informed, fine-grained analysis of some piece of writing" (2016, p. 58).

Distant reading is a potentially nebulous term as Andrew Goldstone indicates:

> the meaning of *distant reading* has undergone a rapid semantic transformation . . . Franco Moretti introduces the phrase to describe "a patchwork of other people's research, *without a single direct textual reading*" (*Distant Reading* 48). Today, however, *distant reading* typically refers to computational studies of text . . . Matthew K. Gold and Lauren Klein employ the term to speak of "using digital tools to 'read' large swathes of text" . . . Ted Underwood embraces "distant reading" as a name for applying machine-learning techniques to unstructured text. (2017, p. 636)

Although the application of computational analysis to literary texts has a history dating back to at least Roberto Busa's 1949 collaboration with IBM on the works of Saint Thomas Aquinas (Busa 1980), scholars such as Katie Trumpener (2009) and Stanley Fish (2012) have argued vociferously against the use of digital techniques, claiming that they are unnecessary, and that

"computational analysis ... is incompatible with the aims of humanistic research" (Hammond, Brooke, and Hirst 2013, p. 1). While not all computational techniques actually fall under 'distant' reading, within the criticisms of distant reading there is a tendency to conflate all digital techniques. For example, Stephen Marche, in 2012 , wrote that:

> Literature cannot meaningfully be treated as data. The problem is essential rather than superficial: literature is not data. Literature is the opposite of data ... The process of turning literature into data removes distinction itself. It removes taste. It removes all the refinement from criticism. It removes the history of the reception of works (2012).

In computational analysis, the choice of tool is often explained as a necessary reaction to scale. If we want to understand world literature (which Franco Moretti is discussing when he first uses the term 'distant reading') it is not possible to read and understand everything, even if there were sufficient time. When commenting on large scale quantitative literary studies, David A. Brewer states that they "remind us of just how broad and varied the literary field of the past actually was, and what a small fraction of it receives scholarly attention of any sort" (2011, p. 161). He goes on to say that it is "accompanied by a sobering reminder that our customary modes of investigation are simply not up to the task of really grasping this broadened field and its forms" (p. 161). Large scale studies seek to explore and understand the broad sweep of literary history — the development of the novel and its genres over time, for example — but this comes "at a cost. In order to be countable (and so graphable or otherwise capable of being traced over time), texts

have to be treated as if they were comparable units" (p. 162), in other words they need to be reduced to a common denominator. This is something that Moretti acknowledges in his description of distant reading: "where distance ... is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes - or genres and systems ... If we want to understand the system in its entirety, we must accept losing something" (Moretti 2013, pp. 48 – 49). The problem for many within literary studies is that there can appear to be a somewhat blasé attitude to what is discarded, and it is important to consider the impact of what is lost.

However, while the large scale study may need to strip back the texts to certain key metrics, this is not the case when considering a medium sized corpus such as the one explored in this thesis.[1] The *middle distance* is a profitable area for exploration as it enables the texts it be considered in their own right, as well as part of a broader corpus. To shed light on texts which fall outside the traditional canon, and compare them with canonical texts, allows us to understand more about the texts, authors and the contexts in which they were written. It is dangerous to assume that texts which have become accepted as part of the canon were the more popular, or more accomplished. The construction of any literary canon reflects the values of those authority figures who are involved in its construction. In the early 20th century this meant often reinforcing a view of literature as, largely, white, male, and English.[2]

[1]Large scale, for the purpose of this thesis is understood to be several hundred texts or more.

[2]This favouring of male writers is evident even in the creation of a specifically Irish canon. Although Edgeworth and Owenson originally "received positive endorsements" (Loeber, Stouthamer-Loeber, and Leerssen 2004, p. 19) they were eventually passed over in favour of the male writers who now take pride of place on the Dublin tourist posters.

The perception for some, it seems, is that to *do* distant reading, one needs to reject close reading and vice versa. While an exaggeration, this antagonistic relationship between close and distant reading may be explained in part by Herrnstein Smith:

> If, as seems to be the case, the practices of close reading have operated in literary studies not as one method among others but as virtually definitive of the field, then how are we to understand a method whose advocates define it in opposition to — and, indeed, as superseding — precisely those practices? (B. H. Smith 2016, p. 58)

Lisa Marie Rhody observes that "most debates over distant reading weigh its merits against that of close reading" (Rhody 2017, p. 660). This can be seen in Matthew Jockers' *Macroanalysis* where he argues that "interpretation is fuelled by observation, and as a method of evidence gathering, observation ... is flawed" (2013a, p. 6). He goes on to suggest that "cherry-picking of evidence in support of a broad hypothesis seems inevitable in the close reading scholarly tradition" (2013a, p. 47). While he acknowledges that "[t]he two scales of analysis, therefore, should and need to coexist ... The literary researcher must embrace new, and largely computational, ways of gathering evidence" (2013a, p. 9), the implication seems to be that the scholar trained in the new methods has no need to compromise.

The lack of a concrete definition of either close or distant reading raises the question: aside from the use of digital tools, how different are the two methods? Herrnstein Smith explains that "attention to microfeatures does not

distinguish the practices of close reading from those of distant reading or the digital humanities" (B. H. Smith 2016, p. 71). Given that both close and distant readings are interpretations of a text or texts which may be:

- 'technically informed'
- 'fine-grained analysis'
- considerations of 'microfeatures' or larger elements

it is not surprising that distant or macro readings define themselves in opposition to the concept of close reading itself rather than the indistinct elements involved.

It also appears that, although the language of distant and macro analysis is littered with imperatives against close reading (Wilkens 2000; Moretti 2005; Saklofske, Clements, and Cunningham 2012; Jockers 2013a), many scholars acknowledge that both methods are needed, and they may be viewed, more usefully perhaps, as points on a continuum. Moretti himself is positioned at the more extreme point on the continuum, distancing himself from close reading, even when it appears that his work would fit Herrnstein Smith's definition of "a technically informed, fine-grained analysis" (2016, p. 58). Moretti argues that "[t]he question, here, seems to be this: should every careful study of words and sentences be considered a version of close reading? Personally, I don't see why" (Moretti 2017, p. 687).

Jockers attempts to separate *analysis* from *reading*: "[t]he former term, especially when prefixed with *macro*, places the emphasis on the systematic examination of data, on the quantifiable methodology. It de-emphasizes the

more interpretive act of 'reading"' (2013a, p. 25). This separation seems unnecessary and may well be misleading, as Johanna Drucker explains: "some scholars mistakenly assume that these computational methods of analysis are objective in contrast to the individuated and situated practices of human reading and interpretation. This objective fallacy is problematic. Designing a text-analysis program is necessarily an interpretive act" (2017b, pp. 630-631). Andrew Goldstone develops this point further: "'Network Theory, Plot Analysis', the last essay in *Distant Reading.* closes with pleas for 'sharing raw material, evidence — *facts*' (240). But that essay, like 'Style, Inc.' and much of the work since, instead makes its case by interpreting data visualizations ... as clues to a cultural totality. These are not facts but readings" (2017, p. 641).

Katherine Bode challenges both Moretti and Jocker's use of *facts*: "Moretti, however, does not share his data. Jockers occasionally publishes the results of data analysis ... but he does not provide the textual data analyzed, even at the level of word frequencies, and he is significantly less open than Moretti about the sources and composition of his datasets" (2017, p. 6). While there may be significant reasons for a lack of transparency, for example copyright or technical challenges such as difficulty of migrating corpora between databases, or political reasons why a scholar may be reluctant to share work which currently remains unrecognised within the academy, this adds to the criticism of distant techniques within the humanities.

2.2 Beyond Close Reading?

In her article 'The Achievement of Scholarly Authority for Women: Trends in the Interpretation of Eighteenth-Century Fiction', Toni Bowers comments that:

> Widespread interest in a new subject or method often builds from an initially slow response to new research that only gradually comes to influence other scholars' work. Pioneering scholars republish previously out-of-print primary texts; deploy innovative interpretative methods toward unlikely textual subjects; or produce provocative rubrics for previously overlooked or dismissed categories of writing, writers, textual production, or readers. Where there has accrued what we might call a critical mass of this kind of groundbreaking research - enough to suggest a significant body of previously obscured work *and* to demonstrate the value of recovering and reading it *and* to suggest appropriate methods for interpreting it — scholarly focus shifts, first to recognize the existence of the new object or method of study, then to take it fully on board. (2009, p. 52)

It would appear that the point has been reached where the 'critical mass' has built to the point where digital techniques are a recognised method, but have not yet reached the point of full acceptance within the literary studies community. At present, the arguments regarding what these techniques may show, and what their value is, have not been satisfactorily answered for those

for whom this type of analysis seems alien and clinical — something this thesis seeks to address.

In reality, the digital offers us a new perspective, an alternative way of reading and interpreting texts, in a similar way to the advent of critical theory.[3] These new lenses through which we can read and interpret a corpus seek to add to our understanding of the texts, not replace previous understandings (Allison et al. 2011; Mueller 2012; Anderson 2013; Clement 2013; Jänicke et al. 2015). To borrow from Clifford Geertz's concept of culture as "webs of significance" which become "a context ... within which [social events, behaviours, institutions, or processes] can be intelligibly - that is, thickly - described" (1994, p. 220), new ways of reading aim to increase what we can say about a text or corpus rather than merely reduce them to a series of sterile numbers. What these methods *do* enable is the ability to step back from the tight focus of close reading and to consider the texts from another angle. The challenge these techniques face is how to quantify, visualise, and interpret these 'webs of significance' between text, context and meaning.

Although there is an acknowledgement that both close and distant methods are needed, a clear sense of a combined method is lacking in these polarised arguments. Rhody points out that:

> Switching readerly perspectives from close to distant to everywhere

[3]Thirty years before Marche's polemic against the digital humanities' approach to literature, an article was published with the title 'Against Theory' which stated "The theoretical impulse, as we have described it, always involves the attempt to separate things that should not be separated ... Our thesis has been that no one can reach a position outside practice, that theorists should stop trying, and that the theoretical enterprise should therefore come to an end" (Knapp and Michaels 1982).

> in between will more likely create networks of reading and representation — communities of selection and refiguration that combine several approaches. To move beyond distant reading, we must learn to create and to occupy many new critical distances — not by throwing distant reading out but by adding new models that demonstrate how we might do it differently. (2017, p. 666)

What digital methods allow is for at least part of the process of analysis to be visible, something that is a relatively new concept in literary studies. This thesis therefore proposes a shift in focus from close and distant reading, and suggests instead considering the combined use of these methods as *enhanced reading*, which has the purpose of further improving the quality and extent of close readings of the texts through the use of digital methods. A core part of this redefinition is the importance of greater openness with regard to the texts, tools, and the parameters used. This provides the opportunity of making textual analysis a little more replicable, allowing the researcher to repeat a particular analysis, and view the source of an interpretation. John Guillory notes that "it always was and always will be possible for other scholars to read what we miss ... and thus arrive at an interpretation that challenges our own. Scholarship is founded on the virtual infinity of the readable textual universe, and that is the nature of the game" (2008, p. 13) (a point echoed more recently by Herrnstein Smith in 2016). It is impossible to create the 'ultimate' interpretation of a text (as David Lodge's character Morris Zapp hopes to do in *Changing Places* (1975).[4]) but what we can strive for is for greater understanding and more detailed or nuanced interpretations of texts.

[4]Lodge notes that the "object of the exercise, as he [Zapp] had often to explain with as much patience as he could muster, was not to enhance others' enjoyment and understanding

A possible reason why digital techniques are so appealing is that they share metaphors which already exist within literary studies. The spatial metaphors used by literary scholars include distance, scale and depth: "metaphors and analogies usually do not rely on the transfer of a single predicate but instead on the transfer of an entire realm or field of concepts . . . [t]he major interest of such metaphorization is the heuristic power that comes with describing and re-describing a lesser-known concept in the light of a better-known experience or insight" (Van Acker and Uyttenhove 2012, p. 261). While individual interpretations differ, the conceptual framework created by these metaphors through which we categorise readings remain constant. These spatial analogies are seen across domains of knowledge suggesting that "spatial concepts provide a foundation for more abstract ideas" (Cooperrider, Gentner, and Goldin-Meadow 2016, p. 2) and that "it remains plausible that spatial analogies lie at the core of the human ability to understand complex relational phenomena" (Cooperrider, Gentner, and Goldin-Meadow 2016, p. 14). This is supported by Alan Baddeley's model of working memory in which he highlights the role of the 'visuo-spatial scratch pad' (Baddeley 1999).

The familiarity of these spatial analogies provides a potential middle path through the contentious landscape of literary studies. It may be possible to consider these differing methods of reading in the context of a broader spatial analogy: a map. Just as digital techniques have allowed us to access maps in a variety of ways — zooming in and out, or applying and removing layers as suited to purpose — so a combination of close and distant techniques allow

of Jane Austen, still less to honour the novelist herself, but to put a definitive stop to the production of any further garbage on the subject" (1975, p. 35)

the same, enabling the creation of enhanced readings of a text or collections of texts. A map will reveal different features depending on the scale and type of map, but still remains a representation of a physical place — even artificially manipulated, topographical maps, like that of the London Underground, are still recognised as being maps.[5] The challenges lie in how to understand, use, and interpret what the different maps reveal. Thus, *Enhanced reading* also raises the necessity of a shared language and consistent methodological approach with regard to the tools used and their output.

A considerable challenge is that currently code, corpora, and databases are not part of recognised scholarship in disciplines which still privilege the monograph over everything else. Reluctance to make the products of this unrecognised work freely available is, therefore, understandable. However, for this to change, and for scholars to start being recognised for this labour intensive background work, it is necessary to share. If the work remains invisible, it is unlikely to ever become recognised, and, without seeing and understanding the techniques used by others, the necessary shared language and consistent methodological approach is unlikely to develop.

[5]There is some discussion in Geospatial Science regarding whether the London Tube Map ought to be regarded as a map or a diagram (Cartwright 2012).

2.3 Visualisation

The significance of visual and spatial metaphors within digital literary studies seems to lead logically to the importance of visualisation. However, visualisation, and its role within literary studies is part of a much larger methodological discussion which Andrew Goldstone refers to as "the *doxa* of reading: the assumption that the primary activity of academic literary study is textual interpretation ... For literary scholars, *doxa* is the belief in the inherent value and meaning of the literary object" (2017, p. 637). He argues that distant reading has evolved to have two forms — one which is a type of reading, and one which attempts to be something else. Essentially, Goldstone is challenging Moretti's assertion that images created "by visualizing empirical findings ... constitute *the specific object of study of computational criticism*; they are our 'text'; the counterpart to what a well-defined excerpt is to close reading" (Moretti 2016, p. 3). For Johanna Drucker (as discussed in Section 2.1), the issues lie closer to the *how* of reading. She argues that computational analysis deals with the *literal* text and therefore ignores the interpretative act and the interpreted text (2017b). Drucker's concern is that some scholars fail to recognise that the decisions made when choosing tools and methods are interpretative in nature and therefore subject to the same biases as traditional textual analysis. Before addressing the specific concerns raised here, it is perhaps wisest to take a step back to consider the nature of visualisation in a broader context.

2.3.1 The Purposes of Visualisation

Visualisations may be divided into those for knowledge communication and those for exploratory purposes (information visualisation). The link between the visio-spatial metaphors outlined above, and the visual nature of memory emphasises an important consideration when applying computational analysis to texts — humans are far better at pattern recognition than computers.[6] Gestalt theories of perception suggest that humans perceive patterns by proximity, similarity, continuity, closure, and connectedness, and have the ability to apply multiple aspects simultaneously. In addition, when considering patterns relating to nineteenth century novels, the expertise of the reader may identify additional patterns between words, texts, or authors, which would be difficult to encode for the benefit of a computer. While a computer may present what *is* (subject to the input of the reader), it is for the reader to decide the interpretation: the 'eureka' moment of interpretation is a particularly human trait. The difficulty faced within the humanities, as noted by Drucker (2017b), is that the use of visualisation is often imported wholesale from other disciplines, and that the difference between communicative and exploratory visualisations is not always made clear.

The core difference between knowledge communication and information visualisation relates to the differences in purpose, whether the visualisation is an end product or part of a larger, exploratory process. Knowledge visualisation "examines the use of visual representations to improve the transfer of

[6]Human prowess in pattern recognition explains why so many websites use captcha asking the reader to identify specific features in a series of images. While this confirms that the viewer is human, this is also vital information for machine learning models.

knowledge between at least two persons or group of persons" (Burkhard 2004, p. 520). It has the purpose of communication and is outwardly facing. Therefore it requires a range of practical considerations regarding focus, content and style. In contrast, information visualisation is part of an ongoing process and may be viewed as more experimental and iterative: "information visualization aims to explore abstract data and to create new insights" (Burkhard 2004, p. 520). Michael Friendly defines it as "the science of visual representation of 'data', defined as information which has been abstracted in some schematic form, including attributes or variables for the units of information" (Friendly and Denis 2009, p. 2). This is the crux of Drucker's argument: through the process of visualising a text it must be first viewed as data and then 'abstracted', each step requiring some aspect of interpretation - for example, what is classed as a word or phrase, or what is deemed important enough to visualise (2017b).

Information visualisation has its roots in John Tukey's 1977 seminal text, *Exploratory Data Analysis*, in which he introduces exploratory data analysis (EDA): "a practical philosophy of data analysis which minimizes prior assumptions and thus allows the data to guide the choice of appropriate models" (Velleman and Hoaglin 2004, p. xv). This has become an almost essential stage in examining data sets. Visualisation is a key aspect of EDA, described as "[d]isplays" (also called "revelations" (Hoaglin 2003, p. 313)) which "visually reveal the behaviour of the data and the structure of the analyses" (Velleman and Hoaglin 2004, p. xv). Essentially, the purpose of EDA is to enable the researcher to familiarise themselves with the data set, note areas of interest, and identify potential problems with the data which may lead to skewing of

the results. However, this type of analysis is merely a starting point for more detailed analysis and it may be misleading to present visualisations created for this purpose as 'facts' (Goldstone 2017), or to present them without detailed explanation.

Hsuanwei Michelle Chen explains that there are "two important aspects to consider: (1) information visualization is used to discover new insights and knowledge from abstract data through graphical means; and (2) information visualization can be considered a representation of data that amplifies cognition" (2017, p. 6). This perhaps explains the difficulty in identifying the sole purpose of a visualisation as it is likely to serve several purposes simultaneously. Chen outlines three subprocesses in the creation of a visualisation: "the data transformation ... which transforms raw data into data tables that offer structure and ease of manipulation. Second, data tables are mapped onto visual structures that include the application of spatial substrates, marks, and graphical properties ... the third subprocess, view transformations, visual structures are transformed into human views, which involves graphic parameters, such as position, scaling, and clipping" (2017, p. 6).

The level of manipulation required to transform data into a visualisation, whatever the purpose, has led Drucker to argue that data should be referred to as 'capta' to emphasise their artificial nature, "[a]n information visualization is not a statement, not simply a presentation of 'ideas' any more than any other articulation" (2017a, p. 903). She goes on to propose that visualisations should be read as enunciations, concluding that:

> The idea that information visualizations can be understood as

> enunciative expressions pushes us to attend to the specific ways graphical features articulate subject positions. The semantically meaningful features of graphical forms — organization, hierarchy, labels, grids, directional view, and spatial systems of perspectival or orthogonal structure — all play a part in encoding ideological beliefs of omniscience, neutrality, power, and powerlessness in visualizations ... Visualization *is* enunciation, and its principles of action — its graphical acts — are everywhere apparent, fully legible, and yet rarely subject to critical analysis. We need to attend to how visualizations do their work if we are to fully understand the work that they do. (2017a, p. 913)

In the analysis of large scale data, visualisations, whatever their perceived drawbacks, are an important tool: "[e]ffective visualisations are vital when trying to synthesise and communicate intangible (multi-dimensional, multi-scale, complex or abstract) information to audiences unfamiliar with the subject matter" (Grainger, Mao, and Buytaert 2016, p. 315). As explored above, the creation of visualisations is a process in which decisions are made at a number of points. "A successful visualisation is able to make smart comparisons, show causality and present multivariate data in a manner that exposes useful information such as size, direction or position. Ideally a visualisation should be able to highlight aspects of the data that were not visible before" (Featherstone and Poel 2017, p. 4). Grainger et al. suggest that "[e]very visualisation, to varying degrees, conveys data and allows users to play with and scrutinise underlying information". They continue that while "some academics have tried to embed different definitions within visualisation, it is perhaps more helpful

to imagine a continuum with presentation and provision of answers at one end, and exploration and the raising of questions at the other" (2016, p. 301).

In her article 'Why Distant Reading Isn't', Drucker cautions that "[t]hese visualizations are representations, elaborately constructed expressions following legible conventions through a series of interpretative decision points that are all concealed in a final statement that passes itself off as a presentation ... The specific semantics and rhetoric of visual epistemological systems are underanalyzed, and the elaborate processes of mediation and remediation are generally overlooked. We are reading (in the cognitive-hermeneutic sense) the artifacts of a process as if they are the actual phenomena. The visualizations are assertions read as declarations" (2017b, p. 633). What Drucker's work highlights is the importance of an open and shared methodological approach which addresses the concerns raised and makes explicit the decisions which lie behind the visualisations. This means not only openly acknowledging the subprocesses involved, as outlined in Chen's article mentioned above (2017), but also raising the profile of the currently unseen coding processes which underpin these subprocesses. As noted above, to develop the methodologies of data analysis and visualisation within the digital humanities requires a shift in what is recognised and rewarded as scholarship. In this thesis, both knowledge communication and information visualisations will be used.

2.3.2 Visualisation in Digital Humanities

Despite the warnings noted above, visualisation has been an integral part of digital humanities since at least the 1990s. While "[c]omputer assisted text

analysis has over fifty years of history in providing tools to help scholars interpret texts" (Sinclair and Rockwell 2007, p. 199), the focus in the past twenty years has increasingly been on how to visualise these analyses (Havre, Hetzler, and Nowell 2000; Paley 2002; Sinclair 2003; Don et al. 2007; Paranyushkin 2011; Scrivner and Davis 2017; Park et al. 2018).

The technological difficulties in creating accessible tools for visualising textual data are often beyond the skills of humanities scholars, even those with some programming experience. This leads to the necessity of inter-disciplinary collaboration; however this too is a "balancing act" as Stefan Jänicke notes: "a crucial issue for us visualization scholars is the *balancing act* to generate valuable contributions for both the visualization and the digital humanities fields due to the diverse needs ... A further crucial issue is that many digital humanists have a prior humanities background, so that the application of complex algorithms and sophisticated visualization techniques — which meet the requirements of a grounded computer science publication — fails as prototypes are likely to be perceived as black boxes" (2016). However, this gap between the disciplinary requirements of visualisation studies and digital humanities is not just in one direction: "a visualization is typically not the end of a research process in digital humanities workflow, and ... the *humanistic gain of knowledge* often comes later. Although a visualization is capable of triggering hypotheses generation, a humanities scholar will [be] unlikely [to] proof a hypothesis predominantly on an image" (2016).

The differences in interdisciplinary knowledge can also prove to be something of an interpretative challenge. In their article 'Learning to Read Data: Bringing out the Humanistic in the Digital Humanities', Ryan Heuser and Long

Le-Khac explain that they were "dumbfounded by a kind of text that for once we had no idea how to read" when examining plots created from a corpus of thousands of nineteenth-century novels (2011, p. 79). The uncertainty felt in using the emerging methodologies that the application of empirical methods make possible includes not just the question of "[w]hat do we do with this kind of evidence?" but also whether "[u]nder the flag of interdisciplinarity are the digital humanities no more than the colonization of the humanities by the sciences?" (Heuser and Le-Khac 2011, p. 79). While this thesis is not the place for yet another discussion of 'what is digital humanities?' (Kirschenbaum 2010; Berry 2012; Terras, Nyhan, and Vanhoutte 2013; Kirschenbaum 2014), this perception of the "scientification of the humanities" (Pawlicka 2017, p. 531) is reflected in the terms which have been adopted. Urszula Pawlicka explains that: "The scientification of the humanities has taken different forms, including reconceiving institutions (building humanities laboratories, or 'labs'), research materials (data instead of text), methods of conducting research (collective practices), applied methodologies (e.g., cultural analytics) and methods of evaluating research (parameterisation, falsifiability, the ranking system of universities). The belief that joins these forms is that the humanities must be empirical, pragmatic and measurable in order to resist its crisis and maintain its position" (Pawlicka 2017, p. 531). While Father Roberto Busa stressed a belief that the: "methods of observation used in the natural sciences should be applied together with the apparatus of the exact and statistical sciences" to the study of language (Busa 2004, p. xviii), the adoption of practices from the sciences has particular implications for use of data, and in particular use of visualisations, as discussed by Drucker (2011), and by Pawlicka:

> visualisation in the humanities, although it is presently used primarily as a presentation tool, can also be used as a tool to clarify, analyse and consolidate ... In the process, however, the 'data' component of data visualisation leads this suite of tools and techniques to not only *extend* our interpretive methods ... but also refine them, including that most venerable and honoured humanities term 'reading'. Once used primarily for written text, 'reading' now refers just as readily to 'data', 'algorithms' and 'patterns'. These different objects of reading require different methods of reading, all dependent on quantitative literacy. (Pawlicka 2017, p. 536)

There is currently no clear tradition or standard methodological approach within computational literary studies, and text-based humanities does not, generally, prepare scholars for interpreting numerical data of this type. A result of this methodological confusion is the prevalence of "three pervasive problems: anecdotal evidence, validation, and interpretation." (Heuser and Le-Khac 2012, p. 80). Ryan Heuser and Long Le-Khac highlight the danger that "a familiar concept is applied too hastily to the data, thus flattening the data's nuances and complexities" , going on to say that scholars need to trust the data: "If the digital humanities are to be more than simply an efficient tool for confirming what we already know, then we need to check this tendency to seek validation" (2012, p. 81). Effectively, this reinforces the need for an agreed and shared methodology for visualising data and interpreting visualisation within the digital humanities.

2.3.3 Austen, Edgeworth, and Owenson

A large proportion of the recent research into the computational analysis of textual data (also known as unstructured data) has focused on non-fiction texts, especially those generated by social media interactions. However, there has been a relatively small, but increasing, interest in applying these techniques to literary texts. Many of these studies have their genesis in the early work carried out in the field of corpus linguistics. The interest in applying computational techniques to the works of Austen, in particular, originate here (Burrows 1986; Burrows 1987; DeForest and E. Johnson 2001; Starcke 2006; Fischer-Starcke 2010). Texts from the nineteenth century have been a core focus for digital analysis. There is a number of reasons for this. Firstly, copyright is not attached to these texts allowing them to be uploaded, stored and manipulated freely. Secondly, the rise in popularity of novels and serials during this period mean that there is a large body of texts available, many of which have not been studied previously - victims of the 'slaughterhouse' ("if we set today's canon of nineteenth-century British novels at two hundred titles (which is a very high figure), they would still be only about 0.5 percent of all published novels." (Moretti 2000, p. 207)). Finally, unlike texts in the eighteenth century and earlier, there are fewer problems with OCR (Optical Character Recognition) readings of the texts as the 'long s' had started to fall out of favour from 1800 (Nash 2001).

While Austen is a popular choice for the application of computational methods, previous studies of Edgeworth and Owenson have only considered them as part of a larger corpus of nineteenth century novels, and therefore, when

considering previous work for the three authors, the majority focus on Austen. The sections below give an indication of the previous published work, which focuses on one or more of the three authors, using the four methods used in this thesis: frequency analysis, topic modelling, word embedding, and network analysis.

Frequency Analysis

Some of the earliest work in computational Austen scholarship was carried out by John Burrows, Mary DeForest and Eric Johnson, and Bettina Fischer-Starcke (Burrows 1986; Burrows 1987; DeForest and E. Johnson 2001; Starcke 2006; Fischer-Starcke 2010), perhaps setting a precedent for Austen's inclusion in computational studies.

Burrows' work in the late 1980s builds on traditional stylometric analysis, which uses word frequency to examine features of the author's style. Using Austen's novels as his focus, Burrows explored Austen's modal verbs, highlighting the difference between characters and between different forms of narrative within and between the novels (1986). Burrows uses word type frequencies to create a frequency profile for a text, which may also be applied to internal textual features, for example narrative and dialect, including the exploration of the idiolect of multiple characters. In contrast to much of the earlier work on frequency, Burrows focuses on the 'function words' such as 'to', 'the', 'and', as well as the more commonly studied 'content words'. Although his work largely falls under stylometrics his later studies go beyond these methods and are more "holistic in emphasis . . . makes no firm distinction

between 'function-words' and 'content-words', between the grammatical and the lexical elements of a vocabulary" (1987, p. 62). Burrows found that the frequency profiles for narrative, character narrative and dialogue map across Austen's six novels, concluding that the method "is able to distinguish, in appropriate degrees, between the narrative styles of different novelists and, likewise between those of the same novelist at different stages of her literary career. Only when the three Chawton novels and the fragment of 'Sanditon', all written in the last four years or so of Jane Austen's life, are compared with each other does the chronological dimension cease to afford a clear overall separation between the two clusters in each graph" (1987, pp. 67-68). Taking advantage of the emerging computational techniques available, he defined Delta as a measure of the difference of an individual text from the rest of the corpus (1987; 1989; 2002).

DeForest and Johnson's paper (2001) focuses on a tool to explore Austen's use of latinate words. They note that "[t]he English language is largely made up of two different language groups: Latinate and Germanic. The choice of a word from either vocabulary has rhetorical implications" (DeForest and E. Johnson 2001, p. 389). *Mansfield Park* proved a particularly interesting example as the use of Latinate phrases highlighted the contrast between the descriptions of Mansfield Park and the Price house in Portsmouth, as well as the occupants. However, a broader exploration of the themes raised by the superficial superiority of the latinate suggests that the veneer of importance the Mansfield party hold is masking flaws that the common Price family lack. DeForest and Johnson's study tagged Latinate and Germanic origin words and annotated the text using a markup language to indicate the speakers in the novels. This

enables the authors to identify the percentage of characters' speeches using each type of word, and reveals the complex structure of Austen's narrative and dialogue.

Fischer-Starcke (2006; 2009; 2010), Michaela Mahlberg (2010), and more recently E. H. Hubbard (2012) and Raksangob Wijitsopan (2013) also apply corpus linguistics techniques to the novels of Jane Austen. Fischer-Starcke's 2006 article (published as Starcke) examines phraseology in *Persuasion*. She comments on the importance of phrases, their frequency and their collocates to "reveal new shades of meaning" (2006, p. 87). She sees corpus linguistic techniques as supporting and extending more traditional readings and states that "the dominance of psychological words as collocates of the novel's most frequent phrase [*she could not*] indicates, as literary critics have frequently discussed, that the characters' inner lives are one of the major foci of the novel" (p. 94). In this article, and her later works, she emphasises a focus not just on the meanings created in the novels but also on identifying how these meanings are created.

Fischer-Starcke's book length study applies a corpus linguistics approach to literary texts, specifically Austen. Comparing Austen's novels with a corpus of contemporary literature from 1740 to 1859, including Edgeworth's (1800) *Castle Rackrent*, noting that the texts available were relatively limited: "the choice of texts results from their availability in electronic form and the fact that they are all novels. The latter makes them comparable to Austen's works" (Fischer-Starcke 2010, p. 28). Fischer-Starcke finds that keyword analysis and exploration of phraseology reveals insights into the role of daughters within the broader network of society. She also highlights the linguistic choices given

to the female characters which, through the use of hedging in conversation enable them to simultaneously appear to conform with societal expectations while also appearing as strong characters. An interesting point made in the conclusion of the book is that the techniques chosen to analyse a text need to be adapted depending on the corpus, emphasising the importance of avoiding a 'one size fits all' approach.

Topic Modelling

Topic modelling has increased in popularity over the past decade and has become a familiar method in the digital humanities, as indicated by an increase in the number of tutorials and texts which use Austen's novels as materials (Jockers 2013a; Jockers 2014; Silge and Robinson 2016; Silge and Robinson 2017). Its application to large corpora of nineteenth century novels has been explored in detail in Matthew Jocker's *Macroanalysis* (2013a). This book examines a broad range of nineteenth century novels, although the identity of all the texts which make up the corpus is unclear.[7] An exploration of the gender of authors (where it is known) highlights a divide between male and female novelists, suggesting that female writers may have largely concentrated their writing on a more limited range of topics and genres, perhaps because of criticism when they did not, as outlined in Chapter 1. As a broader exploration into gender, Jockers examines whether it is possible to determine an author's gender by exploring a range of features present in their works - in a

[7] As Katherine Bode comments "he does not provide the textual data analyzed, even at the level of word frequencies ... and he is significantly less open than Moretti about the sources and composition of his datasets. Indeed, I have discovered only one instance where Jockers indicates the titles and authors he investigates, and then only 106 of the 3,346 (identified in the context of reporting confusion matrices)" (2017, p. 6).

similar manner to the exploration of author attribution. Perhaps unsurprisingly, Austen is clearly defined as 'female'. However, what is interesting is that both Edgeworth and Owenson have characteristics in their writing which erroneously assign them as 'male'; this is explored further in Chapter 4.

Word Embedding

As one of the more recent methods to cross-over to digital humanities from computer science, word embedding has, as yet, been only occasionally used with nineteenth century novels. Lynn Cherny's 'Pride and Prejudice and Word Embedding Distance' visualisation experiments with the novel by training a word2vec model on the novel and replacing the nouns in the text with the closest word in the model. She admits herself that the method and the visualisation itself is far from ideal: "[n]o, it's not a good read. I stuck to 'top match' in the model, deciding this is about Jane's oeuvre themes, not a good novel mashup" (Cherny 2014). An explanation of the vector arithmetic that makes this possible, although fallible, can be found in Chapter 3.

Siobhan Grayson et al. explore the application of a variation of the word embedding algorithm, Novel2Vec, to a corpus of novels by Austen, Dickens and Conan Doyle (Siobhán Grayson et al. 2016). Using datasets compiled "from texts that have been manually annotated to include definitive character names" they found "that syntactically, character vectors are very distinguishable from other grammatical categories of works within each *novel2vec* dataset" (Siobhán Grayson et al. 2016). They conclude that "[o]ur initial results suggest that word embeddings can potentially act as a useful tool in

supporting quantitative literary analysis, providing new ways of representing and visualising well-known literary texts that complement traditional 'close reading' techniques" (Siobhán Grayson et al. 2016).

Grayson et al.'s 2017 paper 'Exploring the Role of Gender in Nineteenth Century Fiction Through the Lens of Word Embeddings' examines the relationship between an author's gender and their selection of words, using a corpus of forty-eight nineteenth century novels. They found that nouns and pronouns could be split into four categories, noting that "Group C consists of family related nouns (singular and plural) by only female authors, nested within a cluster of characters predominately from Jane Austen's novels" (S. Grayson et al. 2017). They "found that there are differences in the ways in which the male and female authors of this corpus use terms relating to contemporary understandings of gender and gender roles (such as 'she', 'lady', 'gentleman' and occupations/professions" (S. Grayson et al. 2017).

Network Analysis

Extraction of networks from texts can be a complex process: "[t]he transformation of textual data into a corresponding network is much easier if the data are structured using some kind of *markup*" (Batagelj, Mrvar, and Zaveršnik 2002). At the heart of the process is the need to extract information from a text in order to present it in an appropriate format for the chosen tool. A common misconception is that the scholar simply runs the information through a computer program and the results are presented (Fish 2012). In reality, the process is far more complex.

Elson and McKeown analyse the conversations in sixty novels and serials, from thirty-one authors, including Jane Austen. Their purpose was to test "the validity of some core theories about social interaction and their representation in literary genres", specifically the correlations between "the amount of dialogue in a novel and the number of characters in that novel" and that "a significant difference in the nineteenth-century novel's representation of social interaction is geographical" (2010, pp. 1014-1016). To create a conversational network, they identified characters using the Stanford NER tagger and trained a categorizer which aimed to assign a speaker to each instance of speech. In order to test the accuracy of their categorizer, Elson and McKeown compared their computer extracted set with a set annotated by humans. They "were able to determine the speaker of 57% of the testing set with 96% accuracy just on the basis of syntactic categorization". However they felt that the fact that "[t]he other 43% of quotes are left here as 'unknown' speakers" was not important as their purpose was to analyse conversational networks rather than individual quotations (2010, p. 1017). Based on their analysis, they found that having more characters in a novel did not reduce the number of conversations and that "[s]urprisingly, the numbers of characters and speakers found in the urban novel were *not* significantly greater than those found in the rural novel" (2010, p. 1020).

Celikyilmaz et al., also writing in 2010, aimed to go further than Elson and McKeown and "uncover patterns from conversations between actors of a corpus that would have otherwise been extremely difficult to discover and to spark off a new approach to social network constructions between actors based on conversation topics" (2010, p. 2). Their study, which also focused on Austen,

aimed to use an unsupervised method of entity extraction. This is not an easy task as we can see from the 43% of unassigned quotations in Elson and McKeown's study, and Celikyilmaz et al. emphasise that the method "requires large amounts of annotated data" (2010, p. 1). However, using the "actor topic model (ACTM) which automatically constructs a network based on dialogue interactions between the characters in a novel without the need for annotated text", they were able to work with unannotated texts (2010, p. 2). This is an important step for researchers who aim to automate the whole process of information extraction. Overall, this study found that it was possible to extract conversations and their speakers and identify social networks of characters based on "topical similarities" (2010, p. 6). Identifying the social networks of characters allows for comparison between texts exploring the representation of status, or gender.

He, Barbosa and Kondrak's 2013 study builds on the work of both Elson and McKeown, and Celikyilmaz et al. focusing on Austen's *Pride and Prejudice* and *Emma*, as well as Chekhov's *The Steppe*. Their work aimed to address the challenge of automating the extraction of speakers and associated speech: "It is challenging to automate because the speakers of the majority of utterances are not explicitly identified in novels" (2013, p. 1312). Unlike the two previous studies mentioned, He, Barbosa and Kondrak trained a system on "relatively small annotated data sets, and subsequently applied to other novels for which no annotation is available". They also challenged Elson and McKeown's "unrealistic" assumption "that all previous utterances are already correctly assigned to their speakers" (2013, pp. 1312-1313). The study identifies three types of

speaker in the novels: the 'explicit' speaker, who is directly assigned to an utterance in the text; the 'anaphoric' speaker, identified by an anaphoric phrase; and the 'implicit' speaker, for whom there is no speaker information in the paragraph. This poses a considerable difficulty for a researcher intending to assign utterances to a specific speaker as "[t]ypically, the majority of utterances belong to the implicit-speaker category" (2013, p. 1314).

This thesis builds on the studies outlined above. However, it aims to go further; by applying the theoretical framework of enhanced reading, the study will explore how the application of a combined close and 'distant' method can provide new insights and greater depth in the study of nineteenth century novels.

3 Enhanced Reading in Practice

> Do not consider me now as an elegant female intending to plague you, but as a rational creature speaking the truth from her heart
>
> Jane Austen, *Pride and Prejudice*

3.1 Corpus Creation

Douglas Biber, in his article 'Corpus Linguistics and the Study of Literature: Back to the Future?', proposes that "it could be argued that a corpus provides the best way to represent a textual domain, and corpus analysis is the most powerful empirical approach for analyzing the patterns of language use in that domain" (Biber 2011, p. 15). To some extent, this study is both corpus-based ("as a source of examples, to check researcher intuition") and corpus-driven ("the corpus itself is the data") (Baker et al. 2008, p. 16), in that the individual corpora provide the data for the study, while also providing examples to reinforce close readings of the texts. The creation of several corpora was necessary: one for each of the chosen authors, and a fourth comparison corpus of novels from the same period. This comparison corpus is comprised of two sub-corpora, one of female authors, and one of male authors. The scope of the corpora was that all texts must have been published for the first time, between 1800 and 1820, and excluded translations of previously published

works. The novels were included if they were listed in Garside and Schöwerling's (2000) *The English Novel 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles. Vol. II: 1800-1829*, which had a rigorous approach to determining whether the novels had been published by using "current short-title catalogues and other on-line resources" and "undertak[ing] extensive further research in individual libraries and special collections" (p. 4), thus ensuring that the novels had actually been published.[1] In addition, Garside and Schöwerling's definition of novels was used which "includes what contemporaries thought of as novels incorporating works categorized as 'novels' in contemporary periodical reviews and under 'novels' headings in circulating library catalogues" (p. 4).

In her article 'Containing Chaos: compiling a corpus of eighteenth century prose fiction', Iris Gemeinböck discusses some of the challenges in using existing corpora; although some examples exist "there are relatively few free corpora of eighteenth century British texts" (2015, p. 2). The more specialised the required corpus, or the more focused the time period, the less likely it is to be able to find an existing corpus which will suit the needs of the study. Gemeinböck refers to the *Corpus of Late Modern English Texts* (*CLMET*), which, although freely available, was not suited to her study of Gothic novels. *CLMET3.1* is a corpus of 333 texts and consists of approximately 35 million words from 212 authors published between 1710 to 1920. The corpus is created from texts freely available from *Project Gutenberg* and *The Oxford*

[1] 'Ghost' texts, which were announced but never published, were not uncommon during this period, as noted in Chapter 1 above, Jane Austen's early version of *Northanger Abbey*, then known as *Susan*, had been announced as being 'in the press' but publication did not take place (Gilson 1982).

Text Archive, providing a variety of text types: narrative fiction, narrative non-fiction, drama, letters, treatise, and 'other'. Contributions from each individual author are limited to a maximum of 200,000 words.[2] For the period examined for this thesis (1800–1820) there are 14 texts, 6 of which are novels, and only four of which would meet Garside and Schöwerling's novel criteria, as two texts were written specifically for children and were excluded from their survey. Of the four remaining texts, two are novels by Austen. Therefore it was necessary to create the four corpora needed for this study.

Where possible, the texts for each of the corpora were sourced from *Project Gutenberg*. Although there are some minor disadvantages to using texts from this source — the specific editions and dates of texts are not indicated, and some texts are available as part of collections or in multiple volumes — the texts are 'clean' and therefore require relatively little pre-processing.[3] A greater challenge is presented by texts, such as those of Owenson, which are not part of the traditional literary canon, and which may not be familiar to a modern audience. Although there is no restriction in the texts submitted to *Project Gutenberg*, the project relies upon volunteers to create and proofread texts. Currently a total of 54,000 texts have been digitised and proofread by volunteers from the *Distributed Proofreaders* site to check for accuracy, each page passing through two proofreaders prior to publication (Hart 2007). It it not surprising, perhaps, that many of the texts chosen by volunteers for digitisation are those which are familiar to a Western reading audience: those

[2] *CLMET3.1* can be accessed at https://perswww.kuleuven.be/~u0044428/clmet.htm

[3] Michael Hart in the 'Project Gutenberg Mission Statement' emphasises that "we do not provide standards of accuracy above those as recommended by institutions such as the U.S. Library of Congress at the level of 99.5%", noting that "most of our eBooks exceed these standards", although this was not the case in the early stages of the project (Hart 2007).

Text Ref	Date	Title	Gutenberg ID	Words
JA_1811_SS	1811	Sense and Sensibility	161	120,764
JA_1813_PP	1813	Pride and Prejudice	1342	122,874
JA_1814_MP	1814	Mansfield Park	141	161,443
JA_1815_E	1815	Emma	158	161,971
JA_1818_NA	1818	Northanger Abbey	121	78,041
JA_1818_P	1818	Persuasion	105	84,158
			Total:	729,251

TABLE 3.1: Austen Corpus

that are canonical texts rather than being representative of publications from this period.

The Austen Corpus

The Austen corpus, as may be expected given Austen's popularity among corpus linguists and digital humanists, was the least problematic to create. For this study the corpus of Austen's six published novels is from *Project Gutenberg* and contains a little over 729,000 words.[4] The texts are named following the format AUTHOR_DATE_TITLE and are listed in Table 3.1, the *Project Gutenberg* text IDs, and word counts are also given.

The Edgeworth Corpus

The Edgeworth corpus was a little more complex as, although all the novels for the study were available from *Project Gutenberg*, a number of them were only available in multi-text collections. To enable tracking by date, the individual texts were extracted (with the exception of the stories in *Popular Tales* which

[4]It is also possible to access Austen's novels using Julia Silge's `janeaustenr` package for R (Silge 2017).

Text Ref	Date	Title	Gutenberg ID	Words
ME_1800_CR	1800	Castle Rackrent	1424	35,365
ME_1801_B	1801	Belinda	9455	182,365
ME_1804_PT	1804	Popular Tales	8720	194,454
ME_1805_MG	1805	The Modern Griselda	9620	22,256
ME_1806_E	1806	Ennui	9439	81,722
ME_1806_L	1806	Leonora	35638	66,818
ME_1809_A	1809	Almeria	9414	29,010
ME_1809_D	1809	The Dun	9439	10,202
ME_1809_MF	1809	Madame de Fleury	9620	62,234
ME_1809_M	1809	Manoeuvring	9414	22,820
ME_1812_A	1812	The Absentee	1473	104,782
ME_1812_EC	1812	Emilie de Coulanges	9620	31,906
ME_1812_V	1812	Vivian	9414	81,217
ME_1814_P	1814	Patronage	8937 9321	243,410
ME_1817_H	1817	Harrington	9107	83,900
ME_1817_O	1817	Ormond	9107	119,314
			Total:	1,371,775

Table 3.2: Edgeworth Corpus

were available as a single text), and saved given the date that they were originally published, using the same format of AUTHOR_DATE_TITLE used for the Austen corpus.[5] Where possible, single text editions have been used for the Edgeworth corpus. The texts in the Edgeworth corpus are listed in Table 3.2.

The Owenson Corpus

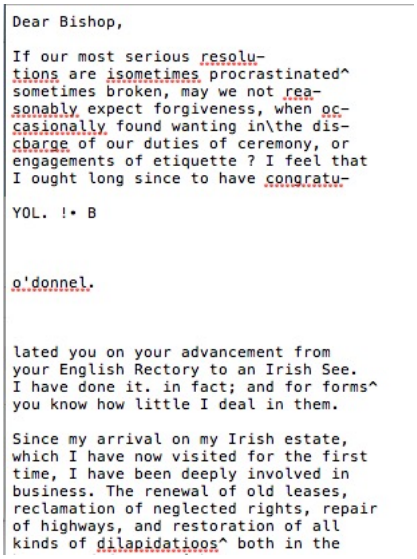
The third author, Sydney Owenson, was rather more problematic as only two of her novels were available from *Project Gutenberg*: *The Wild Irish Girl*,

[5] *Ennui* was originally published in *Tales of Fashionable Life* (*TFL*) volume 1, *Almeria, Madame de Fleury* and *The Dun* in *TFL* volume 2, *Manoeuvring* in *TFL* volume 3, *Vivian* in *TFL* volume 4, *Emilie de Coulanges* in *TFL* volume 5, and *The Absentee* was published in two parts in *TFL* volumes 5 and 6.

and *The Missionary*. As a result, the Owenson corpus, required the greatest amount of additional processing, and contains a greater number of errors than the Austen or Edgeworth corpora. One of the difficulties in studying lesser known authors is that their works are less likely to have been digitised. The *Google Book Project* and other mass digitising projects have made some inroads into texts which may still exist in libraries but not in digital format. However, these projects have human readability as their main aim and therefore the Optical Character Recognition (OCR) scanning may not capture the text in sufficient detail to convert to other formats. The quality of the scan depends ultimately on the quality of the original text. The quality of the paper, depth of colour of the print, font, and foxing (the reddish-brown marks that appear on old paper), all have an impact on the finished scan.

The majority of the novels in the Owenson corpus were sourced from *Archive.org*. The digital copies of these texts are variable due to the quality of the text being scanned and the method of scanning used. To mitigate this, the OCR text was reprocessed by rescanning using Abbyy FineReader Pro for Mac (12.1.6); although this does not lead to a perfect text, the number of errors are reduced significantly as can be seen in Figures 3.1 and 3.2. To further reduce the number of errors a number of manual corrections were made, for example removing page numbers and book titles using the code script seen in Appendix A. It was felt that manually checking the four rescanned novels would be too time consuming and that the remaining level of error was acceptable for the analyses in this thesis.

While it was possible to find reasonable quality scanned copies of Owenson's *Saint Clair*, *The Novice of Saint Dominick*, *O'Donnel*, and *Florence Macarthy*,



Dear Bishop,

If our most serious resolu-
tions are isometimes procrastinated^
sometimes broken, may we not rea-
sonably expect forgiveness, when oc-
casionally found wanting in\the dis-
cbarge of our duties of ceremony, or
engagements of etiquette ? I feel that
I ought long since to have congratu-

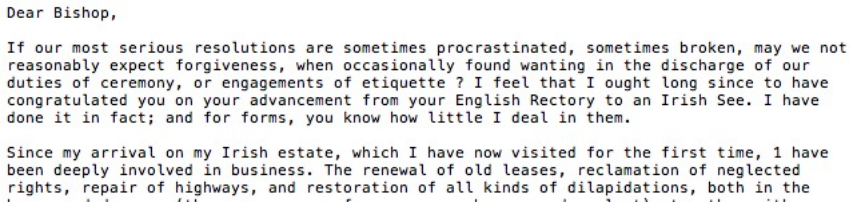
YOL. !• B

o'donnel.

lated you on your advancement from
your English Rectory to an Irish See.
I have done it. in fact; and for forms^
you know how little I deal in them.

Since my arrival on my Irish estate,
which I have now visited for the first
time, I have been deeply involved in
business. The renewal of old leases,
reclamation of neglected rights, repair
of highways, and restoration of all
kinds of dilapidatioos^ both in the

FIGURE 3.1: SO Text as Downloaded



Dear Bishop,

If our most serious resolutions are sometimes procrastinated, sometimes broken, may we not
reasonably expect forgiveness, when occasionally found wanting in the discharge of our
duties of ceremony, or engagements of etiquette ? I feel that I ought long since to have
congratulated you on your advancement from your English Rectory to an Irish See. I have
done it in fact; and for forms, you know how little I deal in them.

Since my arrival on my Irish estate, which I have now visited for the first time, 1 have
been deeply involved in business. The renewal of old leases, reclamation of neglected
rights, repair of highways, and restoration of all kinds of dilapidations, both in the

FIGURE 3.2: SO Text After Reprocessing

the same cannot be said for *Woman; or, Ida of Athens*. As discussed in Chapter 1, this text was viciously criticised when it was published, and was less successful than Owenson's earlier books existing only in a single UK edition (Garside and Schöwerling 2000) and, in her later years, Owenson referred to it herself as a "bad book" (Morgan and Hepworth Dixon 1863a, p. 352), which may in part explain the lack of available copies. The scanned copies of *Woman; or, Ida of Athens* which were available were of variable quality across the four volumes. There was only a single scanned copy of volume 2 which was very poor. Even after rescanning the PDF images using Abbyy FineReader, the resulting plain text file contained far too many errors to be reasonably

Text Ref	Date	Title	Gutenberg ID	Words
SO_1803_SC	1803	St. Clair	-	50,218
SO_1806_NS	1806	The Novice of Saint Dominick	-	230,060
SO_1806_WG	1806	The Wild Irish Girl	54683	103,937
SO_1811_M	1811	The Missionary	48742–48744	83,036
SO_1814_OD	1814	O'Donnel	-	140,519
SO_1818_FM	1818	Florence Macarthy	-	173,532
			Total:	781,302

Table 3.3: Owenson Corpus

corrected using the methods above and therefore the decision was made to exclude this text from the Owenson corpus. The texts in the Owenson corpus are listed in Table 3.3.

The Austen Edgeworth Owenson Corpus

For some of the analyses it was necessary to have a corpus which combined the texts from each of the three authors. The Austen, Edgeworth, Owenson corpus contains 28 texts and approximately 2.9 million words. Figure 3.3 illustrates this combined corpus. It is notable that while Edgeworth and Owenson publish throughout this period, Austen only publishes in the second half, between 1811 and 1818. Edgeworth's texts include the six shortest: *The Dun, The Modern Griselda, Manoeuvring, Almeria, Emilie de Coulanges*, and *Castle Rackrent*, as well as the longest text: *Patronage*.

19th Century Comparison Corpus

In order to identify which aspects of independence found in the novels were typical of novels written during this period, and which were specific to Austen,

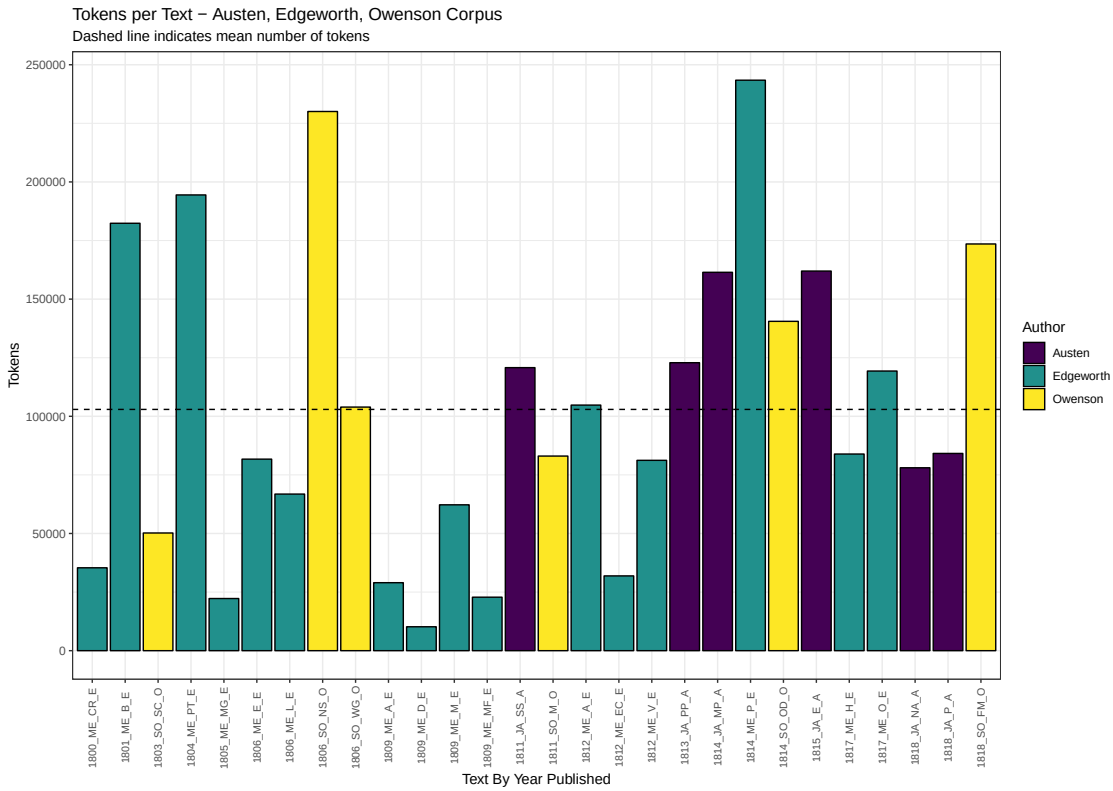


FIGURE 3.3: Tokens Per Text By Year

Edgeworth and Owenson, a comparison corpus was created. As this study focuses on three female authors, two of whom, although popular between 1800 and 1820, have been relatively overlooked until recently, it was important that the comparison corpus was as representative as possible given the availability of source texts. Therefore the decision was made to create a comparison corpus from freely available sources, creating a balance, where possible, between female and male authors, and well known and lesser known texts.

In their survey of novels written between 1770 and 1829, Garside and Schöwerling list the eighteen most prolific novelists from this time (13 female authors and 5 male), responsible between them for 283 novels (2000, p. 64); however being prolific in the early 1800s does not translate to availability today.

While over 1500 novels were published during the period 1800–1820, only a relatively small proportion of these have been digitised, and a smaller number again are available from *Project Gutenberg* or other sites where texts are proofread. This makes sourcing a broad and representative collection of novels for a comparison corpus a challenge.

Algee-Hewitt et al. (2016) highlighted a number of the challenges faced when creating a corpus which is representative of the texts published in a given period including: text availability, human resources, time, usage policies and cost (2016). While physical texts may be available, the cost of scanning texts and the time involved is far beyond the scope of the single scholar. Algee-Hewitt et al. also addressed whether texts in the corpus were representative of the archive as a whole: "Almost certainly not . . . it was perfectly possible that its principle of selection *would make it resemble the canon much more than the archive as a whole*" (Algee-Hewitt et al. 2016, p. 2). David Brewer noted that "[I]n order to be countable . . . texts have to be treated as if they were comparable units" (2011, p. 162). However, he cautioned that "doing so risks distorting the massively different place they occupy in the history of reading" (2011, p. 163). The 'footprint' of a text, as Brewer called it, reflects not only its popularity at the time of publication, but also the significance of the text over a period of time. This echoes the attempts in Algee-Hewitt et al. to plot the popularity of a text against its 'prestige' (2016, p. 4). The difficulty that arises is how 'prestige' or significance are measured (in Algee-Hewitt et al. the measure was "based on the number of mentions a 'primary subject author' in the MLA Bibliography, and on the length of DNB entries" (2016, p. 4)). In the majority of cases these measures will merely reinforce the existing canon,

excluding texts and authors who have been sidelined. To mitigate this in the comparison corpus used for this study, as broad a range as possible of novels and authors was sought, with the guiding principle being that they must be good quality copies — quality, for the purpose of this thesis, being either hand typed or having gone through a process of proofreading such as that employed by *Project Gutenberg*.

The 19th Century comparison corpus is comprised of female and male sub-corpora. Eighteen of the corpus novels were sourced from the Chawton House Library *Novels Online* collection, a collection of hand transcribed novels from early women writers. The focus of *Novels Online* is to make "freely accessible full-text transcripts of some of the rarest works in the Chawton House Library collection" (*Novels Online* 2019), and therefore they represent texts which may not be represented in other online text collections. Only novels where the author was identified as female, either at time of publication or subsequently, were included.[6] Nine additional female novels were sourced from *Project Gutenberg*, two from *ebooks@Adelaide*, and two from *Project Gutenberg Canada*.[7] Table 3.4 lists the thirty-one texts written by female authors, consisting of approximately 3.6 million words. For ease of use, texts in the 19th Century corpus are named using DATE_AUTHOR_TITLE_GENDER.

The 19th Century Male sub-corpus presented a greater challenge, firstly, while

[6] *The Wife* is included in Garside and Schöwerling (2000) Appendix F 'Uncertain reconstituted/unseen titles'. However, the texts in the Chawton House Online collection have been manually transcribed from books physically present in the library, and the validity of this novel is deemed to be confirmed, and it is therefore included in the CF Corpus.

[7] Due to the differing copyright laws in Canada, Australia and the USA there are separate *Project Gutenberg* websites for each country. Although *Project Gutenberg* is by far the largest site, *Project Gutenberg Canada* and *Project Gutenberg Australia* hold some texts from this period which are not available on the US based site.

Text Ref	Date	Title	Words	
1800_Carver_OW_F	1800	The Old Woman	78,989	
1801_Martin_E_F	1801	The Enchantress	45,160	
1801_Opie_FD_F	1801	The Father and Daughter	34,190	
1802_Craik_SN_F	1802	Stella of the North	183,336	
1803_Porter_TW_F	1803	Thaddeus of Warsaw	185,271	
1804_Hunter_UL_F	1804	The Unexpected Legacy	124,544	
1804_Opie_AM_F	1804	Adeline Mowbray	108,057	
1806_Dacre_Z_F	1806	Zofloya	100,833	
1807_Strutt_DR_F	1807	Drelincourt and Rodalvi	123,774	
1808_Wilkinson_CM_F	1808	The Child of Mystery	63,404	
1809_Foster_CE_F	1809	The Corrina of England	58,941	
1809_Mackenzie_G_F	1809	The Irish Guardian	117,681	
1810_Green_RR_F	1810	Romance Readers ...	108,451	
1810_Porter_SC_F	1810	The Scottish Chiefs	287,817	
1811_Brunton_SC_F	1811	Self-Control	186,389	
1812_Foster_SS_F	1812	Substance and Shadow	132,166	
1812_Helme_M_F	1812	Magdalen	94,176	
1812_Jacson_TRN_F	1812	Things By Their ...	96,458	
1812_Stuart_CT_F	1813	Cava of Toledo	210,574	
1813_Cooper_W_F	1813	The Wife	48,168	
1813_Spence_CD_F	1813	The Curate ...	90,172	
1814_Brunton_D_F	1814	Discipline	156,282	
1814_Burney_W_F	1814	The Wanderer	327,099	
1814_Hofland_MW_F	1814	The Merchant's Widow ...	36,949	
1816_Hofland_AB_F	1816	The Affectionate Brothers	25,312	
1816_Lamb_G_F	1816	Glenarvon	152,484	
1817_Selden_V_F	1817	Villasantelle	57,880	
1817_Taylor_R_F	1817	Rachel	28,907	
1818_Ferrier_M_F	1818	Marriage	149,270	
1818_Shelley_F_F	1818	Frankenstein	75,262	
1819_Harvey_AT	1819	Any Thing But ...	124,656	
			Total:	3,612,682

Table 3.4: 19th Century Female Corpus

a greater number of male writers have traditionally been included in the literary canon this in reality only represents a small number of authors from this narrow period (1800–1818). Of Garside and Schöwerling's five prolific male writers, only the works of Sir Walter Scott are obtainable from *Project Gutenberg*; the other prolific male authors from this period, for example Francis Lathom who published 17 novels, are not. The books of some prolific authors, who published with high volume and low reputation presses like the Minerva Press, were often of poor physical quality (J. E. Hill 2006). Initially, it was intended that texts for this corpus would also be sourced from *Archive.org*. However as noted above, as the texts have human readability as a primary aim, the OCR scans are often not suitable for conversion to plain text, Figure 3.4 shows the output of one such text, a page section from Richard Sickelmore's *Osrick; or, Modern Horrors*.

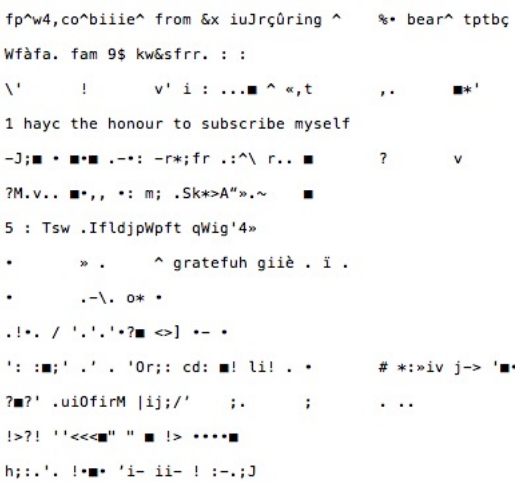


FIGURE 3.4: Screen Shot of OCR to Plain Text

Thirteen texts were sourced from *Project Gutenberg*, one from *Public Library UK*, and two from *Project Gutenberg Australia*. A limit of four books per author was set. Where multiple texts by a single author were available, for

Text Ref	Date	Title	Words
1800_Brown_O_M	1800	Ormond	94,992
1803_Brown_AM_M	1803	Arthur Mervyn	149,880
1803_Brown_EH_M	1803	Edgar Huntley	94,679
1804_Brown_JT_M	1804	Jane Talbot	84,010
1805_Godwin_F_M	1805	Fleetwood	139,580
1810_Shelley_Z_M	1810	Zastrozzi	31,123
1811_Shelley_SI_M	1811	St. Irvyne	31,930
1813_Barrett_H_M	1813	The Heroine	102,391
1814_Scott_W_M	1814	Waverley	205,434
1815_Scott_GM_M	1815	Guy Mannering	171,071
1816_Peacock_HH_M	1816	Headlong Hall	28,542
1816_Scott_TA_M	1816	The Antiquary	173,667
1818_Peacock_NA_M	1818	Nightmare Abbey	27,505
1819_Polidori_V_M	1819	The Vampyre	11,012
1820_Maturin_MW_M	1820	Melmoth the Wanderer	243,635
1820_Scott_A_M	1820	The Abbot	182,697
		Total:	1,772,148

Table 3.5: 19th Century Male Corpus

example Scott who published prolifically during this period, selections were made to ensure as broad a range of dates were covered as possible. As the purpose of this corpus was for comparison, it was decided that, rather than spending considerable time rescanning and cleaning texts from other sources, the use of a smaller number of clean texts would serve the needs of this thesis. The sixteen texts in the 19th Century male sub-corpus are listed in Table 3.5.

The combined 19th Century corpus consists of 47 novels published between 1800 and 1820, with at least one novel per year, and contains approximately 5.4 million words. It includes works from six of the eighteen most productive authors from the period (five female: Barbara Hofland, Sarah Green, Jane Harvey, Amelia Opie, and Anna Maria Porter, and one male author: Sir Walter Scott) (Garside and Schöwerling 2000, p. 64), as well as including a number

of authors and texts which remain popular today. Figure 3.5 illustrates the composition of the 19thC corpus by gender and date. In total, the combined 19thC corpus represents approximately 4% of the 1246 novels published during this period which were 'named', 'identified' or 'implied' as having a female or male author (Garside and Schöwerling 2000, p. 73). Anonymous novels were excluded. Where possible, the novels and authors chosen also aimed to represent geographically novels written during this period: approximately 55% of the novels have an English author, 20% Scottish, 8% Irish or Anglo-Irish, 8% American, and 8% of the authors' origins are unknown. As such, the corpus is deemed to be broadly representative of novels published in the United Kingdom at this time. The full details of the texts in this corpus can be found in Appendix B.

3.2 Computational Analysis

3.2.1 R Programming

There are a number of out-of-the-box tools available for textual analysis, many of them freely available. The *DiRT Directory* and *TAPoR 3* provide information and access to a large number of tools. *TAPoR 3*, "a gateway to the tools used in sophisticated text analysis and retrieval" (Rockwell and Sinclair 2014), for example, lists 499 separate tools for different aspects of textual analysis. Several of these tools are reliant on a strong and stable internet connection, as they work within a browser window, without which results can be lost. Additionally, tools may not be updated regularly or become archived. This

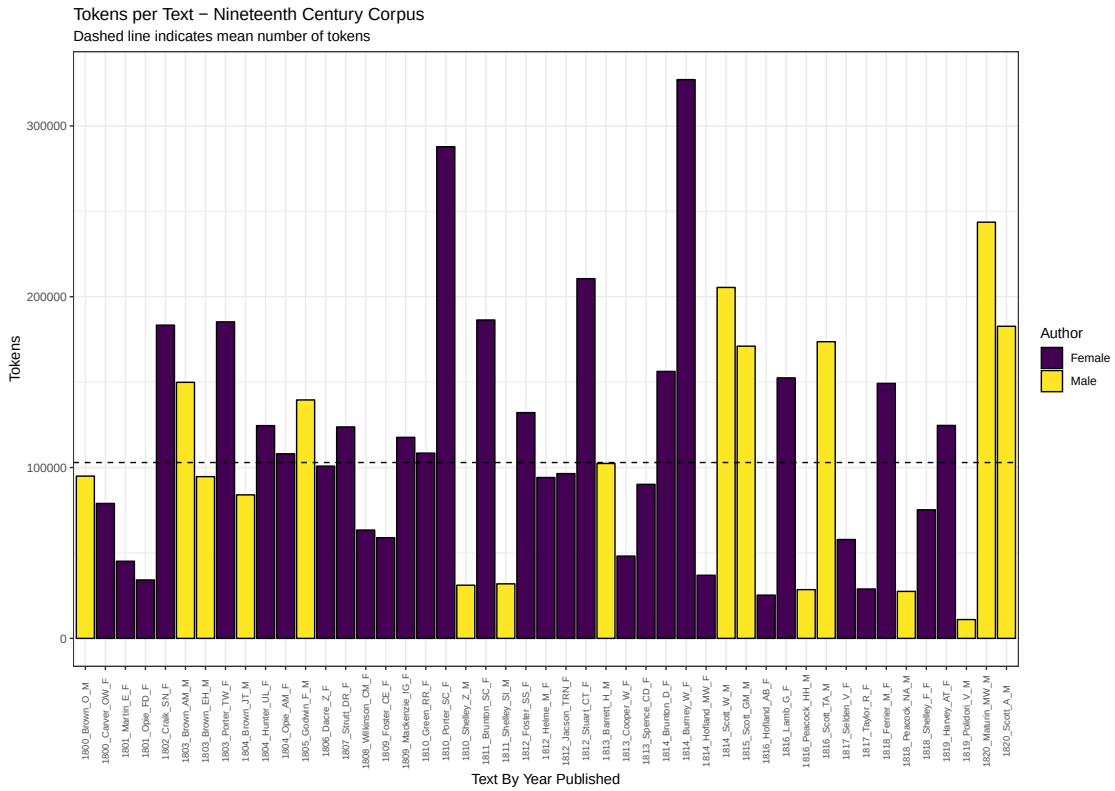


FIGURE 3.5: Plot of 19thC Corpus by Gender and Date

can lead to bugs or a lack of replicability. Although it was possible to use a different tool for each of the analyses employed in this thesis, this would mean switching between a wide number of different interfaces, and potentially having to use another tool for visualisation. Furthermore, the algorithms and parameters used for a particular analysis may not be clear at the surface level of the tool. Where a tool is open access and open source it may be possible to drill down into the code to identify the decisions being made (for example *Voyant Tools* (Sinclair and Rockwell 2019) provides all its code on *GitHub*). However, this assumes detailed knowledge of the algorithm and the programming language used to create the analysis, which may vary from tool to tool. While out-of-the-box tools are useful and effective for an initial analysis, their

lack of customisation rules them out for use in a larger study.

Proprietary software, for example MATLAB or SPSS, was not considered for this study for a number of reasons outlined succinctly by Seth Brown: "proprietary software takes power away from the community and gives it to a single monopolistic entity. When what is best for the community no longer aligns with what is best or most profitable for the entity, problems ensue ... More troubling still, using a closed proprietary language gives the impression that users do not care if their results are reproducible or verifiable" (Brown 2014). Spreadsheets (which are often also proprietary,), for example Excel, were discounted as they are not ideal for advanced problems or large datasets. As some of the guiding principles behind this study were that the data, tools and results should be open access and, as far as possible, replicable - proprietary tools (which charge for use) were excluded.[8]

The decision to use a single programming language, therefore, was taken for simplicity and ease of replicability. For analytics and data science the industry standard tools are R or Python, as shown in the 18th annual KDnuggets Software Poll (Piatetsky 2017). Both languages offer a broad range of packages to carry out data analysis, the main difference between the programs being the approach. R is "an integrated suite of software facilities for data manipulation, calculation and graphical display" (Venables, D. M. Smith, and R Core Team 2017, p. 2), whereas Python is a full-service programming language. In a recent article comparing the two languages, Wayner proposed that "[t]he

[8]While all data came from open access sources, and all analyses and visualisations took place using R, it was necessary to use Abbyy FineReader to reprocess four of the texts. To mitigate this potential difficulty in replicating the creation of the study corpora, in keeping with the guiding principles of the study, the reprocessed corpus texts used are available on request.

difference between Python and R is largely philosophical" (Wayner 2017). R was chosen due to its excellent statistical and analytics abilities, its ability to produce publication quality graphics, and community support, both online and through guides written specifically for an academic audience (Gries 2009; Gries 2013; Jockers 2014). All programming for this study used R and the free IDE (integrated development environment) RStudio.

3.2.2 Vector Space Models

This thesis proposes a shift in focus from large scale "distant" (Moretti 2013) or "Macro" (Jockers 2013a) readings, and suggests instead viewing the use of computational methods as *enhanced* reading, which has the purpose of further improving the quality and extent of close readings of the text, rather than, potentially, replacing them. While the majority of previous studies, discussed in Chapter 2, have focused on a single method of textual analysis, this study uses a combination of methods which come under the umbrella term of vector space models.

Vector space models have their origins in frequency-based information retrieval systems developed for computers, and are a computational model of meaning which rely upon the metaphor of geographic location, using the concept of words existing within a theoretical space where they are represented as a vector of numbers (Sahlgren 2006). These underlying structures of texts, created to enable computers to extract information, have increasingly been used by literary scholars to explore meaning circulation within and between texts.

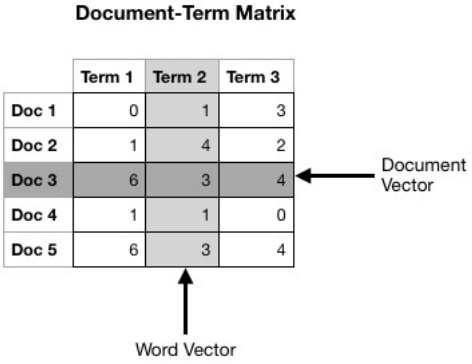


FIGURE 3.6: A Simple Document-Term Maxtrix

Vector space models are matrix-type structures used by computers to make sense of texts and to extract information. At their most straightforward, as seen in Figure 3.6, vector space models can be a raw count of word frequencies within individual documents, and are represented as a series of numbers, a vector. Similarities and differences between texts can be explored by calculating the theoretical distance between the vectors using a variety of mathematical formulae. The vector space model represents documents, or smaller text elements, as points in space which reveal their semantic and/or syntactic relationships. Turney and Pantel's (2010) study identified three main classes of vector space model: the term-document structure, the word-context structure, and the pair-pattern structure. The three classes are unified by their use of "event frequencies" (p. 143), and the underpinning statistical semantics hypothesis: "statistical patterns of human word usage can be used to figure out what people mean" (p. 146).

The term-document structure (Salton, Wong, and Yang 1975) was originally used for automatic indexing. It relates to the bag-of-words hypothesis which

states that "the frequencies of words in a document tend to indicate the relevance of a document to a query" (Turney and Pantel 2010, p. 153). The aim of the term-document structure is to reveal the similarity between documents, for example in the use of Document-Term Matrices. As seen above in Figure 3.6, in a simple Document-Term Matrix, each row represents a document in the corpus and each column represents the frequency of a particular term in that document. The frequency can represent a raw count of the terms, or can be weighted, for example using Term Frequency-Inverse Document Frequency (tf-idf):

$$\frac{t}{d} \times log\Big(\frac{N}{n_t}\Big)$$

where t is the term, d is the document, N is the total number of documents in the corpus and n_t is the number of documents that term t appears in. This reduces the influence of words which appear frequently in the corpus.

The word-context structure, which includes word embeddings, uses the distributional hypothesis that "words in similar contexts tend to have similar meanings" (Turney and Pantel 2010, p. 143). The context is a sliding window of words with a focus term at the centre: "examples of window sizes range from 100 words to just a couple of words" (Sahlgren 2006, p. 68). In this type of model "similar row vectors in the word-context matrix indicate similar word meanings" (Turney and Pantel 2010, p. 148), beyond word similarity, this type of vector space model can be used for a variety of purposes for example: 'word classification', 'automatic thesaurus generation', and 'information extraction' (Turney and Pantel 2010).

The pair-pattern matrix is based on the extended distributional hypothesis

which states that "patterns that co-occur with similar pairs tend to have similar meanings" and the latent relational hypothesis "that pairs of words that co-occur in similar patterns tend to have similar semantic relations" (2010, pp. 148-149). Although the pair-pattern structure has interesting possibilities for the study of literary texts (Lin and Pantel 2001), the technology involved in their creation and analysis have not yet developed fully and, therefore, it is beyond the scope of this study.[9]

Four vector space approaches — frequency analysis, topic modelling, word embedding, and semantic networks — were used in this study. The specific analyses for each of these approaches were chosen to highlight aspects of independence in the novels rather than to represent an exhaustive list of possible analyses or tools. Ute Römer's chapter 'Corpus Analysis in English Studies' (Römer 2006), and Douglas Biber's article 'Corpus Linguistics and the Study of Literature' (Biber 2011), for example, give an indication of the range of analyses which fall under the scope of just frequency analysis. In the sections below a general outline of the methods are included, comprehensive details of each analysis and its output are included in Chapters 4 and 5.

[9]Turney and Pantel also briefly mention other semantic vector space models which go beyond the three structures mentioned above: "higher-order tensor[s]" (Turney and Pantel 2010, p. 151). Whereas a matrix can represent a document and a term, a tensor can simultaneously represent multiple dimensions for example document, term, year and author. These have interesting applications to literary studies in the future as can be seen in machine learning and neural network developments such as the open source *TensorFlow* developed by the Google Brain Team (Google 2017) .

Frequency Analysis

The most basic method of textual computational analysis is frequency analysis, although simple "frequency lists are useful for computing many co-occurrence statistics" (Gries 2009, p. 14). For this study, the terms in the corpora were calculated as raw, relative, and tf-idf weighted frequencies. This enabled a number of analyses to be undertaken to evaluate text and author similarity, lexical co-occurrence, and author word usage, including calculating which words were significantly overrepresented in an author corpus when compared with the 19th Century corpus. The measure of lexical co-occurrence was restricted to collocations (the co-occurrence of terms with a frequency greater than chance) as these best reveal the similarities and differences between the authors' use of words. Collocations can also reveal semantic prosody, "the typical contexts in which particular words occur, as a way of establishing the generalised associations we have for a word, and whether its semantic prosody in a particular text parallels, or contrasts with, its more typical semantic prosody" (Leech and Short 2007, p. 285). This is particularly useful in discerning connoted meaning in ironic text (Römer 2006; O'Halloran 2007; Fischer-Starcke 2009).

Concordances (also known as Key Word in Context) are one of the oldest forms of corpus analysis, having their origins at least as early as the thirteenth century (McClintock and Strong 1872). For this thesis, a Key Word in Context (`kwic()`) function was used to examine the specific context of words within the texts. This was also used to provide context for the results of the other analyses outlined below.

Topic Modelling

Topic modelling is a machine learning model for extracting topics or themes from a collection of texts. While there are topic models which are not probabilistic, for example Non-negative Matrix Factorisation, the most frequently used models in Digital Humanities, and the one used for this thesis, are probabilistic. The topic modelling algorithm calculates the probability that word w in document d is associated with topic t over a series of iterations until a point is reached where the associations are reasonably stable. This is the purpose of the burn-in period (which discards the initial, less accurate results) as well as the multiple iterations. "Gibbs sampling makes it possible to obtain samples from probability distributions ... computing expected values, by defining a conceptually straightforward approximation. This approximation is based on the idea of a probabilistic walk through a state space whose dimensions correspond to the variables or parameters in your model" (Resnik and Hardisty 2010, p. 7). This means that each time the algorithm runs the resulting topics will be slightly different. Therefore topic modelling is classed as non-deterministic and the results from the topic models in this study can be approximated rather than replicated. Blei, Ng and Jordan's Latent Dirichlet Allocation (LDA) algorithm (2003), uses Gibbs sampling, a sampling algorithm "based on Monte Carlo Markov Chain (MCMC) techniques" (Yildirim 2012, p. 1), to extract the topics and their probabilities.

The initial plan for this thesis was to use the `MALLET` package for R (Mimno 2013), which uses R as a wrapper for the Java based MALLET (Machine Learning for Language Toolkit) tool (McCallum 2002). However, MALLET's

reliance on the `rJava` (Urbanek 2018) package proved to be too unstable; updates to the Mac operating system, new versions of Java, and the package itself, caused errors which could not be solved. This led to the decision to use the `topicmodels` R package instead which "[p]rovides an interface with the C code for Latent Dirichlet Allocation" (Grün and Hornik 2011). For each trial the parameters for Gibbs sampling were as follows: initial burn-in period 500 iterations, then 2000 iterations with every 500th iteration being kept over 5 independent runs.

Topic modelling using the `topicmodels` package requires the text to be pre-processed, for example removing punctuation, capital letters, and stop words, and converted into a Document-Term Matrix (DTM). This pre-processing was carried out using the `tm` text mining package. A difficulty when creating topic models is determining the optimum number of topics. The R package `ldatuning` (Murzintcev 2016), which maps the corpus against four separate metrics, was used to identify an appropriate number of topics.

Three tests were carried out on the corpora with the aim of identifying the most effective process for the identification of discourses surrounding independence. The first used the full text of each novel and the standard stop words list. The second used only the singular and plural nouns with standard stop words removed. The third used the singular and plural nouns, broken into chunks of 1000, with a standard stop word list; this was the method chosen for this thesis (Jockers 2013b). The standard stop word list used is in Appendix C. The full process and results are presented in Chapter 4.

Word Embedding

As noted above in Section 3.2.2, word embeddings are based on the linguistic theory, originally proposed by J. R. Firth, that: "you shall know a word by the company it keeps" (Firth 1968). This concept underpins much of corpus linguistics and the more advanced computer algorithms that have developed in the field of computational analysis. In corpus linguistics, collocations rely on reflects the notion that there is something significant about pairs of words that appear together with greater frequency than the average. Word embeddings build on this by considering words which appear in similar contexts. This means that while collocations reveal word pairings which exist within a corpus of texts, word embeddings can provide suggestions of which words *may* appear in a particular context using the model to identify likely candidates. Therefore it is possible to complete analogy tasks, like that mentioned below, even if the target pairing does not occur in the corpus. The results are sensitive to the size and quality of the corpus, and it is important to note that the results may present matches that would be rejected by a human reader. Lynn Cherny's experiment with *Pride and Prejudice*, where she replaces nouns in the novel with the nearest word in a model trained on Austen's novels, presents the opening line as: "It is a case universally acknowledged, that a single woman in defiance of a good sense, must be in use of a son" (2014). An effective but imperfect process.

Like topic models, word embeddings are created using non-deterministic algorithms. Word embedding models are created with a machine learning algorithm, based on an analogy of the human brain. The algorithm takes in a

corpus of texts and represents words as points in a multi-dimensional space. Word meanings and relationships between words are encoded as distances and paths in that space, through the creation of an artificial neural network which aims to reveal the underlying structure of a text or corpus. The relationships between words are encoded as a vector of length n characteristics, n being the chosen number of dimensions (or relationships) between the words in the space. A 'hidden' layer in the neural network provides the output for the word embedding - the vectors - which measure the association between word-context pairs.

This type of analysis requires less pre-processing than topic modelling. Rather than requiring a Document-Term Matrix, it only requires the corpus to be converted into a single plain text file with punctuation removed. "For digital humanists, [word embeddings] merit attention because they allow a much richer exploration of the vocabularies or discursive spaces implied by massive collections of texts than most other reductions out there" (Schmidt 2015).

The word2vec algorithm, originally created by Tomas Mikolov and his colleagues at Google in 2013 (Mikolov, Corrado, et al. 2013), is one of the most commonly used algorithms for creating word embeddings, also known as word vectors. There are two models of the algorithm: continuous bag-of-words (CBOW), and skip-gram. The CBOW model creates a vocabulary from a corpus. It then uses the context of text strings to predict a target word, calculating the probability that each word in the vocabulary is the target. For example, in Figure 3.7 the model uses the context words either side of the focus word (sing, a, of, sixpence) to predict the focus word, which, in this example, we might expect the word 'song' to be amongst the suggestions.

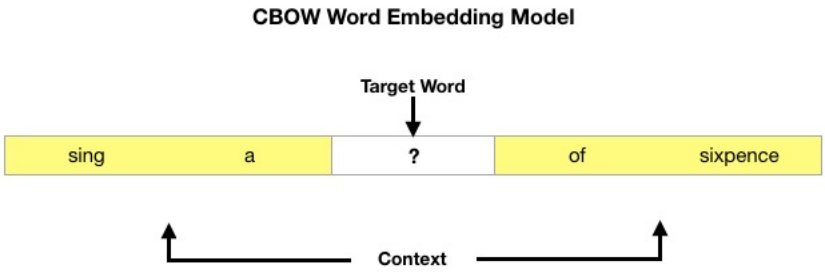


FIGURE 3.7: The CBOW Model

The skip-gram model reverses this, using the focus word to predict the probability for each context word that surrounds it, as can be seen in Figure 3.8. The model can encode multiple semantic and syntactic meanings for the context of each focus word, unlike the CBOW model which averages the possible contexts.

Word embeddings encode both semantic and syntactic relationships, which can be confirmed through vector addition (Mikolov, Yih, and Zweig 2013; Gittens, Achlioptas, and Mahoney 2017). "This unexpected property allows one to use these vectors to answer word-analogy questions *algebraically* . . . e.g., answering the question 'man is to king as woman is to _' by returning the word whose word-vector is nearest to the vector *v(king) - v(man) + v(woman)*" (Gittens, Achlioptas, and Mahoney 2017, p. 70). The *CLMET* corpus provides an example of this using the `wordVectors` package:

```
clmet %>% closest_to(~ 'king' - 'man' + 'woman')
```

which gives the results seen in Table 3.6.

As word embedding models are created through unsupervised learning, they

1	king	0.8053716
2	queen	0.5732614
3	majesty	0.4935590
4	husband	0.4922281
5	woman	0.4828162

TABLE 3.6: Word similarity to 'king' - 'man' + 'woman'

reveal the structure that is latent in the data. Ted Underwood cautions that "it is important to stress that unsupervised models are still shaped by explicit interpretative assumptions" (2018, p. 258). Further, as products of their corpora, word embedding models contain the biases inherent in the original texts, producing analogies such as *doctor - man + woman = nurse* (Bolukbasi et al. 2016; Brunet et al. 2018; Swinger et al. 2018). So, when the model is presented with an analogy its result provides syntactic and semantic information regarding those who created the texts within the corpus, which may reveal the broad biases of a particular period, or the more specific biases revealed by an individual author, all of which are encoded within the model.

This use of vector mathematics also means that the models created can differentiate between polysemic words, for example between river 'bank' and 'bank' the financial institution. As the output of the model is a vector, vector addition or subtraction can be used to combine or reject specific meanings.

An alternative to the word2vec algorithm is Global Vectors for Word Representation (GloVe) created by the Stanford NLP Group (Pennington, Socher, and Christopher D Manning 2014). The main differences between the algorithms is that word2vec is a prediction-based model whereas GloVe is count-based. In their 2014 article, Baroni, Dinu and Kruszewski carried out a "direct comparison between count vs. predictive DSMs [distributional semantic models,

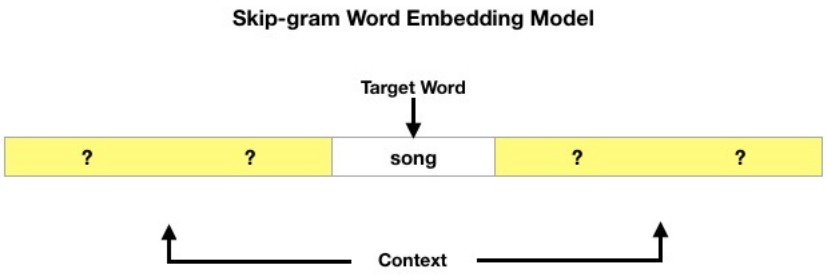


FIGURE 3.8: The Skip-Gram Model

or word embeddings]" (2014, p. 239) concluding that "the predict models are so good that, while the triumphalist overtones still sound excessive, there are very good reasons to switch to the new architecture" (2014, p. 245). In contrast, Pennington, Socher and Manning, the creators of GloVe, argued that "the two classes of methods are not dramatically different at a fundamental level since they both probe the underlying co-occurrence statistics of the corpus" (Pennington, Socher, and Christopher D Manning 2014, p. 1542), a view shared by other recent studies (F. Hill et al. 2014; Schnabel et al. 2015). In light of this, the decision as to which algorithm to use was determined by whether there was an R package available. At the time this study was started, the only R package for word embedding available was `wordVectors` (Schmidt and Li 2015); the package has the benefit of very detailed documentation which is written with non-specialists and Digital Humanists in mind.[10]

The analysis and visualisation was carried out using Schmidt and Li's `wordVectors` package (Schmidt and Li 2015). 300 dimensions were modelled with a window of fifteen words. A function `w2v_analysis()` was written which searched the

[10]In August 2017 an R implementation of GloVe, `text2vec` (Selivanov and Wang 2018) was launched.

word embedding model for a chosen term, reduced the dimensions using `Rtsne` (Krijthe 2015), the Barnes-Hut implementation of t-distributed Stochastic Neighbour Embedding (t-SNE) (Maaten and Hinton 2008; Van Der Maaten 2014), and output a CSV file, word list and plot of the words nearest to the chosen term or terms. The code for the `w2v_analysis()` function can be found in Appendix D.

The finished plots were created using the `ggplot2` package (Wickham 2009) with the coordinates of each word identified by a red marker allowing `ggrepel` (Slowikowski 2016) to adjust the labels to avoid overlap and improve readability. The axes of the graph provide a structure within which the multiple dimensions of the vector space model can be translated. The position of the words in relation to the axes is not significant. Unless a seed is used to ensure replicability, the position of the words will change each time the plot is generated. What *is* significant is the relationship of the words to each other.

Semantic Network Analysis

Once a vector space model has been created it can be explored in greater detail by calculating the similarity between words and creating a semantic network from the results. A semantic network attempts to visualise additional dimensions from a word embedding model, with each word represented by a node and the relationships between the words represented by edges. Whereas dimension reduction such as t-SNE can create a visualisation of the model in two dimensions presenting words clustered by proximity, a semantic network allows these words to be connected to other clusters of terms, allowing

additional semantic and syntactic information to be revealed.

Ryan Heuser (Heuser 2016b) explored the application of word embeddings to abstract nouns and outlined some of the difficulties with dimension reduction. He identified some terms which were physically distant in the t-SNE visualisation, which were actually closer to each other in the vector space of the word embedding model. The benefit of using semantic networks, he suggested, is that "networks usefully appoach distance from a 'monadic' perspective, measuring distance relative to each node or word — which liberates them from t-SNE's impossible task of projecting all the distances between the words onto a single two-dimensional plane" (Heuser 2016b, section 2A). This means that words which are connected to a particular term can "at the same time ... appear in different regions of the network encod[ing] the fact that these words are more typically connected to words in a different region of the semantic space" (Heuser 2016b, section 2A).

The semantic networks created for this thesis use cosine similarity, which uses the geometric definition of vectors as directional lines in space, to calculate similarity between words. Cosine similarity generates a metric which indicates how related two vectors, x and y, are by measuring the angle between them:

$$cos(\theta) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

the resulting value will be between -1 and 1, 1 being totally similar and -1 being totally dissimilar (see Figure 3.9) (Perone 2013). The results were used

to create a similarity matrix where each term was compared to every other term.

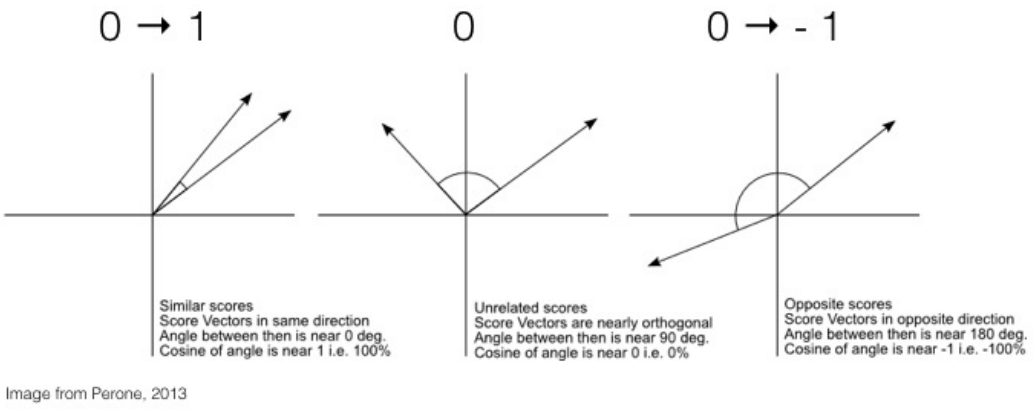


FIGURE 3.9: Cosine Similarity

Creating a network enables not only relationships between words to be viewed, but also provides a method of finding 'meaningful groups' through 'community detection' (Heuser 2016b). Network communities, or groups of terms, can be identified using a clustering algorithm which automatically calculates which terms are related. How this is calculated varies according to the algorithm used and therefore each algorithm will produce slightly different results. The network also allows the relative importance of a term to be calculated using betweenness centrality (the number of shortest paths from each individual node to all other nodes which pass through it) and degree (the number of links a node has to other nodes). The networks in the study were created using the `visNetwork` R package (Almende, Thieurmel, and Robert 2018) and represent the relationship between words nearest to a chosen vector and a corpus. The similarity matrix was filtered to remove terms which were dissimilar by removing those with a score below a threshold of 0.55 and nodes were coloured by community using the `cluster_fast_greedy` algorithm (Clauset, Newman,

and Moore 2004). As outlined above in Figure 3.9, a cosine similarity score of 0 reflects no relationship between terms, and so a threshold of 0.55 was chosen to include words which had a reasonably strong relationship to one another. Setting the threshold too high results in a very sparse network with numerous unconnected nodes, while setting the threshold too low results in an overly dense and unreadable network.

3.3 Conclusion

The application of vector space analyses allow the exploration of the corpora at multiple levels: word, theme, context, and semantic. The first two case studies focus on the term-document vector space structure (Chapter 4), while the second two focus on the word-context structure (Chapter 5). The term-document methods are increasingly used within Digital Humanities, as seen in Chapter 2, and have become accepted tools for the exploration of text. The word-context methods are much more experimental and are at a very early stage in their use in Literary Studies.

4 Term-Document Models

> Your proposal was made to my father — Why was it not made to me? — Men — all men but one — treat women as puppets and then wonder that they are not rational creatures!
>
> Maria Edgeworth, *Vivian*

4.1 Frequency Analysis

Computational frequency analysis has been part of literary studies to some extent since at least the 1960s, largely as a part of Corpus Linguistics (Mosteller and Wallace 1964; Burrows 1986). Frequency analysis allows the empirical exploration of the words that make up a text or a corpus, and as such it requires relatively little interpretation to understand the output, although interpreting its significance may be rather more difficult. In this thesis, frequency analysis is largely used to gain an overview of the make up of the texts and to highlight any initial differences between them. This can then be used to inform the interpretation of later analyses and may account for particular results.

The calculation of frequency, and other linked analyses are some of the most straightforward computational analyses; however there are a number of pitfalls which need to be considered. Different computer packages may identify words in slightly different ways (for example using spacing between words)

and therefore there may be differences in the total number of words and their frequencies. In addition, the method of ingestion for a corpus may introduce some errors in word boundaries, and so it is necessary to consider whether this will have a significant impact on the study.

Word frequencies within a corpus of English texts will usually conform to Zipf's law. Zipf's law proposes that language follows a power law that, given a large sample of words, the frequency of any word is inversely proportional to its rank in the frequency table, so the second most frequent word will appear half as frequently as the most frequent word, the third most frequent word will appear one third as frequently as the most frequent word, and so on (Ferrer-i-Cancho and Elvevag 2010). Therefore, unless the frequencies are normalised in some way, the distribution would not be considered 'normal' in the Gaussian sense (a bell curve), which has implications when selecting methods of computation; for example when calculating distance between text vectors for classification.

For the majority of the analyses in this section, the input is a corpus of plain text files. As any errors introduced, whether through the conversion of files to text format or during the ingestion of texts into the digital corpus, were relatively infrequent, further cleaning to remove these was not deemed necessary.

4.1.1 Grouping Texts

Text classification is a commonly used tool for author attribution (Burrows 2002; Burrows 2003; Hoover 2004a; Hoover 2010; P. W. H. Smith and Aldridge

2011). This type of analysis utilises word frequencies to create an author signature which in turn is used to group known texts. John Burrows' early exploration of Jane Austen's novels in the 1970s highlighted the significance of the words used by an author as a mark of their style. Initially, he hand-counted words, noting that: "The many words I had singled out fell into such revealing patterns that it became clear that Jane Austen wrote so exactly that she was a perfect subject for this approach. And the task became so onerous that hand-counting would not do" (Burrows 2003, p. 6).

While none of the texts in the AEO corpus is of doubtful authorship, classification can be used to reveal similarities between the texts of different authors, and also highlight texts which may not be typical of an author. The continuing popularity of Austen's novels, where those of her contemporaries have fallen out of favour, raises the question: is there something different or unusual about her writing style in comparison with other writers of the period? The first step in answering this question was to run a classification task to explore how Austen's novels grouped with those of Edgeworth, Owenson and those from the 19th Century comparison corpus based on their 'style signal'.

Using the `stylo` package in R, the top 100-500 most frequent words were used to create a Bootstrap Consensus Tree. Classic Delta (Burrows 2002) was used as the distance metric and pronouns were culled. Delta (Δ) is a measure of difference between texts based upon the most common words in a corpus, used in computational stylistics. A "'delta-score' can be defined as 'the mean of the absolute differences between the z-scores for a set of word-variables in an authorial text-group and the z-scores for the same set of word-variables in a target text"' (Burrows 2003, p. 13). Bootstrapping resamples the data and

tests whether the same branches are recreated in multiple iterations of the tree, the final tree reflects the groupings with the greatest probability of occurring. The result of the bootstrap consensus tree can be seen in Figure 4.1.

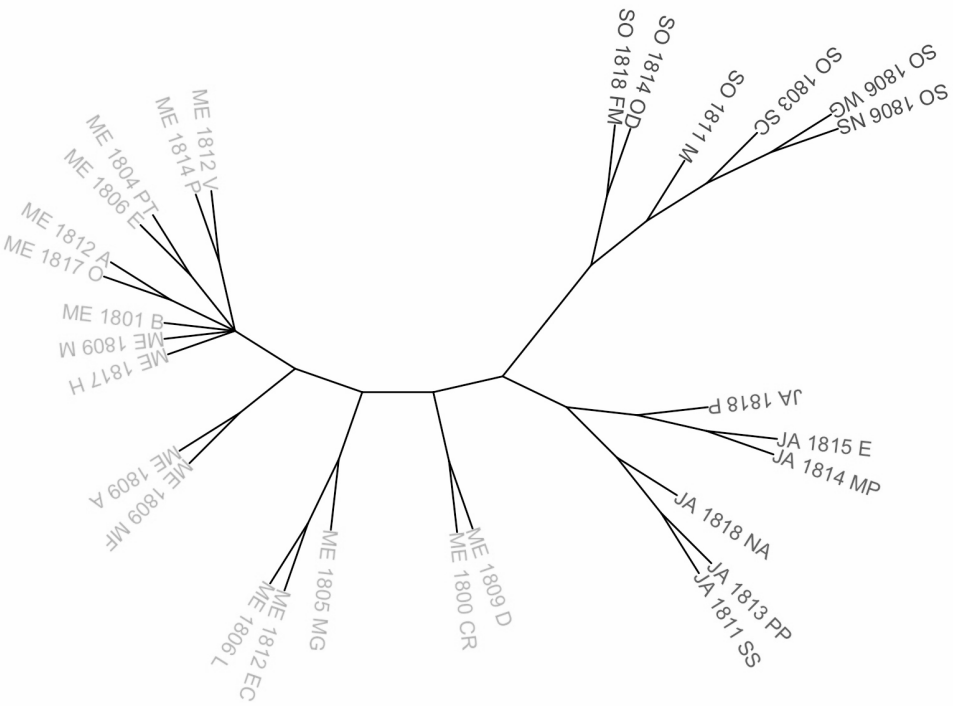


FIGURE 4.1: Bootstrap Consensus Tree - Austen, Edgeworth and Owenson

Figure 4.1 reveals that Austen's texts fall into two main groups which echo the period in which they were written. The early group consists of *Sense and Sensibility*, *Pride and Prejudice*, and *Northanger Abbey*, and the later group is *Mansfield Park*, *Emma*, and *Persuasion* (the Chawton novels). It is notable that the two texts published after Austen's death (*Northanger Abbey* and *Persuasion*) are sub-branches of their group - this may support conjecture that Austen had not completed the preparation of these texts for publication

before her death. Similarly, Owenson's novels are divided into two broad chronological groups (the early group:*St. Clair*, *The Novice of Saint Dominick*, *The Wild Irish Girl*, and *The Missionary*; and the later group: *O'Donnel*, and *Florence Macarthy*). While the Oriental setting and theme for *The Missionary* may be the cause for its separation from the other novels, *The Wild Irish Girl*'s Irish setting, and the designation of 'National Tale' has not caused it to group with Owenson's later National Tales, nor do any of Owenson's National Tales group with those of Edgeworth. This suggests that the signal of the individual authors' style is stronger than their shared themes or genres.

A visualisation of the bootstrap consensus tree for the combined Austen, Edgeworth, Owenson and 19th Century corpora, can be seen in Figure 4.2.

In this visualisation, the works of both Austen and the majority of Edgeworth's are clearly separated from those of other authors. Austen's novels appear in a single group consisting of the two chronological clusters mentioned above, and Edgeworth's are divided in two main groups. Three of Edgeworth's shorter texts (*The Modern Griselda*, *Leonora*, and *Emilie de Coulanges*) are grouped with texts by Taylor, Cooper, Carver, Jacson and Brown, possibly owing to the didactic nature of the texts. In contrast, two of Owenson's texts, *O'Donnel* and *Florence Macarthy*, share a branch with four of Sir Walter Scott's novels, indicating that, while Scott acknowledged the influence of Edgeworth in his introduction to *Waverley*, his style may have been closer to that of Owenson.[1]

[1]Interestingly, this classification also reinforces the argument that Mary Shelley was the sole author of *Frankenstein*, although this was doubted by her contemporary reviewers, and, controversially in the 2007 book *The Man Who Wrote Frankenstein* (Lauritsen 2007). Mary Shelley's *Frankenstein* clusters with her father's 1805 novel *Fleetwood*, and Maturin's 1820 *Melmoth The Wanderer*, rather than those of her husband P. B. Shelley.

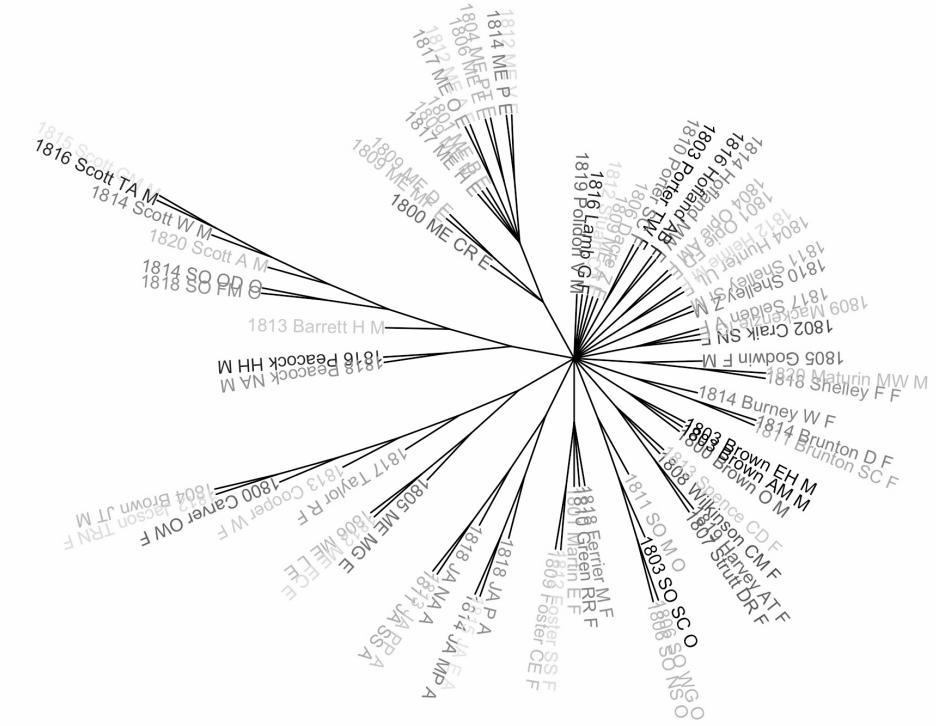


FIGURE 4.2: Bootstrap Consensus Tree - 19th Century Corpus Plus Austen, Edgeworth and Owenson

4.1.2 Type-Token Ratio

As shown in Figure 4.1 and Figure 4.2 above, Austen's 'style profile', based on her most frequently used terms, suggests that her style remained very similar across her novels, and that her style was notably different to that of other authors writing during the same period. To further investigate we need to ask: *how* does Austen's style differ from that of her contemporaries? One method of examining an author's style is to consider the lexical variety used; in contrast to the analysis above this considers all words used rather than just the most frequent. Type-token ratio (TTR) calculates the number of unique words divided by the total number of words in a text, and is a measure of lexical

variety. Comparing the TTR for the Austen, Edgeworth, Owenson corpus combined with the 19th Century comparison corpus, as seen in Figure 4.3, reveals that Austen has a considerably lower TTR than her contemporaries, falling well below the mean of the full corpus, and below the mean of the female authors in the 19th Century corpus. Table 4.1 indicates the mean TTR for each corpus, and shows that Edgeworth's and Owenson's TTR falls closer to that of the male 19th Century authors than either the female 19th Century authors or Austen. This is similar to Matthew Jockers' findings regarding markers of author style: the "gender signal" which identified "features most influential in distinguishing between the male and female authors" for his corpus, words such as "the (M)', 'lady (F)', 'house (M)' and 'thought (F)"' (Jockers 2013a, pp. 94–95). Although the use of TTR is a less sophisticated analysis than that used by Jockers, in both analyses Edgeworth and Owenson were "misclassified" (p. 95). It could be argued that this is because they write about a broader range of ideas than some of the other female novelists of their time, or that their topics were more political. It may also be because their focus was frequently Ireland and the Irish, and because of this their works are differentiated from those of English female writers of the same period.

A notable difference between the male and female subsets of the 19th Century corpus is that the mean for the female TTR is lower than the male. This echoes previous findings (Koppel, Argamon, and Shimoni 2002; Argamon et al. 2003; Jockers 2013a) regarding the influence of gender on an author's style. As a generalisation, it may be argued that an indicator of male style versus female style in the novels in this corpus is that male writers during this period used a higher level of lexical variety.

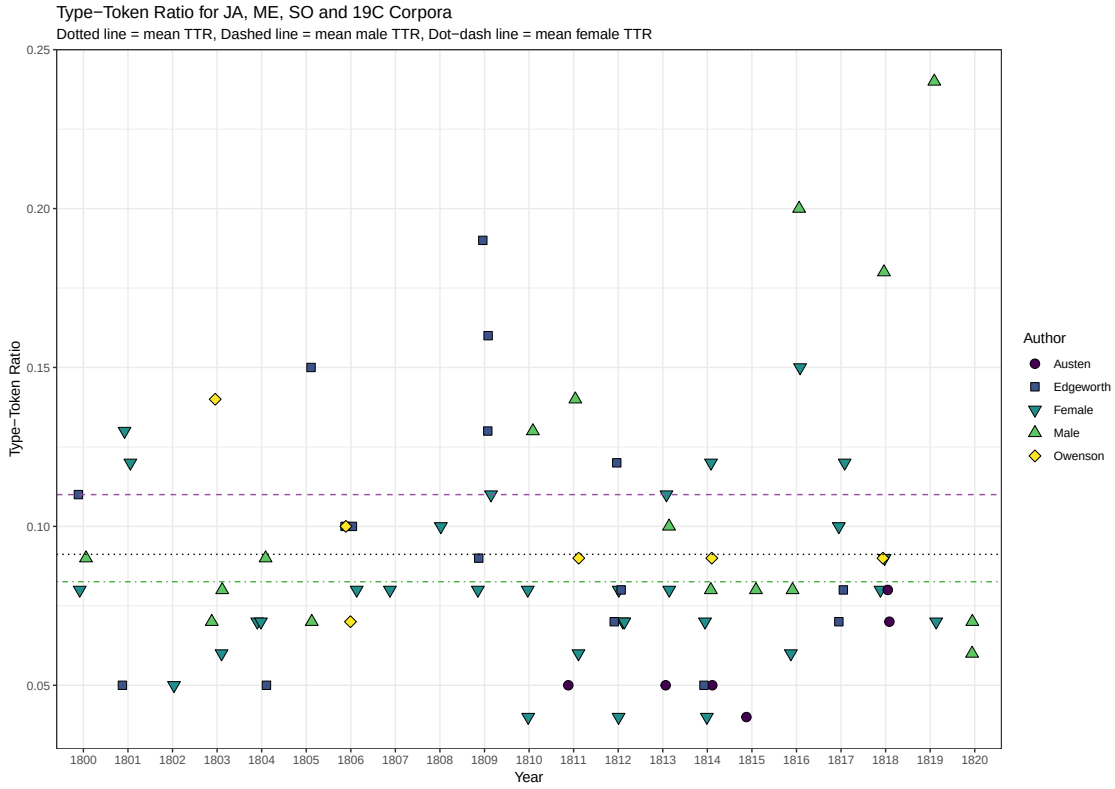


FIGURE 4.3: Type-Token Ratio for Novels in Combined Austen, Edgeworth, Owenson, and 19th Century Corpus

To consider whether the differences between the mean type token ratios (TTR) for each corpus were statistically significant, two statistical tests were carried out. Firstly the homogeneity of variance was assessed using the Fligner-Killeen test, a nonparametric test which is less sensitive to departures from normality. This reveals a p-value greater than 0.05 ($p\text{-}value = 0.108$) which indicates that there is no significant variance between the corpora and therefore an analysis of variance could be calculated. Due to the small sample size for some of the corpora (for example both the Austen and Owenson corpora have only six texts each) this fell below the threshold for using the parametric one-way ANOVA test. Therefore the non-parametric Kruskal-Wallis rank sum test was

Corpus	Mean TTR
Austen	0.057
Edgeworth	0.100
Owenson	0.097
19th Century Male	0.110
19th Century Female	0.083
Combined AEO and 19th Century	0.091

TABLE 4.1: Mean Type-Token Ratio by Corpus

used and this produced a statistically significant p-value ($p\text{-}value = 0.017$). Dunn's multiple comparisons post hoc test, utilising the Bonferroni method for p-value adjustment, was used to identify the specific areas which were statistically different. There was a statistically significant difference between the mean TTR of Austen and Edgeworth ($p\text{-}value = 0.048$), and between those of Austen and male novelists from the same period ($p\text{-}value = 0.026$). This reinforces the suggestion that Austen's low level of lexical variety was something unusual, especially when compared to that of her male contemporaries.

However, although the mean TTR of the Austen and the Edgeworth corpora are significantly different over the 1800–1820 period, Figure 4.3 suggests that this may not be true for the Edgeworth's later novels. After 1812, Edgeworth's lexical variety reduces, while Austen's final two novels (*Northanger Abbey* and *Persuasion*) present an increase, indicating that her sparse lexical style was a deliberate aspect of her writing. Edgeworth's later novels, *Patronage*, *Harrington*, and *Ormond*, match the TTR of Austen's *Mansfield Park*, *Northanger Abbey*, and *Persuasion* respectively.[2] Owenson's TTR also reduces after her

[2]The TTR for each of the novels in the combined Austen, Edgeworth, Owenson and 19th Century corpus can be found in Appendix E.

earlier novels, and settles at a level in line with the mean TTR for the combined Austen, Edgeworth, Owenson and 19th Century corpus. Running the Kruskal-Wallis test on a subset of the combined corpus including only texts from 1812 onwards confirms this, showing no significant differences. This suggests that Edgeworth and Owenson's lexical variety reduced as their style developed towards that of the realist novel, while Austen's published work seems to have been far more consciously edited for publication, explaining the slight rise in TTR in her posthumous novels.

4.1.3 Collocations

Collocations are calculated based on the raw frequency of words and the output is read vertically with the focus term in the centre and the words that appear before (-1) and after (+1) the focus term listed either side, organised first by frequency and then alphabetically.

Austen's collocates of 'independent' (Table 4.2) at position +1 largely reflect the physical and financial: 'fortune', 'property', 'resources', but there are also terms which suggest personal action: 'employments' and 'walk'. Interestingly, the only character name which appears is that of 'Churchill', emphasising the significance of Frank Churchill in *Emma* in Austen's discussion of the impact being independent, or not, could have. The collocates of 'independence' at -1, as seen in Table 4.3, reveals a broad consideration of the impact of independence, while it could be 'happy' or 'tolerable', it could also be 'audacious' or 'conceited'. Austen also refers to 'quiet' or 'small' independence (not shown in

Frequency -1	Word -1	Word	Word +1	Frequency +1
4	and	independent	of	17
3	of		fortune	5
2	quite		and	1
1	alone		as	1
1	an		because	1
1	are		employments	1
1	be		even	1
1	being		i	1
1	churchill		looking	1
1	evil		mrs	1
1	family		besides	1
1	for		property	1
1	fortune		resources	1
1	he		that	1
1	her		there	1

TABLE 4.2: Austen's Top 15 Collocates of 'Independent'

the Table 4.3), suggesting a moderate sum allowing the freedoms and security which Austen herself never quite achieved.

Edgeworth's collocates (Table 4.4) at -1 include adverbs which suggest that being independent could be viewed as being on a continuum which moved from 'quite' or 'generally' independent to 'entirely', 'perfectly', 'truly' or 'wholly'. There is also a sense that being independent was not a fixed status, as one could 'become' independent, or have been 'originally' independent. Her collocates at position +1 reflect the different aspects of being independent, the financial: 'fortune', 'maintenance', 'establishment'; the political: 'Englishman', 'country', 'candidate', 'constituents', and the personal: 'spirit', 'virtue'. Edgeworth's usage presents independence (Table 4.5) as belonging to a person or group: 'my', ''their', 'his', 'own', 'your', as well as being something which might require action ('asserting', 'ensure') or something which is to be aimed

Frequency -1	Word -1	Word	Word +1	Frequency +1
8	and	independence	and	9
5	of		of	8
4	the		a	2
3	that		had	2
2	an		mr	2
2	happy		she	2
2	tolerable		the	2
1	audacious		was	2
1	complete		which	2
1	conceited		at	1
1	considerable		besides	1
1	consideration		beyond	1
1	early		but	1
1	her		by	1
1	his		impossible	1

TABLE 4.3: Austen's Top 15 Collocates of 'Independence'

for ('hopeful').

Owenson's collocates of independent (which can be seen in Table 4.6), in part, echo several of Edgeworth's, one could 'become', be 'equally', 'evidently', 'very' or 'wholly' independent. Again, the words at +1 reveal societal and political differences: 'borough', 'chief', 'class'. However, Owenson's collocations go further than Edgeworth's here, focusing on active personal thought as an aspect of independence: 'intentions', 'judgements', 'principles'. Owenson's collocates of independence (Table 4.7) reveal a more political approach: 'American', 'comparative', 'national'. In the collocates at position +1 there is an implied sense of action: 'carried', 'flows', 'secured'.

Focusing on the collocations which specifically relate to terms associated with the theme of independence can quickly highlight an author's approach to that term. For example, while all three authors discuss education in their

Frequency -1	Word -1	Word	Word +1	Frequency +1
11	an	independent	of	17
4	totally		fortune	5
3	and		spirit	4
3	as		and	3
2	herself		he	3
2	himself		in	3
2	honest		man	3
2	is		country	2
2	me		for	2
2	of		i	2
2	perfectly		maintenance	2
2	the		a	1
1	affections		as	1
1	all		but	1
1	amusement		by	1

TABLE 4.4: Edgeworth's Top 15 Collocates of 'Independent'

Frequency -1	Word -1	Word	Word +1	Frequency +1
17	of	independence	of	10
5	and		and	7
2	my		which	3
2	their		he	2
1	above		I	2
1	an		or	2
1	asserting		their	2
1	ensure		to	2
1	him		a	1
1	his		all	1
1	honest		as	1
1	hopeful		by	1
1	in		did	1
1	independent		establishment	1
1	masculine		for	1

TABLE 4.5: Edgeworth's Top 15 Collocates of 'Independence'

Frequency -1	Word -1	Word	Word +1	Frequency +1
3	and	independent	of	40
3	for		and	2
3	the		besides	1
2	but		borough	1
2	is		chief	1
1	affection		class	1
1	alliance		direction	1
1	an		fortune	1
1	any		intentions	1
1	becomes		interest	1
1	character		judgements	1
1	culpable		nor	1
1	desmond		principles	1
1	equally		there	1
1	essence		sort	1

TABLE 4.6: Owenson's Top 15 Collocates of 'Independent'

Frequency -1	Word -1	Word	Word +1	Frequency +1
8	the	independence	of	11
6	of		for	2
5	and		the	2
3	for		to	2
2	an		already	1
2	her		and	1
1	able		animated	1
1	american		are	1
1	as		as	1
1	comparative		by	1
1	enviable		can	1
1	handsome		carried	1
1	affections		ci	1
1	irritable		ever	1
1	it		flows	1

TABLE 4.7: Owenson's Top 15 Collocates of 'Independence'

novels (as seen in Chapter 1) their collocations reveal that their individual focus is quite different. Edgeworth and Owenson present education as having a national flavour: "English", "European", "Irish", while Austen's discourse centres on the individual. This can be interpreted in a number of possible ways: firstly, that Austen's view is tempered by her own experience, which was much more restricted than that of Edgeworth and Owenson.[3] Secondly, it is worth considering that while Edgeworth and Owenson are writing with a view to the national or international status of education, Austen's interest lies in the effect of education at a personal level. It may also be argued that this narrow focus, the "little bit (two inches wide) of ivory on which I work with so fine a brush"(Jane Austen 1870), is a core part of Austen's style. While her comments appear to focus on the individual, her narrative style implies ideas which are to be considered at a societal level. It is through her collocations that we perceive a much more nuanced attitude: education may be both positive and negative. There is also a much more gendered perspective of education in Edgeworth and Austen's novels. The importance of a broad education, both in content and in morals, for men and women, is emphasised. While we cannot draw a direct line between the works of Wollstonecraft and those of Austen, Edgeworth and Owenson, this does suggest that all three authors were, at least, sympathetic to her calls for an equal education for women. This places all three authors on the more radical side of the arguments in favour of female education.

[3] Although Austen's personal schooling was limited to home and a brief time at the Abbey School in Reading, her brothers James and Henry went to Oxford University. During this time James and Henry produced a weekly humorous periodical *The Loiterer* which ran for about 60 issues. It is hard to believe that Austen did not read this (Geng 2001).

4.2 Topic Modelling

The English language provides the would-be analyst with some challenges. As Marc Alexander explains, when discussing the entries in the *Historical Thesaurus of English*, "62% of English word forms refer to more than one meaning ... 67 [words] have more than 100 possible meanings" (Alexander 2015). This demonstrates the difficulty of interpreting the results of frequency-based analysis. "[U]sing a concordance program that produces keyword-in-context lists is a good start" to addressing this problem, "because such lists offer context, but these lists are hardly practical when it comes to an analysis of anything beyond even a few thousand occurrences of a key word in a given corpus" (citing Biber et al. 2006 Jockers 2013a, p. 121).

To some extent, topic modelling is a reaction to the practicalities of extending analysis to a larger scale. As Jockers explains: "topic modeling provides a scalable method for word-sense disambiguation that specifically addresses the limitations of traditional collocation and KWIC lists ... Unlike KWIC and collocate lists, which require careful human interrogation in order to parse out one word sense from another, topic modeling works in an unsupervised way, inferring information about individual word senses based on their repeated appearance in similar contextual situations" (Jockers 2013a, p. 124).

Topic models are usually created with very large corpora, of hundreds or thousands of texts. A challenge when using a smaller corpus of novels is that character names tend "to bias the model and make it collect collocates that cluster around character names"; this leads to "broad novel-specific topics" rather than an insight into themes which appear in a corpus (Jockers 2013b,

np). That being said, the specific nature of the 'topics' is not as definitive as may be implied by the name of the method. Ted Underwood explores the differing types of 'topic' in his blog post 'What kinds of "topics" does topic modeling actually produce?', arguing that the nomenclature 'topics' is misleading and suggesting that 'discourses' may be more appropriate (2012, np), an idea echoed by Akira Murakami et al. (2017). This point may seem to be somewhat pedantic. However, it is important because understanding the output of a topic model requires the understanding of the nature of the 'topics', how they are achieved, and what they show.

The output of a topic model still requires considerable human intervention in order to understand and interpret it. As Jockers acknowledges, "[t]he resulting 'topics' in a given collection of texts are not provided to the researcher in the form of labeled themes ... but rather as sets of words, which are 'ranked', or 'weighted', according to their probabilities of appearing together in a given topic. The ease with which these resulting word clusters can be readily identified as topics or themes is referred to as topic 'coherence' or topic 'interpretability"' (Jockers 2013a, pp. 124-5). This emphasises the human role in understanding the topic, although it is worth noting that individual topics within a model will contain every word in the corpus, in different positions according to probability. Therefore the differences between topics are due to the position of each word within the topic. This indicates that topics should be seen as less definitive than the term 'topic' may suggest, as the difference between assigning a text to a particular topic rather than another may be based on a very small difference in percentages.

"Interpretability is of key concern to us since we are most interested in investigating and exploring themes that we understand and recognise. This is an important point to dwell on because one criticism of unsupervised models is that they are black boxes" (Jockers 2013a, p. 128). This is at the crux of digital methods in general, and in the more algorithmically driven analyses in particular. The argument against 'black box' analyses is rooted in not knowing how the process works - however, to some extent, this could also be argued to be the case in traditional literary analysis, for how can we fully understand the input which has led to the scholar's interpretation. Jockers' words "we are most interested in" and "themes that we understand and recognise" highlight the possibility of confirmation bias: that we will often find within a text that which we are looking for.The interpretability of the topics perhaps misses the point of what topic modelling is actually doing. The algorithm has no understanding of the texts and their themes, and is not trying to create 'topics' *per se*. Rather, it is calculating a series of probabilities, and therefore, to expect the output to be entirely 'human friendly' is misunderstanding the tool and its limitations.

It would also be prudent at this juncture to address the potential controversy regarding the removal of stop words from the corpus, especially as part of preprocessing the texts for topic modelling. Matthew Jockers hints at this contentious issue in his 2013 blog post where, after discussing the use of a Part of Speech tagger to reduce his chosen corpus to nouns, he warns that "I think this is a good way to capture *thematic* information; it certainly does *not* capture such things as *affect* (i.e. attitudes towards the theme) or other *nuances* that may be *very* important to literary analysis and interpretation"

(Jockers 2013b, np). The topics which emerge from the topic model may be made more 'coherent', or humanly readable, depending on what is excluded from the text (e.g. through the use of stop word lists or using Jockers' 'Secret Recipe' (Jockers 2013b) of removing word classes) and therefore this may indicate the terms or word forms which make up the inherent 'topicness' of the text. In contrast, as illustrated by Burrows and Hoover (Burrows 1987; Burrows 2002; Hoover 2004a; Hoover 2004b; Hoover 2010), these non-topic, excluded words appear to be far stronger indicators of style and, as seen above in Section 4.1.2, gender.

In this thesis, the removal of stop words and other specific parts of speech, has been kept to a minimum as much as is possible. Where stop words have been removed the list of terms is available in Appendix C, other parameters are noted in the text below.

Matthew Jockers highlights that "the quality of the final model ... is largely dependent upon preprocessing" (2013b). Jockers suggests a three step process, firstly removing common stop words, then using a part of speech (PoS) tagger to identify and extract the common nouns, and finally to break the texts within the corpus into chunks. Each of these stages removes some of the noise or 'bias' from the model. However it is important to note what has been removed and consider the possible implications and biases this in turn may create.

The process suggested by Jockers was used as a guideline but altered in a number of areas. Firstly, the texts were tagged using the TreeTagger PoS tagger, rather than the Stanford Tagger. It is important to note that PoS

tagging is not one hundred percent accurate, although accuracy levels may reach the mid to high ninety percent range. The taggers are trained on a bank of tagged texts, in the case of both the Stanford Tagger and TreeTagger the Penn Treebank (Marcus et al. 1999), which may themselves not be completely accurate (Christopher D. Manning 2011). In addition, these treebanks are usually based on large collections of modern texts (for example media texts) and therefore the level of accuracy when applied to nineteenth century novels is likely to be lower than when applied to a modern corpus (Schneider, Hundt, and Oppliger 2016). TreeTagger was chosen as its implementation in R was faster and more stable than the Stanford Tagger or the Spacy Python-based tagger, also, Tobias Horsmann, Nicolai Erbs, and Torsten Zesch's 2015 study which compared the accuracy of 22 PoS speech tagger models, found that TreeTagger was the most accurate for formal written English, with an accuracy of 93.9% (2015). As the corpus used for this thesis is smaller than those typically used for topic modelling, and to reduce the chance of removing too much cogent material, it was decided that both nouns and verbs would be kept, with the exception of proper nouns and the verbs *to be, to do, and to have*. The tags used for the tagged corpus are listed in Table 4.8. The extracted nouns and verbs were then split into chunks of 1000 words.

4.2.1 Combined Corpus

Topic models are "generative models which provide a probabilistic framework for the term frequency occurrences in documents in a given corpus" (Grün and Hornik 2011, p. 1). As such the resulting model is not a fixed entity

PoS Tag	Description
NN	noun, singular
NNS	noun, plural
VV	verb, base form
VVD	verb, past tense
VVG	verb, gerund/participle
VVN	verb, past participle
VVP	verb, present, non-3rd person
VVZ	verb, present 3rd person singular

TABLE 4.8: Part of speech Tags and Definitions

and will generate a slightly different set of topics each time the modelling function is run. The first set of topics allocated to a text represents the topic with the highest probability. However, the topics at positions two, three or lower may not have a significantly lower probability than those at position one. Therefore the model is more usefully viewed as a representation of the text and its themes or discourses, which needs to be interpreted.

The combined 19th Century, Austen, Edgeworth, and Owenson corpus was tagged and chunked as outlined above. A document term matrix (DTM) was created, and the latent Dirichlet allocation (LDA) with Gibbs sampling function in the `topicmodels` package (Grün and Hornik 2011) was used to create the topic model, with the following parameters: `k = 37, nstart = 5, seed = list(123, 420, 224, 10098, 49), best = TRUE, burnin = 500, iter = 2000, thin = 500`. The parameter `k` is the number of topics chosen for the model. This was selected after running the tuner in the `ldatuning` (Murzintcev 2016) package which suggested an optimum number of topics as being 37. The `seed` parameter ensures that the model can be reproduced. The remaining parameters "control how many Gibbs sampling draws are made" (Grün and Hornik 2011, p. 9) and which are saved in the finished model:

`nstart` indicates the number of runs, and `best = TRUE` means that "only the draw with the highest posterior likelihood over all runs is returned" (Grün and Hornik 2011, p. 10) (the posterior likelihood is the probability that a text belongs to a specific topic given all previous information). The final parameters are: the burn in period (`burnin`) which defines the number of runs which are discarded at the start of the process, `iter` which is the number of iterations the model with run through, and `thin` which indicates which of the iterations to keep. The burn in process is an essential part of the topic model, as each time the process runs the algorithm 'learns', or uses information from previous iterations, to calculate the probabilities for the current iteration. This means that the initial runs are less accurate and therefore should be discarded.

Models were created for the suggested 37 topics, as well as 20 topics. The top ten terms for each topic in the 20 topic model can be found in Appendix F. The 37 topic model, seen in Figure reffig:all37, seems to produce a rather more granular set of topics, splitting some of the broader topics in the 20 topic model into subcategories. Some topics, e.g. 'king, country, castle, heart, friend' are very similar across both models (topic 15 in the 37 topic model and topic 18 in the 20 topic model).

Topics 14 and 22 in the 37 topic model are suggestive of the public and private spheres deemed appropriate for men and women: 'lord, lordship, sir, son, family, law, business' and 'lady, conversation, taste, world, fashion, beauty, attention'. This may be viewed as the serious world of the 'lord' and the frivolous and fashionable world of the 'lady'. Female nouns in the model are generally found in relation to other terms relating to the family. A thematic topic relating to marriage (topic 33): 'husband, wife, woman, dear, love, bride,

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
1	house	room	ladyship	sir	sea	brother	eyes	privilege	lady	time	people	convent	friend	lord	king	nothing	mrs	death	honour	moment
2	time	door	doctor	squire	vessel	uncle	feelings	hobby	chevalier	situation	things	abbess	man	lordship	country	time	person	soul	man	night
3	friend	lady	aunt	skull	boat	sister	friends	blow	duke	place	men	time	heart	sir	castle	thing	stranger	heart	horse	light
4	death	hand	dear	tower	shore	nephew	years	acquisition	baron	circumstance	gentlemen	arch	mind	son	heart	man	way	hand	house	church
5	condition	house	countess	wine	land	niece	others	extreme	count	degree	years	bishop	time	family	friend	way	moment	bosom	master	door
6	purpose	night	niece	butler	ship	sir	words	impudence	minstrel	length	ladies	nun	love	lady	man	day	situation	life	country	voice
7	nothing	morning	mamma	ghost	board	curate	days	multitude	chateau	idea	children	order	letter	law	hand	mother	word	voice	boy	man
8	knowledge	woman	picture	system	voyage	ring	spirits	ridicule	king	part	hands	part	happiness	business	prince	father	sight	horror	farm	day
9	city	man	instant	terrace	coast	kind	people	safeguard	lord	mind	times	church	power	commissioner	head	something	air	thy	year	power
10	place	face	bird	music	captain	regiment	circumstances	bustle	father	attention	horses	novice	moment	earl	sword	house	sort	mind	agent	cell

Topic 21	Topic 22	Topic 23	Topic 24	Topic 25	Topic 26	Topic 27	Topic 28	Topic 29	Topic 30	Topic 31	Topic 32	Topic 33	Topic 34	Topic 35	Topic 36	Topic 37
jew	lady	madame	hae	page	religion	eyes	sister	heart	book	sir	men	husband	night	father	money	honor
jewess	conversation	monsieur	sae	youth	priest	thou	aunt	mind	man	man	arms	wife	time	mother	day	stranger
humph	company	mademoiselle	nae	thy	faith	hands	family	soul	country	gentleman	friends	woman	man	child	master	country
emperor	taste	enchantress	highlanders	castle	earth	tears	home	moment	author	antiquary	hands	dear	way	daughter	man	castle
bag	world	friend	ruins	thou	zeal	feelings	day	nature	genius	time	soldiers	love	light	son	business	person
tapestry	fashion	duke	circumstances	church	look	words	brother	love	character	house	troops	bride	place	life	sum	day
toleration	beauty	miss	occasions	man	devotion	feet	pleasure	hand	history	day	people	marriage	ground	time	shop	life
fan	attention	mon	matters	apartment	thou	arms	room	eye	work	country	enemies	adieu	distance	family	sir	place
farce	character	abbé	dogs	grace	altar	lips	subject	life	life	head	followers	grace	spot	house	boy	ould
fatality	thing	hotel	thae	madam	tree	steps	letter	object	name	person	eyes	leonora	side	heart	time	family

FIGURE 4.4: Top Ten Terms for Combined Corpus - 37 Topic Model

marriage', is separate from terms we might associate with marriage today: 'heart, mind, soul', reflecting some of the differences in discussed in Chapter 1. Topics 5 ('sea, vessel, boat, shore, land, ship'), 19 ('honour, man, horse, house, master'), and 36 ('money, day, master, man, business') reflect some the most common routes to financial independence for men, seafaring, land ownership and business.

FIGURE 4.5: 37 Topic Model for Combined Austen, Edgeworth, Owenson, and 19th Century Corpus

Figure 4.5 plots each text segment in the combined Austen, Edgeworth, Owenson, and 19th Century corpus by topic and by sub-corpus. This indicates the topics which are most frequent for each of the authors, and for the male and female writers in the 19th Century corpus. The topics to which the highest

number of text segments were allocated were: Topic 16 ('nothing, time, thing, man, way') for Austen, Topic 13 ('friend, man, heart, mind, time') for Edgeworth, and Topic 29 ('heart, mind, soul, moment, nature') for Owenson. It is notable that Austen and Edgeworth both have a significant number of segments in Topic 7 ('eyes, feelings, friends, years, others, words'), and that they have very few topics in common with Owenson. Indeed, Owenson is the only author whose text segments are allocated to Topic 37 ('honor [sic], stranger, country, castle, person').

4.2.2 The Austen, Edgeworth, and Owenson Corpus

Using the same settings as for the 37 topic model, a 20 topic model was created for the Austen, Edgeworth, and Owenson corpus. The top 20 terms for each of the 20 topics are listed in Appendix G.

Examining the top 20 terms for the 20 topic model reveals several topics which might suggest elements of independence. Topic 13 ('honour, man, house, horse') and Topic 15 ('gentlemen, horses, men, guineas') could be viewed as a topic relating to wealth and financial independence. Topic 7 ('priest, country, castle, book, harp, music, song, history') might suggest a topic relating to national independence, or cliches in Irish literature of the time. Figure 4.6 shows the topics assigned to the novel segments of Austen, Edgeworth, and Owenson. Austen and Edgeworth are shown to have two topics in common, Topic 1 ('people, friends, eyes, feelings, things, words') which could be viewed as a theme of feelings or reflection, and Topic 3 ('man, time, manner, thing, opinion') which could be seen as being related to independent thought. What

is also evident is that, in the smaller corpus, Austen's topics remain consistent across her novels. We can also see a shift in the topics over time, with Topic 19 (room', 'person', 'moment', 'castle', 'honor', 'country') being more significant in Owenson's later novels.

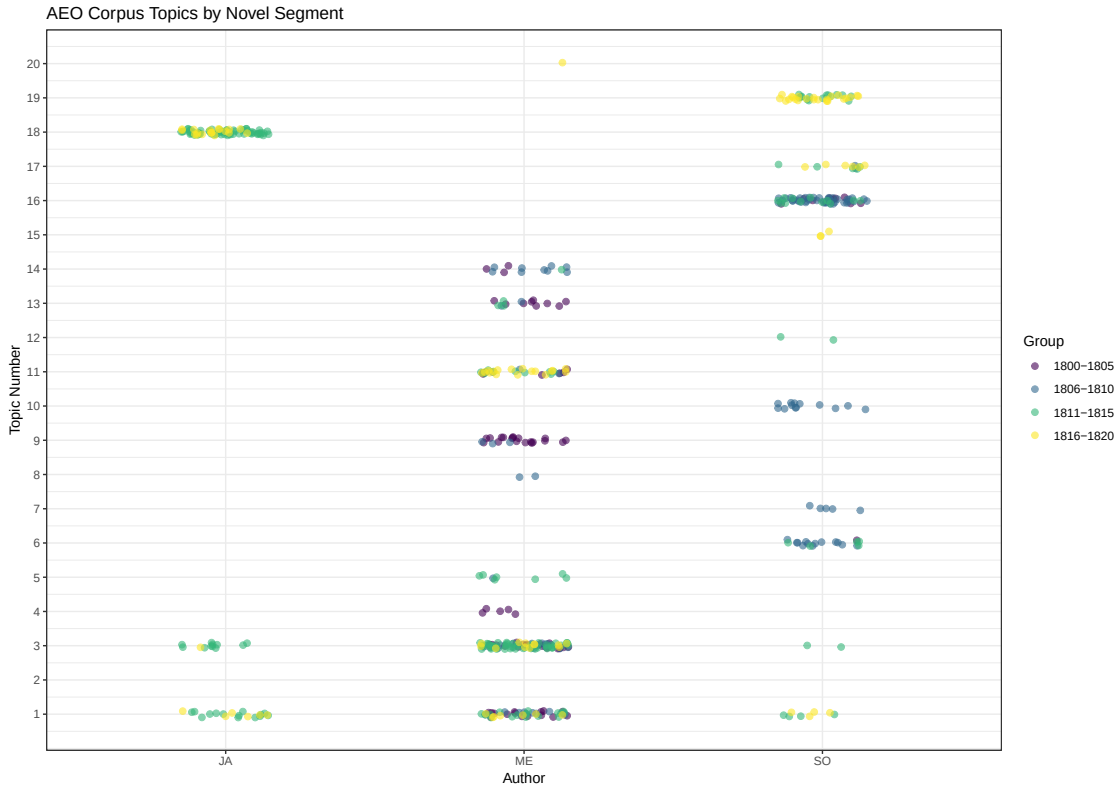


FIGURE 4.6: 20 Topic Model for Combined Austen, Edgeworth, Owenson Corpus

However, the difficulty lies in identifying how, or even if, these topics specifically relate to the theme of independence. Without examining the terms in the context of the individual texts we cannot say whether the topic is actually related to the area of interest. Therefore, for the purpose of this study, topic modelling and frequency analysis were used to gain an overview of the corpora as a whole, and within the context of nineteenth century novels.

The identification and removal of stop words, as discussed above, is something which has been debated throughout the development of text mining methods (Don et al. 2007; Sculley and Pasanek 2008; Turney and Pantel 2010). However there is no consensus on whether they should be removed or not. The question of term reduction raises several concerns, especially amongst traditional, literary scholars, regarding the nature of the text. Can the text still be viewed as inherently the same when a variety of stop words, or even a whole class of words, have been removed? For topic modelling, this decision to reduce the terms used in the model may be due to the desire for more coherent topics, or simply due to the practicalities of applying algorithmic analysis to a large corpus. Helpfully, Jockers illustrates the impact of part of speech tagging, showing the reduction of a text to just nouns, using Austen's *Pride and Prejudice* as an example. "In this manner, Austen's 'It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife. However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters' is transformed into 'truth man possession fortune want wife feelings views man neighbourhood truth minds families property other daughters"' (2013a, p. 133). While this approach may be useful, it is obvious that whittling Austen's opening lines into a string of nouns removes the nuance behind the opening and may suggest meanings that are actually in opposition to what the full text actually means. Therefore, this is a balancing act, and while it may be necessary for achieving clarity in the output of some methods, its impact must be considered.

Ultimately, it is a balance between determining an acceptable level of noise and risking the validity of the model: "if an analyst is too aggressive in removing words, the resulting models may be biased towards what the analyst views is important in a corpus" (A. Schofield, Magnusson, and Mimno 2017, p. 432). This is illustrated by Jockers, in the example above from *Pride and Prejudice*: "I must offer the caveat, therefore, that the noun-based approach used here is specific to the type of thematic results I wish to derive" (2013a, p. 133). It is vital in the name of transparency, therefore, to outline which steps have been taken during preprocessing, and to decide specifically where the line is to be drawn for a particular study. Alongside this, little, if any, research considers the implications of the author's type token ratio in the results achieved through topic modelling. Therefore this might be a fruitful area for further research. The next chapter examines some methods which may help to mitigate these problems.

5 Word-Context Models

> I hate to hear you talking so like a fine gentleman, and as if women were all fine ladies, instead of rational creatures. We none of us expect to be in smooth water all our days
>
> Jane Austen, *Persuasion*

While frequency analysis, topic modelling and other term-document models reveal large scale patterns and potential topics within a corpus of texts or allows individual terms to be targeted; word-context models allow for a simultaneously broad and more focused approach. To use the map analogy from Chapter 2, while term-document models may reveal the contours of a landscape, the word-context model reveals the specific features, so what appeared to be a hillside is revealed to be an ancient hill fort or barrow. The creation of a model for the corpus, where the relationships between the words is mapped within a context window, allows us to ask 'what does this corpus tell us about topic x?', and enables us to be able to explore the discourse surrounding term x.

5.1 Word-Embedding Models

5.1.1 An Independence Lexicon

A challenge in analysing texts is being able to identify whether the words an author uses are unusual in their context or whether they are a typical usage for the time. In order to explore the theme of independence, and where possible to consider the meanings and connotations of independence during this period, a lexicon of terms was created. Using 97 texts from the second period (1780–1850) of *CLMET3.1*, excluding 'CLMET3_1_2_133.txt' and 'CLMET3_1_2_134.txt' (Austen's *Sense and Sensibility* and *Pride and Prejudice*), totalling just over 11 million words, a word embedding model was created with 300 dimensions and a window of 12 words using the skip-gram word2vec model (as outlined in Chapter 3).[1] The 500 closest words to the vector for 'independent' and 'independence' were identified. The lexicon provides us with an insight into the discourse surrounding independence during this period in a range of texts from a broad selection of genres, from letters, treatise and narrative non-fiction, to drama and narrative fiction. With this baseline, we can get an indication of how typical the discourses we find in Austen, Edgeworth, and Owenson's writing were for the period.

The top ten terms for the *CLMET* corpus are compared with the top ten terms for the Austen, Edgeworth and Owenson corpora in Table 5.1. This table reveals some initial insights into the different discourses surrounding independence in the four corpora. The *CLMET* terms are relatively broad, but

[1] Each text in the *CLMET* corpus is named in the format CLMET3_1_PERIOD_TEXT REFERENCE.

CLMET	Austen	Edgeworth	Owenson
independence	independence	independence	independence
freedom	privilege	respectability	distinctions
ennobling	income	greatness	patriots
rights	decorum	professions	faults
integrity	provision	diminishing	political
suppleness	nicety	subordination	asserting
talent	affluence	precluded	affluence
humanized	masters	rudeness	practices
advocating	pecuniary	virtues	supremacy
claim	habits	appreciate	ruling

TABLE 5.1: Comparison of Top 10 Terms in 'independence' word2vec Models

with several terms related to freedom and rights, therefore containing a rather more political aspect than might be found in Austen or Edgeworth's novels. While the three author corpora share some similarities with the broader *CLMET* corpus, especially beyond the top ten terms, each has an individual focus or discourse surrounding independence. Austen's top terms suggest a focus on financial independence (***privilege***, ***income***, ***provision***, ***affluence***), while Edgeworth's seem to focus more on rank and status (***respectability***, ***greatness***, ***subordination***), and Owenson's has a more obvious political aspect (***patriots***, ***political***, ***supremacy***, ***ruling***). This snapshot echoes what we might expect to find based on a close reading of the novels, and supports the use of word2vec as an effective method for identifying relevant textual information.

While making comparisons between the wordlists created by the `wordVectors` package (Schmidt and Li 2015) can provide some initial insights into the discourses surrounding independence presented in the authors' works which

largely support those which may be gained through close reading, looking beyond wordlists of the top ten or twenty words is far more revealing. Using the `w2v_analysis()` function (seen in Appendix D), not only are a list of top terms related to the focus term extracted, but the relationship between the terms is represented in a two dimensional plot. The plot uses `Rtsne` to reduce the 300 dimensions in the model to something human readable, therefore it is a representation of the relationships between the terms rather than an exact replication. As noted in Chapter 3, due to the iterative nature of t-SNE as a dimension reduction tool, the axes are not relevant to the interpretation of the plot; rather, the relationship of the terms in relation with each other are the important area for focus.

The resulting word embedding plots may be used in a number of ways to support reading; while the creation of the plot is unsupervised, the analysis and interpretation is a human endeavour, hence 'enhanced reading'. Firstly, the plot can be explored to identify clusters of terms which relate to a specific aspect of independence. For example in Figure 5.3, which is based on the Austen corpus, there is a cluster of terms grouped around ***education***, including ***improvement***, ***illiterate***, ***talents***, and ***accomplishments***. The focus cluster may be used to explore the corpus in greater detail through close reading, or to compare with word embedding plots for other corpora to identify broader patterns based on the sub-theme. An alternative method is to identify individual terms of interest, either from the plot itself or from the list of terms produced as part of the `w2v_analysis()` function, which may lead to a "serendipitous discovery" (Nicholas and Herman 2009, p. 22). Bridging the gap between the terms within the plot, which have their context hidden

from view, and close reading requires a Key Word in Context (`kwic()`) function which provides a snapshot of the context from which the term has been extracted (as seen in Figure 5.1), allowing the reader to determine whether the term is of use.

fireside

	file	position	left	keyword	right
1	JA_1814_MP.txt	62759	hurried walks and sudden removals from her own	fireside	and many excellent hints of distrust and economy
2	JA_1815_E.txt	5947	win or lose a few sixpences by his	fireside	these were the ladies whom emma found herself
3	JA_1815_E.txt	36977	when he asks people to leave their own	fireside	and encounter such a day as this for
4	JA_1815_E.txt	100396	quit the tranquillity and independence of his own	fireside	and on the evening of a cold sleety
5	JA_1815_E.txt	153947	growing older ten years hence to have his	fireside	enlivened by the sports and the nonsense the
6	ME_1801_B.txt	107368	threw himself into an arm chair by the	fireside	protesting that he was tired to death that
7	ME_1804_PT.txt	127009	of building castles in the air by his	fireside	mrs bettesworth was a vain foolish vixen fond
8	ME_1804_PT.txt	188299	have been safe and snug by my own	fireside	with my wife poor lucy spent her time
9	ME_1806_E.txt	75284	master of the house he lived by the	fireside	disregarded in winter and in summer he spent
10	ME_1812_EC.txt	1287	entrench herself in an arm chair by the	fireside	sprinkled the floor round her with _eau de
11	ME_1814_P.txt	200962	a smiling wife a sympathizing friend a cheerful	fireside	she had musical talents her husband was fond
12	SO_1806_WG.txt	41460	arm and took his usual station by the	fireside	repeating _some claim _ while i was thinking
13	SO_1806_WG.txt	55433	make our offerings round the domestic altar the	fireside	of such fruits as the lingering season affords

FIGURE 5.1: Key Word in Context Output

In the following sections the four focus areas of independence outlined in Chapter 1 will be considered. Section 5.1.2 will explore marriage and conduct through an enhanced reading of *Mansfield Park*, Section 5.1.3 will examine education, employment, wealth, and rank, and Section 5.1.4 will present some of the main characteristics of nation and independence highlighted through the word embedding models.[2]

[2]Where terms are drawn from the word embedding plots they are indicated thus: ***word***.

5.1.2 An Enhanced Reading of *Mansfield Park*

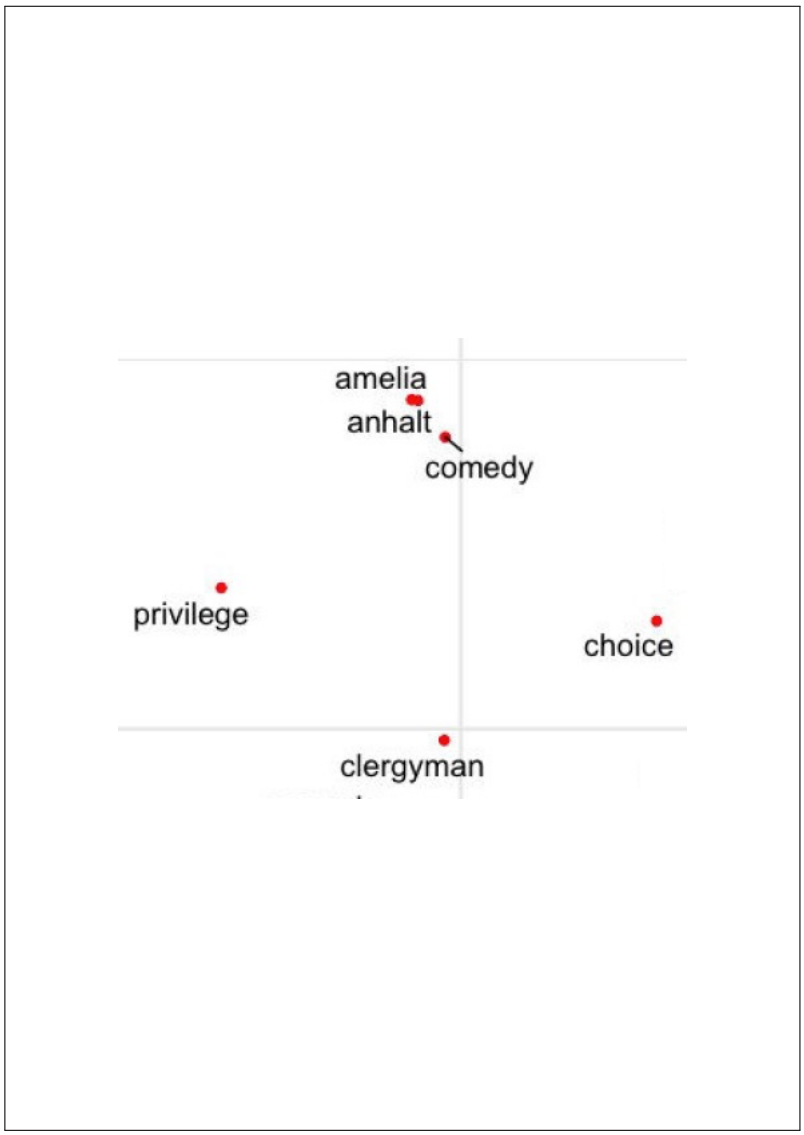


FIGURE 5.2: The *Mansfield Park* Theatricals

Mansfield Park is one of Austen's most challenging novels to interpret and has caused much debate amongst critics (Wiltshire 1997; Kirkham 1997; Tauchert 2003; D. Jones 2004; Jenkins 2006). It was also, as highlighted in Chapter 1, somewhat of a publishing disaster for Austen. The heroine, Fanny Price, has

been presented variously as: "a radically insecure and traumatized personality" (Wiltshire 1997, p. 60), "a pious Christian heroine, or as a character intended to expose the hypocrisy and salaciousness underpinning traditional representations of Christian heroineism" (D. Jones 2004, pp. 115–116), or even "a 'moral tarantula"' who "sits, making negative moral judgements about the actions of others, while doing nothing herself" (Jenkins 2006, p. 346). Certainly, it is difficult to believe that Austen created Fanny as an idealised model of "traditional feminine values and conservative moral thought" (Jenkins 2006, p. 346), especially given her letter to her niece Fanny Knight in which she states that "Pictures of perfection, as you know, make me sick and wicked" (Jane Austen 1884).

The word embedding model seen in Figure 5.3 can be used to provide some insight into the range of representations of independence in the novel. The grouping highlighted in Figure 5.2 — ***amelia***, ***anhalt***, ***comedy***, ***privilege***, ***choice*** — which initially, seems somewhat out of place in a model which focuses on 'independence'. Those familiar with *Mansfield Park* will recognise that it refers to an episode in which the characters take part in a home production of *Lovers Vows*. This is a central scene in the novel extending to eight chapters, spanning the end of the original volume 1 and start of volume 2.

Enhanced reading suggests that, although it may not be immediately obvious, this scene is at the centre of the novel's discourse on independence. We see the poor behaviour of Mary Crawford and the Bertram sisters as a result of uncontrolled independence, and poor moral education, as Austen takes pains to outline early in the novel. Maria and Julia "with all their promising talents and early information" were "entirely deficient in the less common acquirements of

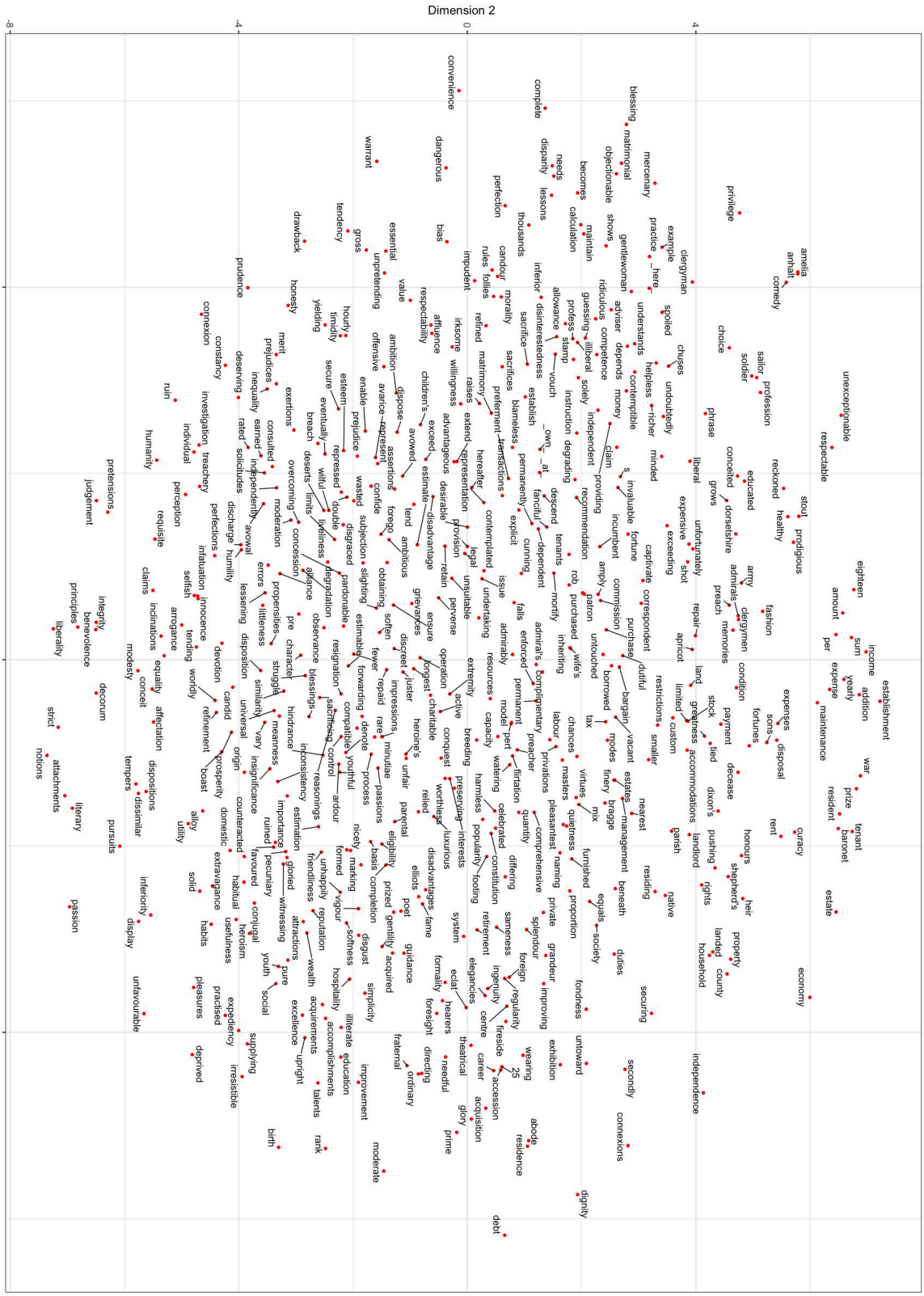


FIGURE 5.3: Word2Vec Model of Austen's Top 500 Terms Related to 'independent' and 'independence'.

self-knowledge, generosity, and humility. In everything but disposition, they were admirably taught" (Jane Austen 1994b, p. 18). Although the girls are intelligent, confident and good-looking, they demonstrate independence but without the moral background and "disposition" necessary. It is this failure in the sisters' education, and that of their brothers, which leads to the negative outcomes the family faces as a results of the events set in motion in this part of the novel. Fanny, on the other hand, is far less favoured than her cousins, with constant reminders of her dependent state. Although she has received a similar education to her cousins, following her move to Mansfield Park, the lack of formal education provided by her parents and her timidity places her at a marked disadvantage to Maria and Julia. It is clear that while Maria and Julia are intended to take their place in fashionable society, this is not the case for Fanny. With Edmund's help she has extended the limited education she has received: "Miss Lee taught her French, and heard her read the daily portion of history; but he recommended the books which charmed her leisure hours, he encouraged her taste, and corrected her judgement: he made reading useful by talking to her of what she read and heightened its attraction by judicious praise" (Jane Austen 1994b, p. 21).

It is the part of the novel where a production of *Lovers' Vows* is planned, however, where many of the characters are revealed in their worst light. The reckless behaviour of Tom Bertram and Henry Crawford, and the stupidity of Mr Rushworth and Mr Yates, hints at a criticism of the traditional model of primogeniture. The image of male independence we are presented with is one which is petulent, self-serving, ridiculous, or, in the case of Sir Thomas, absent. Darryl Jones, building on Roger Sales' work on representations in

Regency England, suggests that the theatre episode represents "a crisis of *masculinity*, implying the emasculation, effeminacy or gender destabilisation of virtually all of the younger generation of men in the novel" (D. Jones 2004, p. 43). Penny Gay emphasises the devilish role of Henry and Mary Crawford's unrestrained independence:

> Henry is happiest when he has persuaded everyone around him to abandon their principles and plunge into the intoxicating world of make-believe, of acting, when anyone may be, or do, anything without reference to the normal restrictions on their behaviour. Henry says (again it is a significant admission; here Austen allows us to see the true springs of his character), "There was such an interest, such an animation, such a spirit diffused! Every body felt it. *We were all alive.* There was employment, hope, solicitude, bustle, for every hour of the day" (p. 227, [author's italics]) — that is, only while acting does Henry feel really alive and purposeful; he has no other 'employment.' And Mary, much later, reiterates these sentiments, emphasising the pernicious power that the theatricals gave her: "If I had the power of recalling any one week of my existence, it should be that week, that acting week . . . I never knew such exquisite happiness in any other. His [Edmund's] sturdy spirit to bend as it did! Oh! it was sweet beyond beyond expression"[p. 353]. (Gay 1995, p. 122)

Examining the character of Fanny Price more closely, we can see that, although adhering to a moral code, she is positioned as a rational and independent thinker. ***Privilege***, ***choice*** and those who have the power to act

upon their choices, are at the heart of the action. While Edmund Bertram, a man who intends to become a ***clergyman*** and who has the independence of age, ***education*** and relative wealth, shifts from the moral and censorious member of the family to falling under the influence of Mary Crawford, it is only Fanny who demonstrates her independence, by refusing to take part despite vicious criticism, and who is ultimately proved to have been in the right. While seeming to uphold Fanny as an image of ***modesty*** and virtue, Austen has created a subversive character, a character who is in turn called to speak yet is silenced, who is weak yet also ***wilful***, who challenges the authority of those who were traditionally viewed as independent, and causes the reader to question their assumptions.

Austen positions Edmund, who up to this point has guided Fanny's moral and intellectual development, in opposition to Fanny. Edmund initially is presented as the family arbiter of moral behaviour in the absence of his father: "the resolution to act something or other seemed so decided, as to make Edmund quite uncomfortable. He was determined to prevent it if possible ...'I think it would be very wrong ...it would be highly injudicious ...It would shew great want of feeling on my father's account ...and it would be imprudent ...as to acting myself ... *that* I absolutely protest against"' (1994b, pp. 127–131). However, within the turn of a page we see his convictions begin to waver. On being told that Mary Crawford wants to be part of the theatricals, Austen informs the reader that "Edmund, silenced, was obliged to acknowledge that the charm of acting might well carry fascination to the mind of genius" (1994b, p. 132). From here it is a slippery slope before he is persuaded into taking part in the performance himself:

> I cannot think of it with any patience; and it does appear to me an evil of such magnitude as must, *if possible*, be prevented ... There is but *one* thing to be done, Fanny. I must take Anhalt myself ... It is not at all what I like ... No man can like being driven into the *appearance* of such inconsistency ... I see your judgement is not with me. Think it a little over ... Put yourself in Miss Crawford's place ... it would be ungenerous, it would be really wrong to expose her to it. Her feelings ought to be respected ... If you are against me, I ought to distrust myself, and yet — (1994b, pp. 157–158).

Fanny's judgement of Edmund is revealed through free indirect speech: "To be acting! After all his objections — objections so just and so public! After all that she had heard him say, and seen him look, and known him to be feeling. Could it be possible? Edmund so inconsistent!" (1994b, p. 160). Her view is clearly that this is far more than the "*appearance*" of inconsistency.

In contrast to Edmund's actions, Austen positions those of Fanny, without directly emphasising the comparison. While Fanny acknowledges that "[f]or her own gratification she could have wished that something might be acted, for she had never seen even half a play ... everything of higher consequence was against it" (1994b, p. 134). Fanny's position as dependent within the family places her in a position of powerlessness and obligation; however she remains constant in her decision repeating "no", "impossible", "cannot", and "you must excuse me" (1994b, pp. 149-150) while being pressured by her cousin Tom. Mrs Norris cruelly and publicly highlights Fanny's dependent position in the household: "I shall think her a very obstinate, ungrateful girl, if she does not

do what her aunt and cousins wish her; very ungrateful, indeed, considering who and what she is" (1994b, p. 150). Notably, Edmund's inaction during these exchanges is highlighted by Austen, firstly: " Edmund, who was kindly observing her; but unwilling to exasperate his brother by interference"; then, after Mrs Norris's outburst: "Edmund was too angry to speak" (1994b, p. 150), leaving Miss Crawford to comfort her.

A postscript to the theatricals appears in Chapter 20. Following Sir Thomas's return and condemnation of the acting in general, and Edmund's judgement in particular, we learn that "Edmund's first object the next morning was to see his father alone, and give him a fair statement of the whole acting scheme, *defending his own share in it* as far only as he could then, in a soberer moment, feel his motives to deserve ... He was anxious, *while vindicating himself*, to say nothing unkind of the others" (my italics) (1994b, p. 188). Such is our romantic hero: inconsistent, easily manipulated, slow to defend the weakest member of the family, and determined to justify his actions to his father. Austen presents us with a flawed hero, not truly worthy of Fanny's devotion.

From this point in the novel, we see a subtle shift in the character of Fanny Price - it is here where she first demonstrates her "independence of spirit" (1994b, p. 320) which allows her to reject Henry Crawford's proposal even in the face of her uncle's attempts to convince her through the classic elements of emotional blackmail — fear, obligation and guilt. Sir Thomas declares her "wilful and perverse" (1994b, p. 321) for her rejection, both terms which appear in Figure 5.3; the word ***wilful***, with its connotations of rebellion against authority, also appears amongst the terms related to independence seen in Figure 5.3. The term is used as a criticism of women who refuse to comply

with male expectations. Rather than suggesting that women without personal independence should accept an unwelcome proposal of marriage, Austen makes it clear that she disagrees with the speakers and champions the intellectual independence of her female characters to make their own choices.

Fanny Price knows her own mind even if those who know her best do not realise or accept this. Edmund in particular, while claiming to acknowledge her as a rational being in fact undermines her, firstly during the theatricals: "I see your judgement is not with me. Think it a little over . . . Put yourself in Miss Crawford's place . . . it would be ungenerous . . . Her feelings ought to be respected . . . If you are against me" (1994b, pp. 157–158). The second time occurs after her repeated rejection of Henry Crawford:

> 'So far your conduct has been faultless, and they were quite mistaken who wished you to do otherwise . . . This, we know must be a work of time. But' (with an affectionate smile) 'let him succeed at last, Fanny, let him succeed at last. You have proved yourself upright and disinterested, prove yourself grateful and tender-hearted; and then you will be the perfect model of a woman, which I have always believed you born for' (1994b, p. 351).

Not only does Edmund assume that Henry's proposal will eventually be successful, despite Fanny's protestations, he also seems to believe that her actions are for the sake of conforming to an idealised image of womanhood. Fanny's response of "never, never, never! he never will succeed with me" (1994b, p. 351), causes him to question whether she is behaving rationally: "so very determined and positive! This is not like yourself, your rational self" (1994b,

p. 351). Here, Fanny seems to represent "the stilled domestic" and is "refused a public social voice" when she asserts her own independence (R. Batchelor 1994, p. 368). Austen uses Edmund's dialogue and her ironic style to reveal just how little he knows about Fanny's 'rational self'. "I wish he had known you as well as I do ... Between us, I think we should have won you. ... could not have failed ... as I firmly believe ... cannot suppose ... You must have some feeling ... You must be sorry' ... You have moral and literary tastes in common ... I am myself convinced ... I am perfectly persuaded' (1994b, pp. 352–353). Austen allows Fanny to respond in dialogue, although Edmund does not appear to fully understand her meaning, setting out her (and Austen's) beliefs on the subject:

> 'I *should* have thought,' said Fanny, after a pause of recollection and exertion, 'that every woman must have felt the possibility of a man's not being approved, not being loved by someone of her sex, at least, let him be ever so generally agreeable. Let him have all the perfections in the world, I think it ought not to be set down as certain that a man must be acceptable to every woman he may happen to like himself' (1994b, p. 357).

Austen's own comment is reserved until the end of the book after Maria Bertram's (now Mrs Rushworth) affair with Henry and fall from society. Austen demonstrates that Fanny's reading of Crawford's character was correct: "Would he have persevered, and uprightly, Fanny must have been his reward, and a reward very voluntarily bestowed, with a reasonable period from Edmund's marrying Mary. Had he done as he intended, and as he knew he ought ... But" (1994b, p. 473). The subtext Austen indicates is that Henry

Crawford's character would not change, and that Fanny would only marry him in a world where Edmund was already married to someone else. The turning of the paragraph on the word "but" is one employed by Austen in other novels, for example in *Emma* "Nothing was wanting but to be happy when they got there" (1994a, p. 277) when she wishes to highlight the difference between an idealised situation and reality.

Whether we agree with Fanny's moral distaste for acting, or whether her actions may be clouded in part by jealousy is not the point. Austen highlights Fanny's right to hold her own views, to be an independent, rational being, whether she is rejecting acting or the unwanted offer of marriage from Henry Crawford. It could be argued that her heroic actions in the face of those with greater power leads to her winning her reward at the end of the novel, Edmund, the man of *her* choosing. As Ashley Tauchert explains: "The Austen marriage, I contend, offers an (admittedly unfashionable) alternative narrative route to rational female autonomy" (2003, pp. 149–150).

Here we have Austen suggesting that the poor, dependent, daughter of a sailor is the moral superior to those who represent the traditional social hierarchy. In contrast to Beth Tobin's view, this episode therefore could be viewed as a clear criticism of the very structures Austen is claimed to support. The significance of using the word embedding model to draw out this interpretation is key. Yes, it may be possible to identify this specific scene as being related to independence through close reading. However, the word embedding model draws attention to the scene *because* of its relationship to independence - therefore implying that the connection is there, at some level, whether it is immediately identified through close reading or not.

5.1.3 Education, Employment, Wealth and Rank

While the word embedding plot can be used to highlight textual examples to support a close reading, the plots also highlight areas which may not be easy to identify through close reading alone, or sections of the text where the significance in relation to a particular theme may not be automatically recognised. It is important to note, however, that close reading is an essential part of this method in order to make sense of the points raised, and to provide the context necessary to understand the term's association with the chosen theme.

Education and Employment

Figure 5.4 presents a visualisation of the word2vec model of the top 500 terms closest to the vector for 'independent' and 'independence' in Edgeworth's novels. In Figure 5.4, from the top left corner towards the bottom right are terms which are related to education (***enlightened***, ***intellectual***, ***knowledge***, ***talents***) through ***morals*** and ***moralists*** near the centre of the plot, towards a series of negative terms towards the lower right hand side (***sneer***, ***mortify***, ***jilted***, ***ungallant***, ***rashness***). The area towards the top right of contains terms related to employment, from ***manufacturer*** and ***tradesman*** at the centre top of the plot, to ***barristers***, ***litigation*** and ***attorneys*** towards the right.

Edgeworth's emphasis on the importance of education is well known through her work with her father and step-mother, Honora Sneyd, on *Practical Education* published in 1798. "Edgeworth, herself belonging to the Irish Protestant

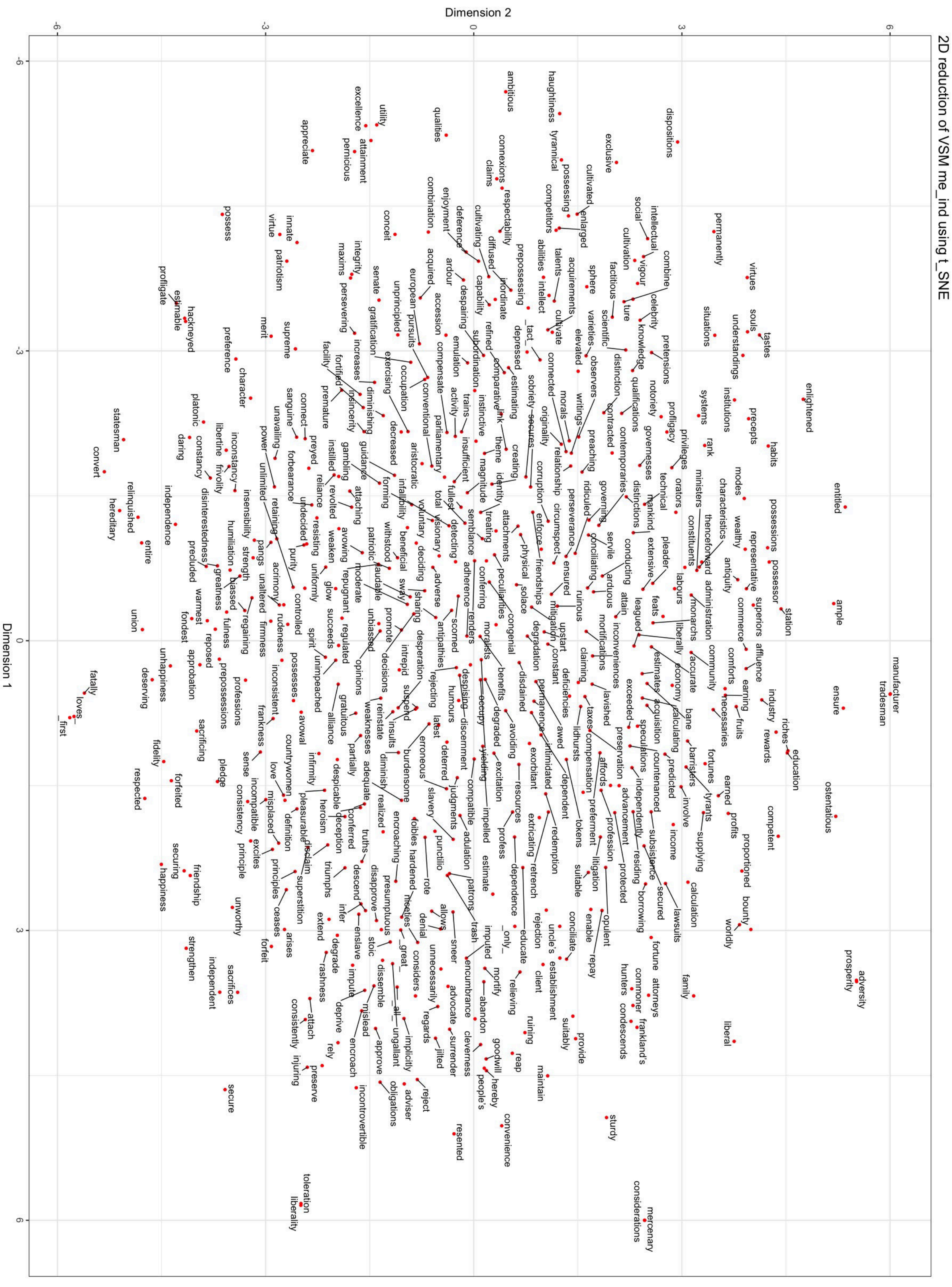


FIGURE 5.4: Word2Vec Model of Edgeworth's Top 500 Terms Related to 'independent' and 'independence'

Ascendancy, was heir to the liberal and rationalist thought of the Enlightenment. This brought her to share many progressive ideas of the time and put her out of sympathy with contemporary conservative discourses, centred on religion, tradition, and 'time-hallowed' institutions"(Lacaita 2001, p. 149). This positions Edgeworth in conflict with many traditional views of education, in particular with regard to the education of women and the lower classes.

Edgeworth's *Patronage* presents contrasting philosophies of male education and the subsequent impact they have on the fortunes of the Percy and Falconer families. The Percy sons are given a broad education enabling them to go into a profession of their choosing: the military, medicine and the law. The Falconers believe firmly in the patronage system and manoeuvre themselves to take advantage of connection and influence where ever possible (***ambitious***, ***connexions***, ***claims***). Eldest son, Buckhurst Falconer, has had an expensive education and his father desires that he goes into the Church, emphasising the prospect of a lucrative living of "nine hundred a year" (1814, p. 80) on the death of the existing incumbent. Buckhurst's entreaty that he is not suited to the Church is met bluntly by his father's insistence that:

> I cannot afford to purchase you a commission, and to maintain you in the army ... And I don't well see what else you could do. You have not application enough for the bar ... You have nothing to expect from my death — I have not a farthing to leave you — my place, you know, is only for life — your mother's fortune is all in annuity, and two girls to be provided for — and to live as we must live — up to and beyond my income — shall have nothing to leave. Though you are my eldest son, you see it is in vain to look

> to my death — so into the church you must go, or be a beggar — and get a living or starve. (1814, pp. 84–86)

This is echoed by Mr Percy who, having outlined the financial arguments already put forward by Buckhurst's father, goes on to suggest that Buckhurst "was bound to exert himself to obtain an independent maintenance in another line of life; that he had talents which would succeed at the bar, if he had application and perseverance sufficient to go through the necessary drudgery at the commencement of the study of the law" (1814, p. 99). The emphasis is clear: either Buckhurst must follow the path intended for him by his father and his education, or he will have to exert considerable effort in order to excel in another profession. Buckhurst is insistent that he will work hard and study for a career in Law, with the hope of marrying Caroline Percy in the future. Edgeworth emphasises the deficiencies in Buckhurst's broader education when, despite assurances to Mr Percy that he will immediately begin his studies, he postpones them for a ball and subsequently follows his father's plan for ordination. Ultimately, Buckhurst's ***mercenary considerations***, his desire for financial independence, are at the expense of his personal independence.

Buckhurst's sense of entitlement and desire for an easy income and a leisurely life is in stark contrast to the sons of the Percy family. The ability and willingness to work hard to adapt, and where necessary make one's own way in the world is shown to be important in gaining lasting independence. When the Percy family lose their estate and fortune through ***litigation*** the positive quality of the sons' previous education is presented:

> From Alfred and Erasmus Mr. Percy had at this trying time the

> satisfaction of receiving at once the kindest and the most manly letters, containing strong expressions of gratitude to their father for having given them such an education as would enable them, notwithstanding the loss of hereditary fortune, to become independent and respectable. What would have been the difference of their fate and of their feelings, had they been suffered to grow up into mere idle lounging gentlemen, or four-in-hand coachmen! In different words, but with the same spirit, both brothers declared that this change in the circumstances of their family did not depress their minds, but, on the contrary, gave them new and powerful motives for exertion (1814, pp. 315–316).

Like Mary Wollstonecraft before her, Edgeworth also addresses several of the popular concerns regarding female education. In *Patronage* the contrasts between the Percy and Falconer women are clear. The Falconer daughters receive the fashionable type of education, with the sole purpose of making an advantageous match and secure an ***establishment***.

The Percy daughters Caroline and Rosamond on the other hand are well educated and well read, with a mother who teaches them ***forbearance*** and ensures that they do not believe themselves above domestic tasks: "It has by some persons been thought, that women who have been suffered to acquire literary tastes, whose understandings have been cultivated and refined, are apt to disdain or to become unfit for the useful minutiae of domestic duties. In the education of her daughters Mrs. Percy had guarded against this danger"

(1814, pp. 324–325).[3] As a result, when the family fortunes change for the worse they are happy to take on domestic tasks and gain greater pleasure from their leisure activities:

> At first they had felt it somewhat irksome, in their change of circumstances, to be forced to spend a considerable portion of their time in preparations for the mere business of living, but they perceived that this constraint gave a new spring to their minds, and a higher relish to their favourite employments. After the domestic business of the day was done, they enjoyed, with fresh delight, the pleasures of which it is not in the power of fortune to deprive us" (M. Edgeworth 1814, pp. 325–326).

For Owenson, education is at the heart of an individual's moral life. She makes it clear that a solid and balanced education prepares one for the ***arduous inconveniences*** and ***deficiencies*** life may present.

Examining the word embedding plots allows the reader to identify possible thematic similarities, or differences, across a corpus of an author or multiple authors' work, aiding the selection of focal points for close reading. The word embedding plot for the 500 words closest to 'independent' and 'independence' for Owenson's novels also includes a number of terms relating to education: ***teaches***, ***instruct***, ***wisdom***, ***philosophers***. Many of her heroines are learned rather than just well educated. In *The Novice of Saint Dominick*, Imogen's education, although almost entirely academic, is suited for

[3]This echoes Vicesimus Knox's challenge to readers that: "Neither is there sufficient reason to conclude, that she whose intellectual acquisitions enable her to entertain her husband, and to form the minds of her children, must be incapable or unwilling to superintend the table, and give a personal attention to domestic oeconomy [sic]" (1990, p. 109).

someone destined to spend their life in a convent. Our first introduction to Imogen is working as an amanuensis for "the pious and learned lady Magdelaine de Montmorell" (1806a, p. 1). Imogen's narrative journey from novice to Countess de St. Dorval to Duchess de Beauvilliers is aided by her education. Owenson concludes the novel saying that "the liberal cultivation of the female mind is the best security for the virtues of the human heart" (1806b, p. 395).

Owenson emphasises both ***education***, and the need to ***instruct***. In *The Wild Irish Girl*, Glorvina has had a broad and varied education, both in the traditional feminine accomplishments and in the broader 'masculine' subjects of history, Latin and business. She is educated to a level where she is able to educate both Horatio, and his father before him, into a broader appreciation of Ireland and Irish culture, reversing the role so commonly seen in Austen's novels. Glorvina is the higher intelligence, she is the one who tries to rein in the financial excesses of her father, something echoed in Austen's *Persuasion* by the character of Anne Eliot. The failure of both Glorvina and Anne is more of a criticism of the men and the society which tells them that women are not worthy of being listened to than the actions of the women themselves.

The three authors present a clear argument for women to receive a broad education to enable them a greater level of independence, echoing Priscilla Wakefield's 1798 *Reflections on the Present Condition of the Female Sex; with Suggestions for its Improvement*

> The necessity of directing the attention of females to some certain occupation is not so apparent, because custom has rendered them dependant upon their fathers and husbands for support; but as

> some of every class experience the loss of those relations, without inheriting an adequate resource, there would be great propriety in preparing each of them, by an education of energy and useful attainments, to meet such disasters, and to be able, under such circumstances, to procure an independence for herself. There is scarcely a more helpless object, in the wide circle of misery which the vicissitudes of civilized society display, than a woman genteelly educated, whether single or married, who is deprived, by any unfortunate accident, of the protection and support of male relations (1990, p. 123).

Wealth and Rank

Figure 5.5, shows terms related to wealth, employment and independence in the Austen corpus which are an expected aspect in the topic of independence in her novels. Within this cluster, there are several sub-clusters: ***landed***, ***property***, ***county***, ***household*** suggests the traditional route to independence through property. The second sub-cluster refers to finance, including words such as ***income***, ***sum***, ***amount***, ***yearly*** and ***expense***. Another two sub-clusters suggests the increasing number of professions gaining independence through employment: ***army***, ***admirals***, ***clergymen***, ***war***, and ***prize***. These terms appear to reinforce Beth Tobin's view of Austen as "an apologist for the landed classes" (Tobin 1990, p. 229). However, although this view may be supported through close readings of the text, as seen in Chapter 1, and as explored in Section 5.1.2 above, this is a far too simplistic view of Austen's work.

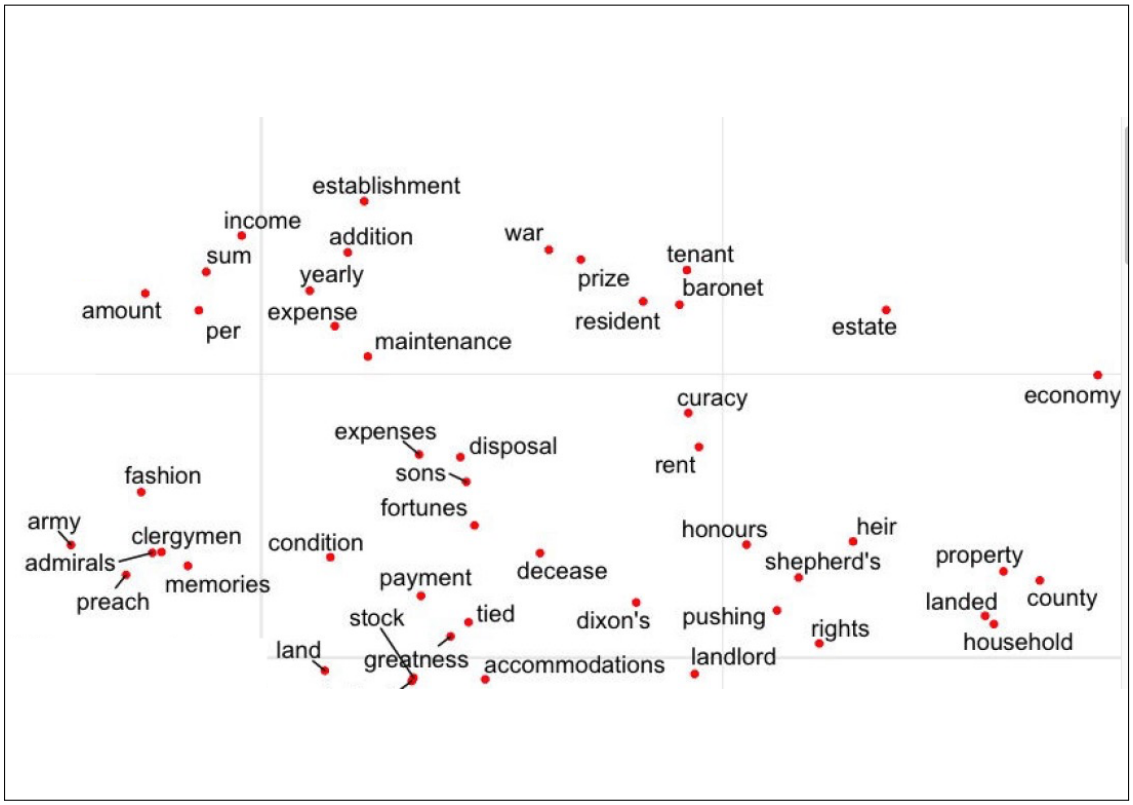


FIGURE 5.5: Cluster of Wealth, Employment and Independence Terms in Word2Vec Model of the Austen Corpus

A term which seems somewhat out of place in a discourse on independence is ***fireside***. We might perhaps expect ***fireside*** to be a term associated with the domestic, female sphere, and in fact we see both ***fireside*** ("make our offerings round the domestic altar . . . the fireside" (Owenson 2008, p. 139)) and chimney ("fancy not that I am one of your chimney-corner, household goddesses" (M. Edgeworth 1806, pp. 171–172)) used in this way. However, as can be seen from the Key Word In Context output in Figure 5.1, this is not the case. In all but two examples, the ***fireside*** is explicitly, or implied to be that of a man. John Knightley in *Emma*, for example, is surprised at Mr Weston's decision "to quit the tranquillity and independence of his own fire-side" after a day on

business in London (Jane Austen 1994a, p. 229). This suggests that the term ***fireside*** has a gendered aspect to it, representing property ownership and the luxury of leisure time afforded to men who can afford it.

Extending our analysis to other related terms, ***fire***, ***grate***, and ***hearth***, suggests that the independence associated with these terms is not merely a gendered one, but also an indicator of rank. In *Mansfield Park*, Fanny's position as a dependent member of the family is highlighted through her lack of a fire, the most simple and essential comfort, "Mrs Norris having stipulated for there never being a fire in it on Fanny's account" (Jane Austen 1994b, p. 154) and Fanny's acute awareness that this is not acceptable for others, "Fanny ... endeavoured to shew herself mistress of the room by her civilities, and looked at the bright bars of her empty grate with concern" (Jane Austen 1994b, p. 171). Once Sir Thomas rectifies this, Fanny's consideration of a fire being a "too great indulgence and luxury" (Jane Austen 1994b, p. 332) for her, highlights the inequality and neglect amongst the wealth of Mansfield Park. For Owenson, the ***hearth*** has both national and class implications. The "earthen hearth" (Owenson 2008, p. 27) and "fine turf fire" (Owenson 1814b, p. 204) emphasise at once the poverty of many of her characters, the wealth of natural resources in Ireland, and the warmth of the people.

Characters of high rank, in the novels of all three authors, are frequently presented as dissipated, out of touch or prejudiced, in some cases both physically and morally diseased (see Figures 5.3, 5.4, and 5.6). Lady Delacour, in *Belinda* is such a character, the physical disease of her breast a metaphor for her lack of ***morals*** and ***inconstancy***. However her independence is destroying

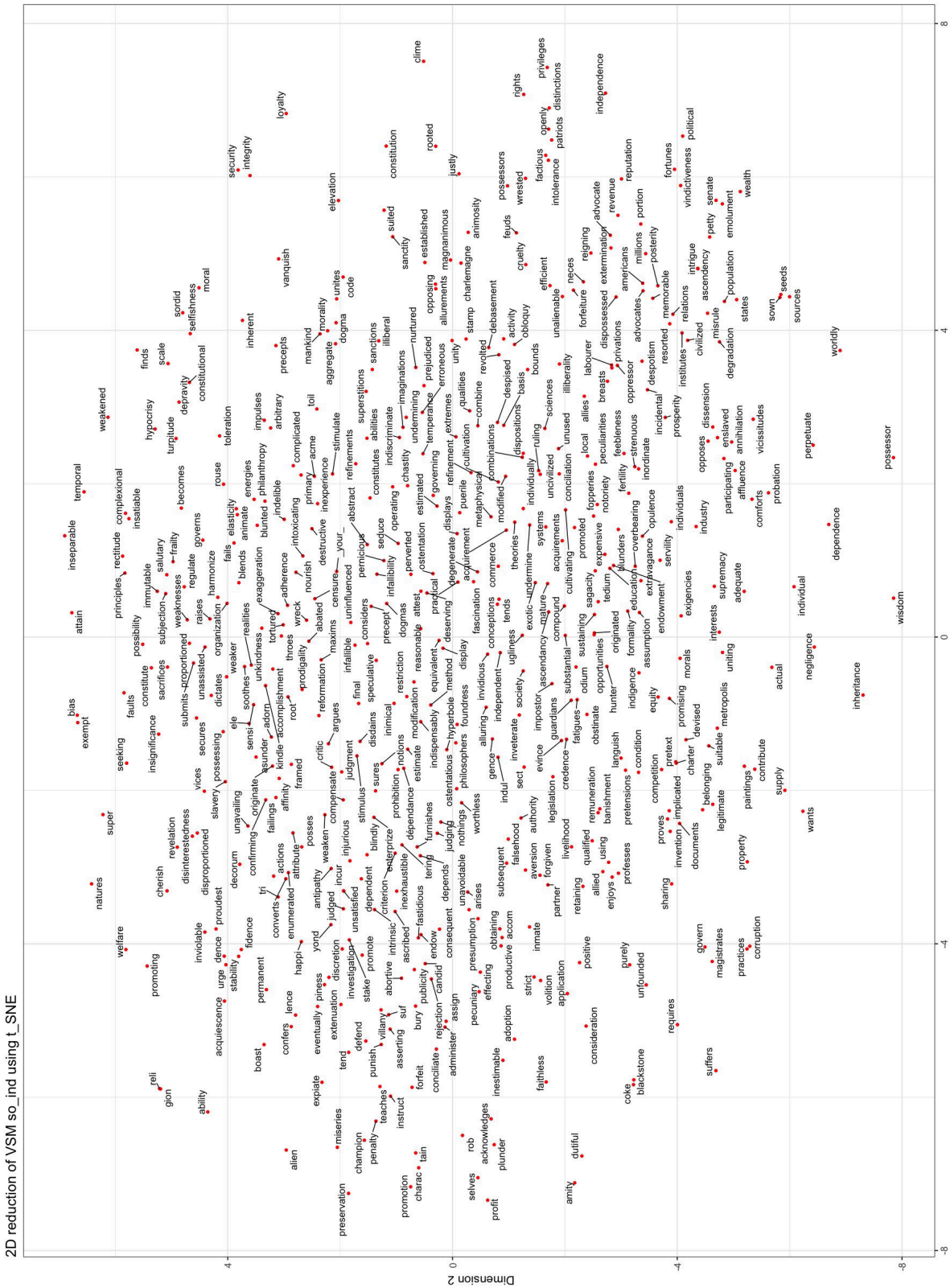


FIGURE 5.6: Word2Vec Model of Owenson's Top 500 Terms Related to 'independent' and 'independence'.

her: "my mind is eaten away like my body, by incurable disease — inveterate remorse — remorse for a life of folly — of folly which has brought me all the punishments of guilt" (M. Edgeworth 2008a, p. 32). Sir Walter Elliot in *Persuasion*, Lady Catherine de Burgh in *Pride and Prejudice*, and Lady Singleton and Lady Llanberis in *O'Donnel* also represent the excesses of the ruling classes, financially, morally and through misplaced pride in their rank. Therefore it is not surprising that each of the word embedding plots for independence contain numerous negative terms (for example as seen in Figure 5.7 and Figure 5.8).

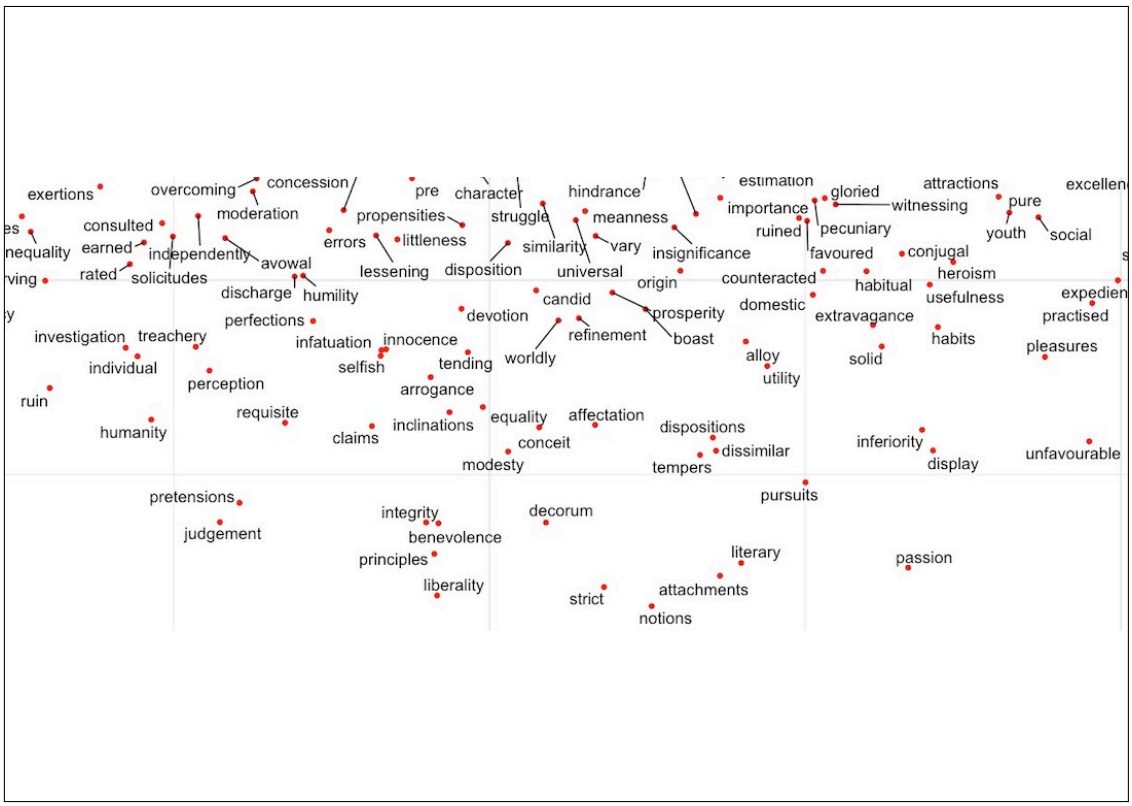


FIGURE 5.7: Cluster of Positive and Negative Terms in Word Embedding Model of the Austen Corpus

Austen, for example demonstrates the vanity of Sir William Lucas:

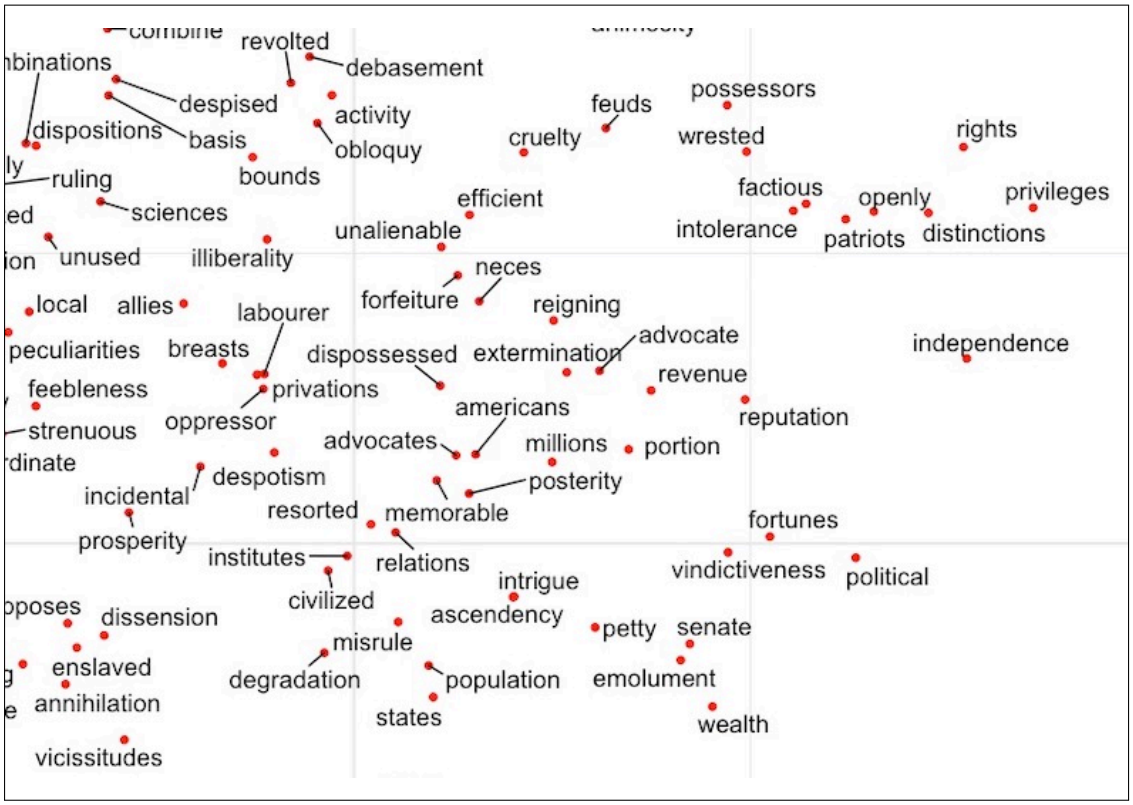


FIGURE 5.8: Cluster of Positive and Negative Terms in Word Embedding Model of the Owenson Corpus

Sir William Lucas had been formerly in trade in Meryton, where he had made a tolerable fortune, and risen to the honour of knighthood by an address to the king during his mayoralty. The distinction had perhaps been felt too strongly. It had given him a disgust to his business, and to his residence in a small market town; and, in quitting them both, he had removed with his family to a house about a mile from Meryton, denominated from that period Lucas Lodge, where he could think with pleasure of his own importance, and, unshackled by business, occupy himself solely in being civil to all the world (Jane Austen 1985, p. 65).

Sir William's rejection of his business, without considering the financial constraints he faces (having a reasonably large family of 'several children' (1985, p. 65) including at least three daughters), indicates his foolishness. He attempts to live the life of an aristocrat without the land and wealth provided by the established estates of Mr Darcy or Lady Catherine De Bourgh. The fact that he cannot provide his eldest daughter Charlotte with sufficient money to attract a suitable husband leaves Charlotte with little option but to pursue and accept the ***unfavourable*** marriage proposal of Mr Collins, who she finds ridiculous, a mercenary ***calculation*** to save herself from future poverty.

5.1.4 Nation and Independence

Austen's novels seem a world away from questions of nationality and conflict. However, as outlined in Chapter 1, during the period 1800–1820 questions of nation and national identity were almost ubiquitous, with the French Revolution and the 1798 Irish rebellion still in recent memory. The Napoleonic and Peninsular wars also took place during these years. Threats were not all external: treasonous plots to kill King George III, the assassination of Prime Minister Spencer Perceval, and attacks on industrial machinery by the Luddites provided a far from peaceful background. Virginia Woolf claimed that:

> The Napoleonic wars did not affect the great majority of those writers at all . . . Jane Austen and Walter Scott. Each lived through the Napoleonic wars; each wrote through them. But, though novelists live very close to the life of their time, neither of them in all

> their novels mentioned the Napoleonic wars. This shows that their model, their vision of human life, was not disturbed or agitated or changed by war (Woolf 1947)

This is far from the truth. The census of 1801 indicates the strength of the armed forces, including the navy, regular troops and militia, as 470,596 men (Great Britain Historical GIS Project 2017a). The following census, in 1811, showed an increase to 640,500, or 10% of the male population (Great Britain Historical GIS Project 2017b). Troop movements, the building of Martello towers as defences, and a steady stream of prisoners of war, would have been hard to ignore. Even harder to ignore would be the casualty figures - more than 300,000 casualties, mostly from wounds and disease (White 2015). With two brothers commanding naval ships during this period, and another brother who had been part of the Oxfordshire militia, Austen would have been fully aware of the challenges and dangers faced (Tomalin 1998).

The word embedding plot for Austen's 500 words nearest to 'independent' and 'independence', seen in Figure 5.3 includes the terms ***sailor***, ***soldier***, ***army***, and ***war***. *Pride and Prejudice* presents a world where the militia are billeted in local towns: "At present, indeed, they were well supplied both with news and happiness by the recent arrival of a militia regiment in the neighbourhood; it was to remain the whole winter, and Meryton was the headquarters" (1985, p. 75). Brighton is a large military camp and staging post for the war in France, and troops are stationed as far afield as Newcastle - the destination for the disgraced and newly married Lydia and Wickham.

In *Mansfield Park*, Austen presents the naval town of Portsmouth in great detail. The Price family, in contrast to the wealthy Bertrams at Mansfield Park, are a naval family: the father is a former Lieutenant in the Marines, William a midshipman looking for promotion, and younger brother Sam is preparing to join the navy. Like Austen's brothers, the Price boys join the navy young, at about the age of twelve. *Persuasion*, one of the two novels published after her death, shares this appreciation of the navy, her main character Anne Elliot eventually marrying a naval officer Captain Wentworth. The novel, however, ends with a harsh reality: "the dread of a future war [was] all that could dim her sunshine. She gloried in being a sailor's wife, but she must pay the tax of quick alarm for belonging to that profession which is, if possible, more distinguished in its domestic virtues than in its national importance" (1994d, p. 254).

Even *Emma*, Austen's most restricted novel in terms of setting, does not lack a military presence: Mr Weston in his youth "had satisfied an active, cheerful mind and social temper by entering into the militia of his county, then embodied" (1994a, p. 12) becoming Captain Weston. Although a necessary plot element, for Captain Weston to meet and marry Miss Churchill leading to the birth of their son Frank, the fact that the militia was embodied suggests a cycle of war and peace which runs beneath the surface of the novel. Austen's novels present a nation at odds with itself, torn between the order and peace of "English verdure, English culture, English comfort" (1994a, p. 272) and the constant reality of a nation at war.

While the horrors of the physical conflicts are never touched on in any detail,

the constant presence of military references and characters across the novels of Austen, but also of Owenson and Edgeworth, present the reality of a country that had been at ***war*** for many years. The Napoleonic Wars had an ongoing impact on the nation, not least through the need to raise taxes: it was estimated that the cost of war between 1793 and 1815 was £831.45 million (Hartwell 1981, p. 141). The decision of the widowed Mrs Dashwood to sell the horses and carriage, and to only take two servants were financial as well as pragmatic as horses for pleasure use and male servants were subject to taxation, something which was far more pressing for those on fixed or reduced means. It may be possible to surmise that this was also the reason that the Price family in *Mansfield Park* kept only female servants, and appear to have only a limited supply of candles, which were also deemed a luxury with strict licensing during this time (Dowell 1888).

In *O'Donnel* Owenson's eponymous hero is presented in a similar fashion to the military characters William Price (*Mansfield Park*) and Captain Wentworth (*Persuasion*), although at somewhat more of a disadvantage as an Irish man. Lacking a fortune to support himself, he is obliged to make his way in the world using his skills. "Again, when his fortunes had fallen with those of an empire ... he had left ... to draw, for the *first time* his sword under the consecrated shadow of his country's banner. In a still more mature period of his life ... he had again gone forth to earn subsistence by his sword" (1814b, p. 295). As a soldier in peacetime he has to consider a move to Europe, part of a larger tradition of 'wild geese' Irish exiles. The impact of taxes on male servants is also evident where O'Donnel is no longer able to maintain McRory as his servant when he moves to England: not only would O'Donnel be subject

to tax he would have to pay a higher rate of taxation as a bachelor.

All three authors criticise the social and political structures of their day in their novels, notably Edgeworth's exploration of "cynicism and corruption in contemporary British life" in *Patronage* (Connolly 2005, p. 66). Edgeworth "was inclined to give national identity a sociocultural foundation, and to construe it as rootedness in place (against absenteeism) and educated spiritual cosmopolitanism (against 'national character', a mark of the uneducated lower classes)" (Lacaita 2001, p. 149). In doing so she emphasised the role of education in the creation of national identity. However, Edgeworth's view was generally that of the ***enlightened*** and ***cultivated*** Anglo-Irish improving the situation of their tenants, as seen, for example, in *The Absentee*.

Owenson too criticises the vagaries of the legal system which increases the wealth of some while doing nothing to protect those in society who were not in a position to look after themselves. However she makes a much more emphatic political point: for her, the governing ranks are not merely foolish, or individually immoral, they are corrupt en masse. In a letter to Alicia LeFanu (18th January 1810) she writes:

> I hear of nothing but politics, and the manner in which things are considered, give me a most thorough contempt for the '*rulers of the earth*;' I am certain that the country, its welfare or prosperity, never, for a moment, make a part in their speculation; it is all *a little miserable system of self-interests*, paltry distinctions, of private pique, and personal ambition. I sometimes with difficulty keep in my indignation. (qtd. in McAuley 2016, p. xi)

Like Austen's Frank Churchill in *Emma* or Edgeworth's Buckhurst Falconer in *Patronage*, Horatio in *The Wild Irish Girl* is a classic example of the idle, independent male, more used to seeking entertainment than making the most of his ***privileges***.

The importance of influencing her reading public, and her success in doing so, is acknowledged by Owenson:

> I said, laughingly, "With a Tory father, a Tory uncle, an aristocratic mother, how comes it that all your young men, bred up in absolution, should be such liberals! who converted them?" She [Margaret, Lady Chapman] smiled, and said, "Why, then, to tell the truth, it was your Ladyship." "I? Why, I have talked so little to them since they grew up." " You have talked enough, and written more than enough to make them what they are!" It is thus we women, the secret tribunal of society, can mine and countermine. (Morgan and Hepworth Dixon 1863b, pp. 377–8)

In *The Wild Irish Girl*, Owenson aims to ***instruct*** both her fictional character Horatio and the reader. Owenson's positioning of her own comments, in an intrusive series of footnotes, allows her to imply a ***moral*** superiority, due to her lack of prejudice against Ireland and the Irish, as well as a greater sense of honesty. Horatio and his father, the two main narrative voices in the novel, are unreliable and deceitful in contrast to Owenson. However, we are gradually shown Horatio's development and, through his first person narrative and his changing viewpoints, feel a sense of understanding and sympathy for him. This was key for Owenson, as if her English readers were alienated

from the narrator their interest, as well as the success of the novel, would be jeopardised, and the opportunity to raise the question of Ireland lost.

There is a marked difference between the romanticised and sentimental view of Ireland presented in *The Wild Irish Girl*, and the more realist, political commentary we find in *Florence Macarthy*. The terms ***corruption***, ***dispossessed***, ***misrule***, ***vindictiveness***, and ***plunder***, which all appear in Owenson's word embedding plot, are most evident in her novel *Florence Macarthy*, a novel that ends with the maxim "IRELAND CAN BE BEST SERVED IN IRELAND" (2012, p. 367). Written in 1818 the novel is "[s]et in the years immediately after the Union, the novel nonetheless addresses its own moment: the famine and fever outbreak of 1817-18 and agrarian unrest in Munster" (Connolly 2011, p. 75). As in *The Wild Irish Girl*, *Florence Macarthy* makes use of extensive footnotes and notes to ***instruct*** the reader in the historical oppression of Ireland and the Irish people. Owenson carefully constructs her comments to leave the reader no doubt as to who the ***oppressor***[s] are and the evil of their actions. For example, she notes that "*The rebels* of those days were chiefly such men as *Lord Barry*, who, sooner than give up their families to massacre, and their property to plunder, set fire to their houses, and took shelter in woods and fastnesses" (Owenson 2012, p. 104).

Owenson's view, as presented in *Florence Macarthy*, is that the entire ruling system of Ireland by the English is corrupt; indeed, she explains that ***corruption*** is essential for success within the system. "The principle of representative government, founded in the positive equality of all men before the law" (Owenson 2012, p. 179) is proved to be a sham when those appointed to uphold the law are themselves corrupt: "Baron Boulter . . . any where *but in Ireland*, might

have been as respectable in his public character as he was pleasant and courteous in his social deportment. But ... like other aspiring barristers ... he submitted to the necessary probation of political corruption, graduated with success, and rose to professional eminence" (p. 191). She presents this ***corruption*** as a "well organized system" (p.179) which "commenced even before the reign of George the First" and which had been "perseveringly directed to overthrow the constitution" (p.179)

The central character of Conway Crawley (a thinly disguised caricature of Owenson's nemesis John Wilson Croker), embodies ***corruption***:

> If ever there was a man formed alike by nature and education to betray the land that gave him birth, and to act openly as the pander of political corruption, or secretly as the agent of defamation, who would stoop to seek his fortune by effecting the fall of a frail woman, or would strive to advance it by stabbing the character of an honest one, who would crush aspiring merit behind the ambuscade of anonymous security, while he came forward openly in the defence of that vileness which rank sanctified and influence protected, that man was Conway Crawley. (p.128)

As a Tory politician, Croker may also have found Owenson's support of Whig politicians and Irish nationalism a threat to his political party and his own emerging political career, but this satirical portrait no doubt continued to fuel his animosity.

In addition to her criticisms of the law and the ruling classes, Owenson goes as far as to suggest that Ireland had been deliberately destabilised in order

to pass the Act of Union: "rebellion, excited for the purpose of effecting a ruinous union, called forth all the worst passions of humanity, and armed petty power with the rod of extermination; placing torture at the disposal of personal vindictiveness, and making falsehood, treachery and corruption, the stepping-stones to power, office, and emolument." (2012, p. 120).

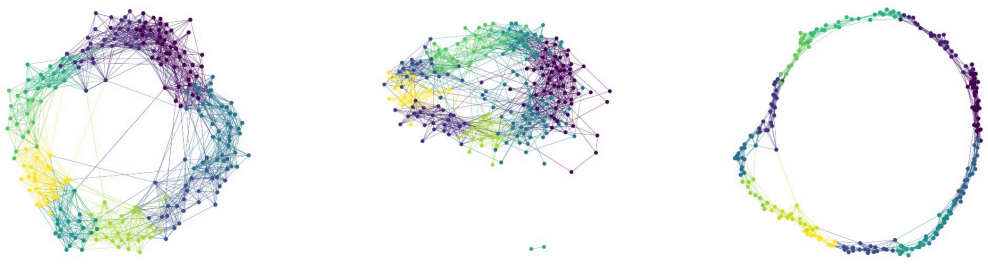
5.2 Semantic Networks

As discussed in Section 3.2.2 above, word embeddings can be used to build semantic networks of related terms, capturing links between words. Each word represented by a node is connected to related terms using lines (or edges), and present a method of exploring the meanings circulating within the corpus. An additional dimension can be added to the network graph by colouring the nodes according to 'community', this can be a manual process with a small network, or an automated method of finding communities of terms through 'community detection' (Heuser 2016a). Community detection, using an algorithm such as `cluster_fast_greedy` allows an additional dimension of the word embedding model to be viewed, and is an unsupervised method, the 'communities' indicate words which appear in similar contexts but are not necessarily linked (Clauset, Newman, and Moore 2004).

For the analysis in this thesis, the `cluster_fast_greedy` algorithm created by Clauset, Newman and Moore was used. The `cluster_fast_greedy` algorithm is a hierarchical agglomeration algorithm which uses a bottom up approach to community detection. The algorithm groups nodes within a network starting with each node in its own community and gradually adding nodes to create a

"*community structure*" which is defined as: "the gathering of vertices [nodes] into groups such that there is a higher density of edges within groups than between them" (Clauset, Newman, and Moore 2004, p. 6). The colouring of nodes according to community allows key concepts within the network to be viewed, as well as terms which link concepts to one another.

Figure 5.9 was created using the `visNetwork` R package (Almende, Thieurmel, and Robert 2018) and represents the relationships between the 250 words nearest to an 'independence/independent' vector and each of the author corpora. The words with a cosine similarity of 0.5 or more are shown. The colours represent communities of words and the original network is an HTML file which allows nodes to be moved or highlighted. This interactivity is an essential part of analysing the graph as it means that the interrelationships between the words (the nodes) can be seen without being obscured by the words themselves. On zooming in, each node is labeled and clicking on the node highlights its links to other nodes (its ego network).



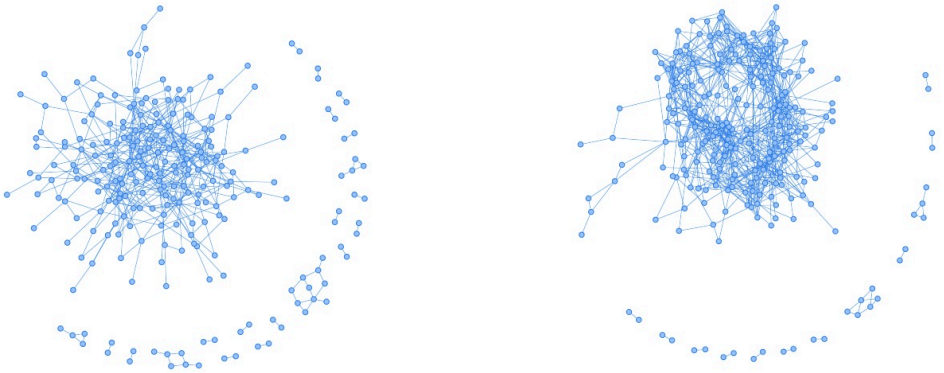
(A) Austen Corpus (B) Edgeworth Corpus (C) Owenson Corpus

FIGURE 5.9: Network Graphs of top 250 'independent/independence' Terms in Austen, Edgeworth, and Owenson Corpora

The three network graphs are all examples of small world networks, where "there are highly clustered neighborhoods and it is possible to move from one node to another in a relatively small number of steps" (Bales and S. B. Johnson 2006, p. 455). The Owenson corpus has very highly clustered communities, 14 in total, each highlighted in a different colour. There are almost no links to other clusters except at the community edges, suggesting that Owenson's discussions of independence are relatively compartmentalised. Austen's corpus produces a network which is less clustered than that of Owenson, with more links between the 9 coloured communities. Austen's communities contain several nodes which have multiple links to other nodes (hubs). This suggests that in Austen's novels, many aspects of independence are interrelated. Edgeworth's network has fewer large hubs and is divided into a large giant component and a smaller satellite group. There are multiple connections between the 9 communities suggesting that Edgeworth's references to independence are less separated than those of the other two authors.

In comparison, the network graphs for the 19th Century and the CLMET (Figure 5.10) corpora could both be described as scale-free, or centralised, networks with small world properties. Both corpora consist of a giant component and a series of small satellites. Unlike the author networks, there are only a few hubs, while most of the nodes have only a few connections. The number of communities is also much larger with 35 for the 19th Century network, and 20 for the *CLMET* network (this is why the networks are a single colour in Figure 5.10).

Examining the three author networks in greater detail reveals that the terms



(A) 19th Century Corpus (B) *CLMET* Corpus

FIGURE 5.10: Network Graphs of top 250 'independent/independence' Terms in 19th Century and *CLMET* Corpora

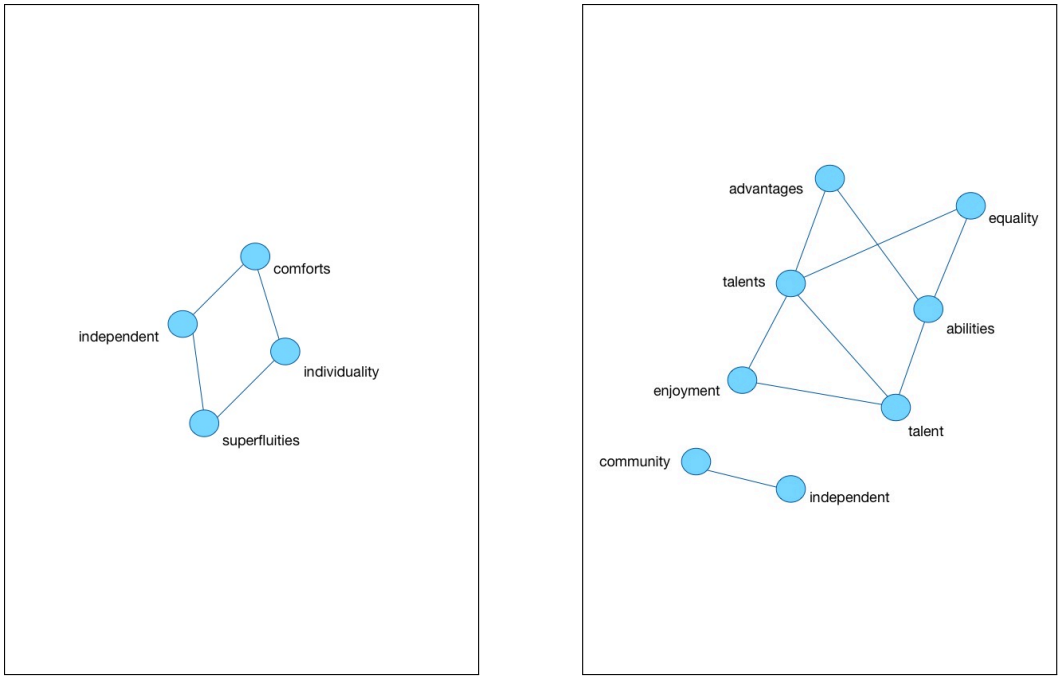
with the highest degree (the number of links to other nodes) differ considerably. For Austen it is 'capacity', for Edgeworth 'subordination', and for Owenson 'regulate'. Edgeworth and Owenson's highest degree terms echo the focus noted in Section 5.1.1 above, with 'subordination' linked to a discussion of rank, and 'regulate' linked closely to the legal and military controls applied to Ireland. Austen's use of the term 'capacity' is interesting as she uses it in relation to employment on a number of occasions in her novels. In *Emma* for example, John Knightley jokes with Emma about the impact pay has on Post Office clerks' abilities to decipher bad handwriting: "the clerks grow expert from habit they must begin with some quickness of sight and hand and exercise improves them if you want any farther explanation,' continued he, smiling

they are paid for it that is the key to a great deal of capacity" (Jane Austen 1994a, p. 224). While in *Northanger Abbey*, Austen bemoans the fate of the author, her own method of gaining some little independence: "there seems almost a general wish of decrying the capacity and undervaluing the labour of the novelist" (Jane Austen 1994c, p. 25). This seems to suggest that ability and intelligence are core to Austen's view of achieving independence, the focus being on the actions of the individual rather than an inherited fortune or title.

Betweenness centrality (which measures the influence of a node as a bridge between different parts of the network based on the number of shortest paths which pass through it) can also reveal the central focus of the author regarding independence. The term with the highest betweenness centrality in Austen is 'respectability', in Owenson 'legitimate'. In the Edgeworth network the term with the highest betweenness centrality is the same as the term with the highest degree — 'subordination': "Mad. de Fleury thought that any education which estranges children entirely from their parents must be fundamentally erroneous; that such a separation must tend to destroy that sense of filial affection and duty, and those principles of domestic subordination, on which so many of the interests, and much of the virtue and happiness, of society depend" (M. Edgeworth 1833, p. 66). For Edgeworth, it seems, at the heart of independence is education and order, both within the family and within society. As was mentioned in Section 5.1.3, Edgeworth indicated that "well-educated individuals should be able to adjust", whether to a change in personal fortunes or revolution (Ó Gallchoir 2005, p. 82).

Finally, focusing on the node for 'independent' in each of the networks is, in itself, quite revealing. Figure 5.11 shows how sparse the connections are in the

19th Century and the *CLMET* networks. 'Independent' links only with 'comforts' and 'superfluities', and 'community' respectively. Edgeworth's network reveals something similar, with only the words 'subordination', 'dependence', 'dependent' and 'advocate' in the ego network for 'independent' (Figure 5.12).



(A) 19th Century Corpus (B) *CLMET* Corpus

FIGURE 5.11: Section of Network Graph Showing 'independent' in the 19th Century and *CLMET* Corpora

In contrast, the 'independent' ego networks for Owenson and Austen have far more connections, each of which reveal their individual preoccupations and suggest a more overt interest in independence. Owenson's 10 terms (Figure 5.13), link to broad political and theoretical concepts: 'political', 'principles', 'regulate', 'independence', 'constitution', 'boast', 'cherish', 'consideration', 'moral', 'mankind', reflecting the political nature of her writing and, perhaps, the security her success afforded her. Austen's 16 terms (Figure 5.14),

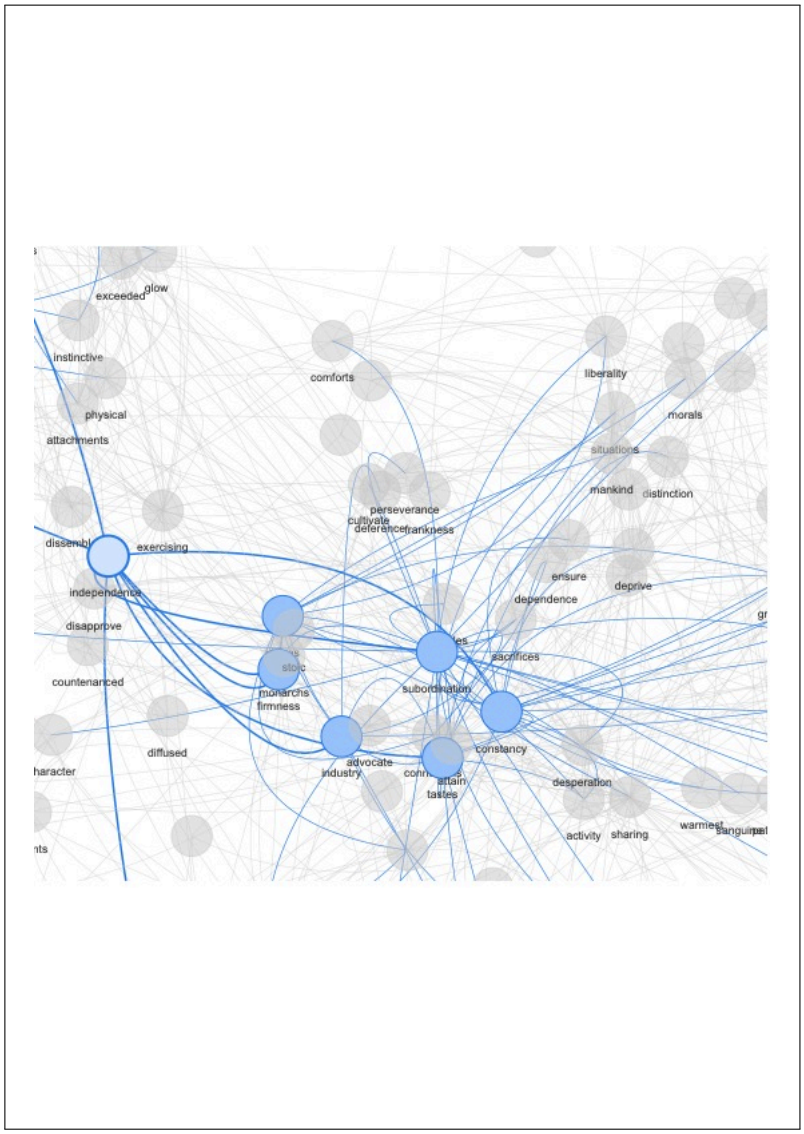


FIGURE 5.12: Ego Network for 'independence' in Edgeworth Corpus

however, reflect a more personal side to independence: 'capacity', 'property', 'profession', 'illiterate', 'value', 'undoubtedly', 'inferior', 'rank', 'fashion', 'principles', 'matrimony', 'abode', 'residence', 'establishment', 'birth', 'respectability'. It may be that Austen's own lack of financial security is reflected in the terms she used.

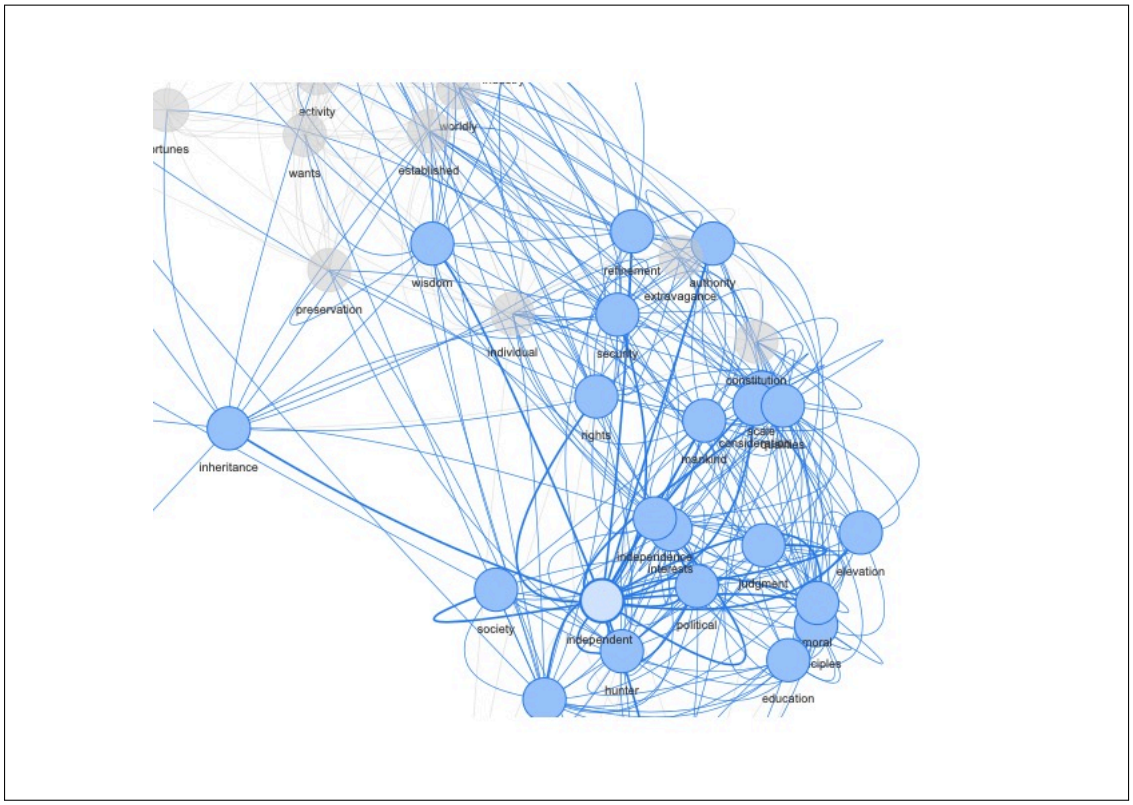


FIGURE 5.13: Ego Network for 'independent' in Owenson Corpus

Semantic networks offer some interesting methods for examining a corpus. However, this process, certainly within literary studies, is in its infancy and therefore this section provides only an indication of what the method is able to achieve.

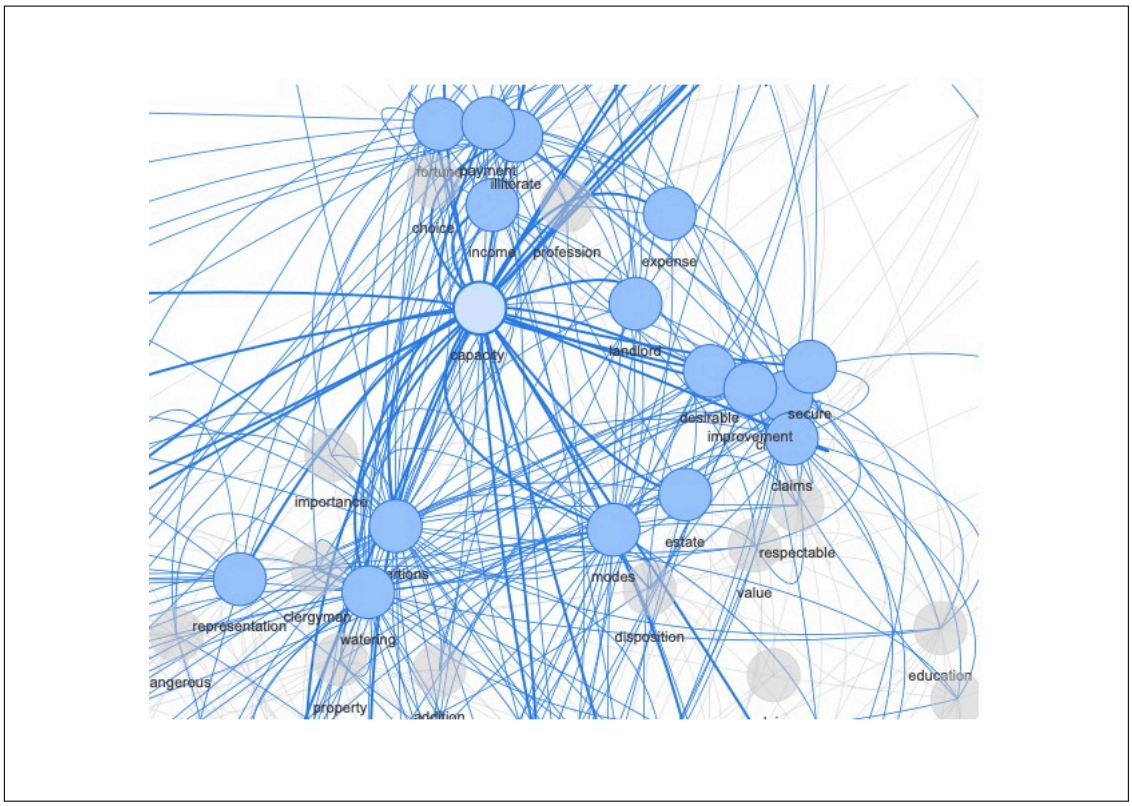


FIGURE 5.14: Ego Network for 'capacity' in Austen Corpus

Conclusion

> I don't care a fig about being popular or unpopular. I am sick of that stuff and intend to be more savagely independent than ever.
>
> Sydney Owenson, *Lady Morgan's Memoirs*

In her speech to the United Nations on 12th July 2013, Malala Yousafzai stated: "The extremists are afraid of books and pens ... The power of the voice of women frightens them" (2013). Austen, Edgeworth and Owenson all lived in a time where women were largely "refused a public social voice" (R. Batchelor 1994, p. 368) leading to them expressing their social and political ideas through the novels they wrote. In some cases, like that of Owenson, we know from their writings that this was their intended aim: "A novel is especially adapted to enable the advocate of any cause to steal upon the public, through the by-ways of the imagination, and to win from its sympathies what its reason so often refuses to yield to undeniable demonstration" (Owenson qtd. in Ferris 2002, p. 13).

All three authors shared the belief in the value of the novel and the work required to write them, as seen above in Section 5.2. For example, in Austen's defence of the novel in *Northanger Abbey*: "there seems almost a general wish of decrying the capacity and undervaluing the labour of the novelist, and of slighting the performances which have only genius, wit, and taste to recommend them" (Jane Austen 1994c, p. 25). Owenson echoes her: "that even to

write a *novel* requires a considerable portion of general information, knowledge, and intelligence, besides talent" (Morgan and Hepworth Dixon 1863a, p. 267). What is abundantly clear is that they were fiercely proud of the work they produced, and had no qualms in emphasising the intellectual capacity needed to become a successful author. It is telling that Austen uses the word 'capacity', with its central position within her semantic network of independence (Figure 5.14), as part of her defence; for Austen, independence was not just about having sufficient means to live on, although this is often at the forefront of her novels. It was about being intellectually independent.

The desire expressed in the novels of Austen, Edgeworth, and Owenson is that of women wanting to be viewed as 'rational creatures', echoing the words used by Mary Wollstonecraft before them. In each woman's work many different aspects of independence were emphasised, and presented, on the surface, in very different ways: from the subtle and cautious Austen to the outspoken Owenson. It is perhaps not surprising that this emergent call for recognition faced increasing criticism, and that Owenson, in particular, was gradually erased from the public consciousness. As Susan Faludi notes: "[a] backlash against woman's rights is nothing new. Indeed, it's a recurring phenomenon: it returns every time women begin to make some headway towards equality, a seemingly inevitable early frost to the brief flowerings of feminism" (1993, p. 66).

As Ina Ferris writes: "the innovative power of the national tale . . . lies precisely in its status as . . . a second-order genre" (2002, p. 11). It may be argued further, that the power of the novels written by Austen, Edgeworth, and Owenson, not just the national tales, had such an impact because the novel

was deemed to be a frivolous genre. What each of the authors has in common is a call to the internal power of intellect rather than the external markers of wealth, rank and accomplishment. Austen's heroines, for example, "are not, especially in the later novels, allowed to get married at all until the heroes have provided convincing evidence of appreciating their qualities of mind, and of accepting their powers of rational judgement, as well as their good hearts" (Kirkham 1997, p. 31). Equally, despite the melodramatic nature of the novel's ending, *The Wild Irish Girl* only allows Horatio to marry Glorvina when he has achieved a true appreciation of her country, and her intellect.

However, although the three authors share an aspiration to intellectual equality, it does not necessarily equate to them valuing each other's work. While Austen's appreciation of Edgeworth's novels is in little doubt, her "judgements of those publishing contemporaneously with herself are complicated by a new element of irony and indirection: that of feigned or exaggerated envy and rivalry ... [her] own sex is exempt neither from her serious literary judgement nor from her outrageous teasing ... The flamboyant *Wild Irish Girl* ... would be worth reading in cold weather if only 'the warmth of her Language could affect the Body" (Grundy 2008, pp. 222–202). Edgeworth, a "mild reformist, a writer [who supported] rational domesticity and the improvement of female education" (Weiss 2007, p. 442), and did not desire a "rational overturning of the social order" (Ross 2014, p. 386), was critical of the overtly political nature of Owenson's novels: "in *Patronage* (1814), she was to caricature Lady Morgan as a governess turned author, and later still to deplore 'a shameful mixture' in Lady Morgan's *Florence Macarthy* (1818) of 'the highest talent and the lowest malevolence and the most despicable disgusting affectation and *impropriety*

. . . Oh that I could prevent people from ever naming me along with her — either for praise or blame"' (Edgeworth qtd. in R. Tracy 1985, p. 7).[4] In this regard, and through her depictions of the native Irish, as an Enlightenment thinker she embodies the tensions inherent in Enlightenment thought, as well as revealing a desire to defend the hierarchical Anglo-Irish position of power against potential threats:

> Just as Enlightenment thinkers' belief in natural equality was in tension with their construction of contemporary hierarchies of cultural and racial differences, so their stress on the social malleability of human nature and the shifting social position of women was in tension with their emphasis on natural differences between the sexes. Depictions of women as a civilising and progressive influence, and indeed women's actual contributions to Enlightenment culture as writers and hostesses of intellectual salons, sat uncomfortably alongside the association of women with nature, the ideas of historical progress as the assertion of masculine reason over female irrationality, and fears of the corrupting influence of ignorant and frivolous women on contemporary society. (Midgley 2007, p. 15)

This study has found that it is possible to consider Austen and Owenson as more closely aligned in their beliefs for a more equal society beyond the scope

[4]This does not appear to have been the view of Edgeworth's father, writing in a letter to Owenson in 1806 he declares" "I think it is a duty, and I am sure it is a pleasure, to contribute as far as it is in my power, to the fame of a writer who has done so much, and so well, for her country" (Morgan and Hepworth Dixon 1863a, p. 294).

of wealth and rank, although not necessarily their politics. As Ina Ferris suggests, Owenson's controversial nature was in part because she "pushed questions of femininity and the nation beyond the prevailing domestic model into speculative and improvisational terrain that confounded standard analogical conflations of the female and national body" (2002, p. 12). While Margaret Kirkham writes of Austen that "Scott had spoken of the author of *Emma* as a woman, in associating the attack upon sensibility with 'these times of revolution', he perhaps reveals a partly conscious awareness that the point of view of this author implies a radical criticism of the society in which she lives" (1997, p. 75).

Although the ideological stance of each author was often quite different, in part due to the privilege (or lack thereof) afforded to them due to their socio-economic positions, all three authors appear to have agreed that "neither birth nor the accoutrements of title and status accurately represented the individual", rather it was the "essential qualities of mind" that truly defined independence (Armstrong 1987, p. 4). While this may not seem particularly radical from a modern perspective, this belief that people, especially women, should be valued for themselves rather than the traditional trappings of wealth and rank was a challenge to the patriarchal hierarchy. The call to challenge the *status quo* and reject what had been traditionally valued was a subversive act that had the potential to call into question the right of those in power to hold that power. What is more, ideas which were expressed in such a palatable way had the potential to succeed where other, more formal, methods would not, as Owenson wrote: "[w]ith respect to my Romances, they found their way into boudoirs and drawing-rooms, where better and sterner Propaganda might

have been rejected" (Morgan 1851, p. 6). That Owenson cites her influence in the notably female spaces of "boudoirs and drawing-rooms" means that it is perhaps not surprising that this emergent call for recognition as a political being in her own right faced such condemnation.

Significance of Research

This study set out to create a theoretical framework combining close and distant reading techniques with a view to their application to small and mid-sized corpora. Enhanced reading (Chapter 2) offers a way to understand the transitions between these techniques, viewing them as part of an extended set of methods with the central aim of developing greater understanding of the texts studied. The practical application of enhanced reading outlined in Chapters 3, 4, and 5 demonstrates the value of this type of methodological approach in literary studies.

The four broad methods used in this thesis each have their strengths and weaknesses. Topic modelling and frequency analysis are most effective for examining a corpus as a whole. They can provide a high level understanding of an unknown corpus, reveal groups of similar texts for further study, or confirm themes or topics suggested by initial close reading. Some of the negatives of topic modelling have already been discussed in Chapter 4, and, overall, topic modelling is less effective for examining small corpora or individual texts as what it may reveal can be easily identified through close reading, and by its nature requires a large amount of text.

Word embedding and semantic network analysis, as shown in this thesis, may be used to explore texts at a range of scales, providing insight that may be difficult to gain from close reading alone. Both methods allow for serendipitous discoveries as, rather than searching for a single word, they allow the exploration of associated terms. In contrast to topic modelling which treats the words in each text as a 'bag of words', word embeddings using the skip-gram method does not. This means that not only do word embeddings help to mitigate some of the concerns raised in Chapters 3 and 4 regarding the removal of terms from the corpus being examined, they also maintain the contextual information from the original text. Word embeddings also allow for word sense disambiguation by enabling 'vector rejection' to discount unwanted word meanings. However, both word embeddings and semantic networks (and topic modelling to a lesser degree) require a level of technical knowledge that may be intimidating to a researcher with little or no past programming experience. The output of a word embedding model requires dimension reduction (in this thesis using t-SNE), to be readable by humans. While there are a growing number of tutorials and materials created with the purpose of bridging this gap, it is time consuming and for some an off-the-shelf solution may be a better option. Having said this, these methods allow far greater flexibility than a packaged product can provide.

The strength of these methods at the mid and small scale is when they are used in combination, as a valuable toolbox for use in conjunction with close reading. It is important to remember that at the heart of corpus analysis are texts, and therefore our interpretation of the output of any computational tool should always refer back to the text. To make sense of the output of

each of these methods requires knowledge of the the original novels. While a KWIC (key word in context) programme can provide a quick insight into the context of the chosen term, close reading is necessary to fully understand how the word is used within the text. This is why these methods work best as part of an iterative process using enhanced reading by moving repeatedly between distant and close reading to gain greater understanding of the corpus.

This research extends our knowledge of applying vector space models to literary texts, and has gone some way to document the efficacy of the word-context model in the analysis of small and mid-sized literary corpora. While a large proportion of digital humanities research has centred upon European or American subjects, the development of methods such as those used here makes this type of research much more egalitarian. As Mary Beard notes: "Women, let's remember, are not the only ones who may feel themselves 'voiceless"' (2017, p. 36). Methods which encourage focus on these 'voiceless' or misinterpreted authors, not as part of a broad sweep of 'literature' or 'culture', but raised to attention in their own right, require a shift in focus.

The size or make-up of a corpora, as discussed above, may be a decision which reduces the diversity of the texts being studied, or limits who may have the necessary tools to analyse it. Insisting on 'clean' or pre-annotated texts can place their study beyond the reach of many scholars who lack the funds to gain access to a previously cleaned collection, or the time to carry out the annotation themselves. The tools chosen to analyse the corpus may also limit those who are able to carry out this type of analysis, proprietary software may be unavailable outside the university, and unaffordable for the individual scholar.

There is a need to consider the impact of the choices we make on the inclusivity and accessibility of our research, as Julia Flanders and Fotis Jannidis point out: "while the terms 'data' and 'modeling' may be new, many of the activities and intellectual frameworks they entail are familiar and deep-rooted", they go on to emphasise that "in a more specific sense, our models represent the shaping choices we make in representing and analysing the materials we study" (Flanders and Jannidis 2018, p. 3). We have a responsibility to ensure that the texts and authors we study help to reflect literature in its broadest sense. To be committed to inclusivity requires us to consider what we are doing to ensure that the methods we use are accessible to all researchers, not just those with large grants or budgets available to them. The enhanced reading techniques here have been shown to work effectively for corpora as small as 6 novels, and applied to text collections where the quality of the OCR is less than perfect. The focus on tools which are free and open source means that enhanced reading is accessible without licences or subscriptions.

However, for methods such as these to flourish the academic system needs to develop and change what it values. A considerable challenge is that currently code, corpora and databases are not part of recognised scholarship in disciplines which still privilege the monograph over everything else. Reluctance to make the products of this unrecognised work freely available is, therefore understandable, and yet the tutorial, software development and blog posts that inform much of this study are invaluable. However, for this to change, and for scholars to start being recognised for this 'background' work, it is necessary to bite the bullet and share: if the work is invisible it is unlikely to ever become recognised.

Recommendations for Further Research

The current study was, by design, limited to the study of a single area of discourse, that of independence, and it is by no means exhaustive. Therefore further in depth studies to apply these techniques to other themes within the focus novels, and to other small to mid-sized corpora are recommended. The impact of an author's lexical variety on topic models (as noted in Chapter 4) should also be reviewed to explore whether a low lexical variety, such as that noted in Austen, has a potentially adverse effect on the topic models produced.

Despite its exploratory nature, Section 5.2 offers some insight into the efficacy of this method when applied to relatively small collections of literary texts. There is scope to examine the use of semantic networks for literary analysis further. Specifically, the inclusion of terms with a negative cosine similarity to the focus term (e.g. of -0.55 to -1) could be considered with a view to exploring antonyms or negatively associated terms in addition to the synonyms and positively associated terms reviewed in this thesis. A combined network of this type may help to provide even greater insight into the ideology reflected in a corpus. An additional area for future study is an exploration into the types of community detection algorithms used for highlighting nodes on a semantic network, and how they might be used to extend our understanding of relationships between terms. One of the limitations for the application of semantic networks is the current limitations for visualising, and thus exploring, the finished results. Even with additional RAM a limit has to be set on the number of nodes and links explored. Future developments in virtual and augmented reality, and the advent of cloud-based workstations enabling far

greater processing power, could provide some exciting developments for the analysis of semantic networks.

A Text Cleaning Script

As all but two of Owenson's novels had to be sourced from Archive.org and rescanned using Abbyy FineReader there are a number of OCR errors which remain - the majority of these were page numbers and random letters used as notes for the bookbinders. This script carries out minimal cleaning by removing individual letters (with the exception of a, i, A and I), and digits. The files were then re-saved.

```r
# To remove random letters and page numbers from Owenson

# Identify folder containing texts
input.dir <- "so_corpus"

# Read the name of all .txt files
files <- dir(input.dir, "\\.txt")

# Read in file - change number in brackets as needed
x <- scan(file = files[6], what = "char",
                  sep = "\n", quote = "", comment.char = "")
# To remove random letters
x <- gsub(" *\\b[^aiAI]{1}\\b *", "", x)

# To remove numbers
x <- gsub("\\d", "", x)

# To remove headers
x <- gsub("DOMINICK", "", x) # Specific capitalised words

# Write cleaned file back to folder
write(x, paste0("clean/", files[2]))
```

FIGURE A.1: Text Cleaning Script

B Nineteenth Century Corpus

Date	Author	Title	Source
1800	Carver	The Old Woman	Chawton
1801	Martin	The Enchantress	Chawton
1801	Opie	The Father and Daughter	Gutenberg
1802	Craik	Stella of the North	Chawton
1803	Porter	Thaddeus of Warsaw	Gutenberg
1804	Hunter	The Unexpected Legacy	Chawton
1804	Opie	Adeline Mowbray	Gutenberg
1806	Dacre	Zofloya	eBooks@Adelaide
1807	Strutt	Drelincourt and Rodalvi	Chawton
1808	Wilkinson	The Child of Mystery	Chawton
1809	Foster	The Corrina of England	Chawton
1809	Mackenzie	The Irish Guardian	Chawton
1810	Green	Romance Readers and Romance Writers	Chawton
1810	Porter	The Scottish Chiefs	UPenn Digital Library
1811	Brunton	Self-Control	Gutenberg
1812	Foster	Substance and Shadow	Chawton
1812	Helme	Magdalen	Chawton
1812	Jacson	Things By Their Right Names	Chawton
1813	Stuart	Cava of Toledo	Chawton
1813	Cooper	The Wife	Chawton
1813	Spence	The Curate and His Daughter	Chawton
1814	Brunton	Discipline	Gutenberg
1814	Burney	The Wanderer	Gutenberg
1814	Hofland	History of a Merchant's Widow	Gutenberg Canada
1816	Hofland	The Affectionate Brothers	Gutenberg Canada
1816	Lamb	Glenarvon	eBooks@Adelaide
1817	Selden	Villasantelle	Chawton
1817	Taylor	Rachel	Chawton
1818	Ferrier	Marriage	Gutenberg
1818	Shelley	Frankenstein	Gutenberg
1819	Harvey	Any Thing But What You Expect	Chawton

Table B.1: 19th Century Female Corpus - Full Title and Source

Date	Author	Title	Source
1800	Brown	Ormond	Gutenberg
1803	Brown	Arthur Mervyn	Gutenberg
1803	Brown	Edgar Huntley	Gutenberg
1804	Brown	Jane Talbot	Gutenberg
1805	Godwin	Fleetwood	Public Library UK
1810	Shelley	Zastrozzi	Gutenberg Australia
1811	Shelley	St. Irvyne	Gutenberg Australia
1813	Barrett	The Heroine	Gutenberg
1814	Scott	Waverley	Gutenberg
1815	Scott	Guy Mannering	Gutenberg
1816	Peacock	Headlong Hall	Gutenberg
1816	Scott	The Antiquary	Gutenberg
1818	Peacock	Nightmare Abbey	Gutenberg
1819	Polidori	The Vampyre	Gutenberg
1820	Maturin	Melmoth the Wanderer	Gutenberg
1820	Scott	The Abbot	Gutenberg

Table B.2: 19th Century Male Corpus - Full Title and Source

C Standard Stop Word List in `tm`

Package

The standard stop word list that is part of the `tm` package is listed below:

"a, about, above, across, after, again, against, all, almost, alone, along, already, also, although, always, am, among, an, and, another, any, anybody, anyone, anything, anywhere, are, area, areas, aren't, around, as, ask, asked, asking, asks, at, away, b, back, backed, backing, backs, be, became, because, become, becomes, been, before, began, behind, being, beings, below, best, better, between, big, both, but, by, c, came, can, cannot, can't, case, cases, certain, certainly, clear, clearly, come, could, couldn't, d, did, didn't, differ, different, differently, do, does, doesn't, doing, done, don't, down, downed, downing, downs, during, e, each, early, either, end, ended, ending, ends, enough, even, evenly, ever, every, everybody, everyone, everything, everywhere, f, face, faces, fact, facts, far, felt, few, find, finds, first, for, four, from, full, fully, further, furthered, furthering, furthers, g, gave, general, generally, get, gets, give, given, gives, go, going, good, goods, got, great, greater, greatest, group, grouped, grouping, groups, h, had, hadn't, has, hasn't, have, haven't, having, he, he'd, he'll, her, here, here's, hers, herself, he's, high, higher, highest, him, himself, his, how, however, how's, i, i'd, if, i'll, i'm, important, in, interest, interested, interesting, interests, into, is, isn't, it, its, it's, itself, i've, j, just, k, keep, keeps, kind, knew, know, known, knows, l, large, largely, last, later, latest,

least, less, let, lets, let's, like, likely, long, longer, longest, m, made, make, making, man, many, may, me, member, members, men, might, more, most, mostly, mr, mrs, much, must, mustn't, my, myself, n, necessary, need, needed, needing, needs, never, new, newer, newest, next, no, nobody, non, noone, nor, not, nothing, now, nowhere, number, numbers, o, of, off, often, old, older, oldest, on, once, one, only, open, opened, opening, opens, or, order, ordered, ordering, orders, other, others, ought, our, ours, ourselves, out, over, own, p, part, parted, parting, parts, per, perhaps, place, places, point, pointed, pointing, points, possible, present, presented, presenting, presents, problem, problems, put, puts, q, quite, r, rather, really, right, room, rooms, s, said, same, saw, say, says, second, seconds, see, seem, seemed, seeming, seems, sees, several, shall, shan't, she, she'd, she'll, she's, should, shouldn't, show, showed, showing, shows, side, sides, since, small, smaller, smallest, so, some, somebody, someone, something, somewhere, state, states, still, such, sure, t, take, taken, than, that, that's, the, their, theirs, them, themselves, then, there, therefore, there's, these, they, they'd, they'll, they're, they've, thing, things, think, thinks, this, those, though, thought, thoughts, three, through, thus, to, today, together, too, took, toward, turn, turned, turning, turns, two, u, under, until, up, upon, us, use, used, uses, v, very, w, want, wanted, wanting, wants, was, wasn't, way, ways, we, we'd, well, we'll, wells, went, were, we're, weren't, we've, what, what's, when, when's, where, where's, whether, which, while, who, whole, whom, who's, whose, why, why's, will, with, within, without, won't, work, worked, working, works, would, wouldn't, x, y, year, years, yes, yet, you, you'd, you'll, young, younger, youngest, your, you're, yours, yourself, yourselves, you've, z "

D w2v_analysis Script

The `w2v_analysis` analyses a chosen term in a vector space model. The function takes 6 arguments:

- vsm — a vector space model
- words — a character vector of focus words
- seed — an integer
- path — the path to the folder you want files saved to
- ref_name — the reference name for the exported files
- num — the number of nearest words you wish to examine

The function will create a vector which is the average of the words input and will output a wordlist of the n nearest words, a csv file of the words and their positions, and a plot of the 2D reduction of the vector space model using the Barnes-Hut implementation of t-SNE. The points for each word are marked in red so the labels can be moved by `ggrepel` for ease of reading. An HTML network graph for the chosen word will also be created `set.seed` is used to ensure replicability.

```r
w2v_analysis2 <- function(vsm, words, seed, path, ref_name, num) {
        # Set the seed
        if (!missing(seed))
                set.seed(seed)
        # Identify the nearest 10 words to the average vector of search terms
        ten <- closest_to(vsm, vsm[[words]])
        # Identify the nearest n words to the average vector of search terms and
        # save as a .txt file
        main <- nearest_to(vsm, vsm[[words]], num)
        wordlist <- names(main)
        filepath <- paste0(path, ref_name)
        write(wordlist, paste0(filepath, ".txt"))
        # Create a subset vector space model
        new_model <- vsm[[wordlist, average = F]]
        # Run Rtsne to reduce new VSM to 2D (Barnes-Hut)
        reduction <- Rtsne(as.matrix(new_model), dims = 2, initial_dims = 50,
                           perplexity = 30, theta = 0.5, check_duplicates = F,
                           pca = F, max_iter = 1000, verbose = F,
                           is_distance = F, Y_init = NULL)
        # Extract Y (positions for plot) as a dataframe and add row names
        df <- as.data.frame(reduction$Y)
        rows <- rownames(new_model)
        rownames(df) <- rows
        # Save dataframe as .csv file
        write.csv(df, paste0(filepath, ".csv"))
        # Create t-SNE plot and save as jpeg
        ggplot(df) +
                geom_point(aes(x = V1, y = V2), color = "red") +
                geom_text_repel(aes(x = V1, y = V2, label = rownames(df),
                                    size = 8)) +
                xlab("Dimension 1") +
                ylab("Dimension 2 ") +
                theme_bw(base_size = 16) +
                theme(legend.position = "none") +
                ggtitle(paste0("2D reduction of VSM ", ref_name, " using t_SNE"))
        ggsave(paste0(ref_name, ".jpeg"), path = path, width = 24,
               height = 18, dpi = 100)
        # Create a network plot of the words
        sim <- cosineSimilarity(new_model, new_model) %>% round(2)
        # convert those below threshold to 0
        sim[sim < max(sim)/2] <- 0
        g <- graph_from_incidence_matrix(sim)
        edges <- get.edgelist(g)
        # Name columns
         colnames(edges) <- c("from", "to")
         g2 <- graph(edges = edges)
         g2 <- simplify(g2) # removes loops
        data <- toVisNetworkData(g2)
         visNetwork(nodes = data$nodes, edges = data$edges, main = "Network of Terms") %>%
                visOptions(highlightNearest = T) %>%
                visSave(paste0(ref_name, ".html"))
        return(ten) # replace with new_list if running as stand alone.
}
```

FIGURE D.1: w2v_analysis Script

E Type-Token Ratio By Text

Date	Title	TTR
1811	Sense and Sensibility	0.5
1813	Pride and Prejudice	0.05
1814	Mansfield Park	0.05
1815	Emma	0.04
1818	Northanger Abbey	0.8
1818	Persuasion	0.7

Table E.1: Austen Corpus - Type-Token Ratio

Date	Title	TTR
1800	Castle Rackrent	0.11
1801	Belinda	0.05
1804	Popular Tales	0.05
1805	The Modern Griselda	0.15
1806	Ennui	0.1
1806	Leonora	0.1
1809	Almeria	0.13
1809	The Dun	0.19
1809	Madame de Fleury	0.16
1809	Manoeuvring	0.09
1812	The Absentee	0.07
1812	Emilie de Coulanges	0.12
1812	Vivian	0.08
1814	Patronage	0.05
1817	Harrington	0.08
1817	Ormond	0.07

TABLE E.2: Edgeworth Corpus - Type-Token Ratio

Date	Title	TTR
1803	St. Clair	0.14
1806	The Novice of Saint Dominick	0.07
1806	The Wild Irish Girl	0.1
1811	The Missionary	0.09
1814	O'Donnel	0.09
1818	Florence Macarthy	0.09

TABLE E.3: Owenson Corpus - Type-Token Ratio

Date	Author	Title	TTR
1800	Carver	The Old Woman	0.08
1801	Martin	The Enchantress	0.13
1801	Opie	The Father and Daughter	0.12
1802	Craik	Stella of the North	0.05
1803	Porter	Thaddeus of Warsaw	0.06
1804	Hunter	The Unexpected Legacy	0.07
1804	Opie	Adeline Mowbray	0.07
1806	Dacre	Zofloya	0.08
1807	Strutt	Drelincourt and Rodalvi	0.08
1808	Wilkinson	The Child of Mystery	0.1
1809	Foster	The Corrina of England	0.11
1809	Mackenzie	The Irish Guardian	0.08
1810	Green	Romance Readers and Romance Writers	0.08
1810	Porter	The Scottish Chiefs	0.04
1811	Brunton	Self-Control	0.06
1812	Foster	Substance and Shadow	0.07
1812	Helme	Magdalen	0.08
1812	Jacson	Things By Their Right Names	0.07
1813	Stuart	Cava of Toledo	0.04
1813	Cooper	The Wife	0.11
1813	Spence	The Curate and His Daughter	0.08
1814	Brunton	Discipline	0.07
1814	Burney	The Wanderer	0.04
1814	Hofland	The History of a Merchant's Widow	0.12
1816	Hofland	The Affectionate Brothers	0.15
1816	Lamb	Glenarvon	0.06
1817	Selden	Villasantelle	0.1
1817	Taylor	Rachel	0.12
1818	Ferrier	Marriage	0.08
1818	Shelley	Frankenstein	0.09
1819	Harvey	Any Thing But What You Expect	0.07

Table E.4: 19th Century Female Corpus - Type-Token Ratio

Date	Author	Title	TTR
1800	Brown	Ormond	0.09
1803	Brown	Arthur Mervyn	0.07
1803	Brown	Edgar Huntley	0.08
1804	Brown	Jane Talbot	0.09
1805	Godwin	Fleetwood	0.07
1810	Shelley	Zastrozzi	0.13
1811	Shelley	St. Irvyne	0.14
1813	Barrett	The Heroine	0.1
1814	Scott	Waverley	0.08
1815	Scott	Guy Mannering	0.08
1816	Peacock	Headlong Hall	0.2
1816	Scott	The Antiquary	0.08
1818	Peacock	Nightmare Abbey	0.18
1819	Polidori	The Vampyre	0.24
1820	Maturin	Melmoth the Wanderer	0.06
1820	Scott	The Abbot	0.07

Table E.5: 19th Century Male Corpus - Type-Token Ratio

F 20 Topic Model for Combined Corpus

The top ten terms for the combined Austen, Edgeworth, Owenson, and 19th Century corpus 20 topic model are shown below.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
1	heart	page	person	door	heart	eyes	father	people	men	man	country	man	man	time	convent	spot	woman	king	character	time
2	mind	thy	stranger	night	eye	tears	mother	friends	arms	gentleman	castle	thing	house	sister	church	distance	company	country	man	friend
3	love	youth	hand	moment	mind	feelings	child	ladies	hands	antiquary	honor	time	money	day	religion	sea	air	castle	life	length
4	soul	thou	situation	bed	thy	words	daughter	things	eyes	horse	day	friend	time	thing	power	light	thing	man	society	situation
5	friend	time	moment	hand	air	hands	son	feelings	feet	country	person	love	day	brother	life	side	conversation	heart	mind	purpose
6	happiness	castle	friend	voice	soul	arms	family	eyes	friends	time	family	mind	master	man	time	night	hand	friend	world	house
7	hand	man	door	time	moment	thoughts	time	days	people	honour	stranger	letter	gentleman	home	light	sun	party	hand	age	attention
8	life	church	name	death	duke	thou	wife	children	mountains	friend	head	moment	honour	subject	day	ground	table	prince	nature	knowledge
9	passion	art	air	head	emotion	lips	husband	men	words	head	land	woman	business	friend	moment	rock	morning	head	taste	person
10	bosom	madam	woman	man	influence	hours	house	words	walls	family	life	opinion	thing	family	priest	stranger	head	count	country	degree

FIGURE F.1: Top Ten Terms for Combined Corpus - 20 Topic Model

G Topic Model for Austen, Edgeworth, Owenson Corpus

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
1	people	father	man	ladyship	lordship	eyes	priest	lord	day	lady	thing	religion	honour	love	gentlemen	heart	stranger	sister	room	madame
2	friends	mother	time	lady	commissioner	feelings	country	con	house	chevalier	and	zeal	man	friend	horses	mind	way	nothing	person	mademoiselle
3	eyes	daughter	manner	room	family	thou	castle	self	time	count	time	faith	house	heart	men	life	road	time	moment	mlle
4	feelings	son	thing	lord	lord	arms	book	castle	money	heart	sir	earth	horse	happiness	guineas	moment	sea	room	castle	monsieur
5	things	child	opinion	dress	son	hands	harp	tion	master	baron	room	look	time	husband	persons	soul	journey	brother	honor	mamma
6	words	ring	day	wit	business	tears	music	ennui	man	duke	man	death	way	mind	years	hand	country	father	country	nun
7	days	jew	life	play	sir	lips	song	life	night	king	hand	feeling	master	woman	hands	nature	town	way	life	countess
8	ladies	heiress	character	fashion	count	emotions	history	pre	boy	minstrel	moment	force	country	passion	things	air	light	home	table	abbé
9	years	imagination	world	world	father	feet	genius	nurse	door	chateau	but	destruction	son	power	sons	thy	mountain	house	countenance	society
10	men	will	nothing	night	law	passions	language	bed	work	friend	morning	light	gentleman	wife	tenants	object	party	day	head	hero
11	others	general	friend	dear	moment	powers	taste	ing	child	novice	way	order	boy	affection	people	emotion	night	something	letter	property
12	children	jewess	family	aunt	country	sufferings	order	pas	home	page	word	tree	lady	letter	orders	influence	boat	family	place	snow
13	spirits	army	fortune	taste	brother	sentiments	conversation	indolence	thing	lord	you	devotion	day	moment	travellers	eye	sun	evening	hand	sum
14	minutes	obstacle	person	part	court	senses	name	ment	hand	convent	that	enthusiasm	head	esteem	parts	bosom	distance	anything	look	opportunity
15	letters	boy	company	lordship	minister	talents	province	mind	way	duchess	night	brow	agent	friendship	strangers	spirit	time	mother	nothing	benefactress
16	circumstances	box	conversation	grace	colonel	rocks	soul	trouble	life	mademoiselle	door	temple	land	conduct	books	tender	chaise	moment	door	month
17	manners	officer	mother	conversation	captain	ruins	window	cave	hour	eye	friend	cast	family	sensibility	guests	character	rock	happiness	world	princess
18	times	apothecary	mind	mamma	majesty	virtues	fact	dis	sir	court	body	multitude	nothing	lover	feet	voice	order	aunt	stranger	sentiment
19	hands	line	course	book	letter	scenes	nurse	ter	business	cheek	voice	man	year	virtue	slaves	day	shore	side	family	gratitude
20	thoughts	childhood	pleasure	admiration	marriage	principles	interest	thorn	mine	tear	woman	error	life	honour	times	love	traveller	place	air	criticism

FIGURE G.1: Top Twenty Terms for Austen, Edgeworth and Owenson Corpus - 20 Topic Model

Bibliography

Adburgham, Alison (1972). *Women in Print*. London: George Allen and Unwin Ltd.

Alexander, Marc (2015). *Keynote - Data Visualisation for the Digital Humanities*. Belfast.

Algee-Hewitt, Mark et al. (2016). *Canon/Archive. Large-scale Dynamics in the Literary Field*.

Allison, Sarah et al. (2011). *Quantitative Formalism: an Experiment*.

Almende, B. V., Benoit Thieurmel, and Titouan Robert (2018). *visNetwork: Network Visualization using 'vis.js' Library*. URL: https://cran.r-project.org/package=visNetwork.

"An Essay on Reading , Addressed to the Ladies" (1810). In: *Belfast Monthly Magazine* 4.23, p. 407.

Anderson, Emily Hodgson (2013). "Why We Do (or Don't) Argue About the Way We Read". In: *The Eighteenth Century* 54.1, pp. 125–128.

Argamon, Shlomo et al. (2003). "Gender, genre, and writing style in formal written texts". In: *Text - Interdisciplinary Journal for the Study of Discourse* 23.3.

Armstrong, Nancy (1987). *Desire and Domestic Fiction: A political history of the novel*. Oxford: Oxford University Press.

"Art. IX. Emma; a Novel. By the Author of Sense and Sensibility, Pride and Prejudice, etc." (1816). In: *The Quarterly Review* 14.Oct 1815 - Jan 1816, pp. 188–201.

Austen, James and Henry Austen (1789). "The Loiterer". In: *The Loiterer* 27.

Austen, Jane (1870). *XIV - "Letter to Edward Austen" 16th December 1816.*

– (1884). *LXXXIV - "Letter to Fanny Knight" 23rd March 1816.* (Visited on 10/02/2017).

– (1985). *Pride and Prejudice*. London: Penguin.

– (1994a). *Emma*. London: Penguin.

– (1994b). *Mansfield Park*. London: Penguin.

– (1994c). *Northanger Abbey*. London: Penguin.

– (1994d). *Persuasion*. London: Penguin.

Axelrad, Arthur M. (1994). ""Of which I avow myself the Authoress...J. Austen": The Jane Austen-Richard Crosby Correspondence". In: *Persuasions: The Jane Austen Journal* 16, pp. 36–38.

Baddeley, Alan D. (1999). *Essentials of Human Memory*. Hove: Psychology Press.

Baker, Paul et al. (2008). "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press." In: *Discourse & Society* 19.3, pp. 273–306.

Bales, Michael E. and Stephen B. Johnson (2006). "Graph theoretic modeling of large-scale semantic networks". In: *Journal of Biomedical Informatics* 39.4, pp. 451–464.

Bank of England (n.d.). *A Brief History of Banknotes*. URL: http://www.bankofengland.co.uk/banknotes/Pages/about/history.aspx.

Baroni, Marco, Georgiana Dinu, and German Kruszewski (2014). "Don't count , predict ! A systematic comparison of context-counting vs . context-predicting semantic vectors". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Pp. 238–247.

Batagelj, Vladimir, Andrej Mrvar, and M Zaveršnik (2002). "Network analysis of texts". In: *University of Ljubljana, Institute of Mathematics, Physics and Mechanics, Department of Mathematics, Preprint series* 40.

Batchelor, Jennie (2010). "Women writers , the popular press and the Literary Fund , 1790 – 1830". In: *Women's Work*. Manchester University Press, pp. 144–188.

Batchelor, Rhonda (1994). "The Rise and Fall of the Eighteenth Century's Authentic Feminine Voice". In: *Eighteenth-Century Fiction* 6.4, pp. 347–368.

Bauerlein, Mark (2005). "Reading at Risk". In: *The Journal of the Midwest Modern Language Assocation* 38.1, pp. 101–106.

Bautz, Annika (2007). "Early Nineteenth-Century Readers of Jane Austen". In: *Literature Compass* 4/5.April 1815, pp. 1412–1427.

Beard, Mary (2017). *Women & Power: A Manifesto*. London: Profile Books Ltd. and London Review of Books.

Belanger, Jacqueline (1998). "Educating the Reading Public : British Critical Reception of Maria Edgeworth ' s Early Irish Writing". In: *Irish University Review* 28.2, pp. 240–255.

Belanger, Jacqueline (2007). *Critical Receptions: Sydney Owenson, Lady Morgan*. Bethesda: Academia Press, LLC.

"Belinda, By Maria Edgeworth." (1802). In: *The Monthly Review; or Literary Journal* 37, pp. 368–374.

Berry, David M (2012). "Introduction". In: *Understanding Digital Humanities*. Ed. by David M Berry. London: Palgrave Macmillan.

Biber, Douglas (2011). "Corpus linguistics and the study of literature: Back to the future?" In: *Scientific Study of Literature* 1.1, pp. 15–23.

Blei, David, Andrew Ng, and Michael Jordan (2003). *Latent dirichlet allocation*.

Bloch, Ruth H. (2003). "Untangling the Roots of Modern Sex Roles: A Survey of Four Centuries of Change". In: *Gender and Morality in Anglo-American Culture, 1650-1800*. University of California Press, pp. 42–54.

Bode, Katherine (2017). "The Equivalence of "Close" and "Distant" Reading; Or, Toward a New Object for Data-Rich Literary History". In: *Modern Language Quarterly* 78.1, pp. 1–26.

Bolukbasi, Tolga et al. (2016). "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings". In: *CoRR* abs/1607.0, pp. 1–25.

Botting, F. (2005). *Gothic*. London: Taylor & Francis.

Bowers, Toni (2009). "The Achievement of Scholarly Authority for Women: Trends in the Interpretation of Eighteenth-Century Fiction". In: *The Eighteenth Century* 50.1, pp. 51–71.

Boyle, Gerry E. and Patrick T. Geary (2003). *The Irish Currency Report of 1804*. Maynooth.

Brewer, David A. (2011). "Counting, Resonance, and Form, A Speculative Manifesto (with Notes)". In: *Eighteenth-Century Fiction* 24.2, pp. 161–170.

Brown, Seth (2014). *The future of statistical analysis is in open-source programming , not domain-specific proprietary software* . URL: http://blogs.lse.ac.uk/impactofsocialsciences/2014/02/24/on-the-future-of-statistical-languages/ (visited on 11/22/2017).

Brunet, Marc-Etienne et al. (2018). "Understanding the Origins of Bias in Word Embeddings". In: *CoRR* abs/1810.0, pp. 1–20.

Burke, Edmund (1951). *Reflections on the French Revolution and Other Essays*. London: J. M. Dent & Sons Ltd.

Burkhard, Remo A (2004). "Learning from Architects : The Difference between Knowledge Visualization and Information Visualization". In: *IIEEE Proceedings of the Eighth International Conference In Information Visualization IV 2004*, pp. 519–524. DOI: 10.1109/IV.2004.1320194.

Burrows, John F. (1986). "Modal verbs and moral principles: an aspect of Jane Austen's style". In: *Literary and Linguistic Computing* 1.1, pp. 9–23.

– (1987). "Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style". In: *Literary and Linguistic Computing* 2.2, pp. 61–70.

– (1989). "'An ocean where each kind\ldots{}': Statistical analysis and some major determinants of literary style". In: *Computers and the Humanities* 23.4–5, pp. 309–321.

– (2002). "'Delta': a measure of stylistic difference and a guide to likely authorship". In: *Literary and Linguistic Computing* 17.3, pp. 267–287.

Burrows, John F. (2003). "Questions of Authorship : Attribution and Beyond : A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001". In: *Computers and the Humanities* 37.1, pp. 5–32.

Busa, Roberto (1980). "The Annals of Humanities Computing : The Index Thomisticus". In: *Computers and the Humanities* 14.2, pp. 83–90.

– (2004). "Foreword: Perspectives on Digital Humanities". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens, and John Unsworth. Blackwell, Oxford.

Butler, Marilyn (1997). *Jane Austen and the War of Ideas*. Oxford: Clarendon Press.

Caplan, Clive (1998). "Jane Austen's Banker Brother: Henry Thomas Austen of Austen & Co., 1801-1816". In: *Persuasions: The Jane Austen Journal* 20, pp. 69–90.

Cartwright, William (2012). "Beck ' s representation of London ' s Underground system : map or diagram ?" In: *Geospatial Science Research 2*.

Celikyilmaz, Asli et al. (2010). *The actor topic model for extracting social networks in literary narrative*.

Chapone, Hester (1990). "Letters on the Improvement of the Mind, addressed to a Young Lady". In: *Women in the Eighteenth Century: Constructions of femininity*. Ed. by Vivien Jones. London: Routledge.

Chen, Hsuanwei Michelle (2017). *An Overview of Information Visualization*. Tech. rep., pp. 5–7.

Cherny, Lynn (2014). *Visualizing Word Embeddings in Pride and Prejudice*. URL: http://blogger.ghostweather.com/2014/11/visualizing-word-embeddings-in-pride.html.

Clauset, A., M. E. Newman, and C. Moore (2004). "Finding Community Structure in Very Large Networks". In: *Physical Review E* 70.6 066111.

Clement, Tanya (2013). *Text Analysis, Data Mining, and Visualizations in Literary Scholarship*.

Colley, Linda (1992). *Britons: Forging the Nation 1707-1837*. New Haven: Yale UP.

Connolly, Claire (2000). "'I accuse Miss Owenson': The Wild Irish Girl as Media Event". In: *Colby Quarterly* 36.2, pp. 98–115.

– (2003). "Introduction: Ireland in Theory". In: *Theorizing Ireland*. Palgrave Macmillian. Chap. 1, pp. 1–3.

– (2005). "'A big book about England'? Public and private meanings in Maria Edgeworth's Patronage". In: *The Irish Novel in the Nineteenth Century: Facts and Fictions*. Ed. by Jacqueline Belanger. Four Courts Press.

– (2011). *A Cultural History of the Irish Novel, 1790-1829*. Cambridge, UK: Cambridge University Press.

Cooperrider, Kensy, Dedre Gentner, and Susan Goldin-Meadow (2016). "Spatial analogies pervade complex relational reasoning: Evidence from spontaneous gestures". In: *Cognitive Research: Principles and Implications* 1.1, p. 28.

"Defence of Mary Wollstonecraft Godwin" (1803). In: *The Anti-Jacobin Review and Magazine, or, Monthly Political and Literary Censor* 15.

DeForest, Mary and Eric Johnson (2001). "The density of Latinate words in the speeches of Jane Austen's characters". In: *Literary and linguistic computing* 16.4, pp. 389–401.

Dickens, Charles (1839). *Oliver Twist (vol.3)*. 2nd. London: Richard Bentley.

Dodd, William (1990). "An Account of the Rise, Progress, and Present State of the Magdalen Hospital, for the reception of Penitent Prostitutes. Together with Dr. Dodd's Sermons. . ., 1770". In: *Women in the Eighteenth Century: Constructions of femininity*. Ed. by Vivien Jones. London: Routledge.

Don, Anthony et al. (2007). "Discovering Interesting Usage Patterns in Text Collections: Integrating Text Mining with Visualization". In: *CIKM '07 Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, pp. 213–221.

Dowell, Stephen (1888). *History of Taxation and Taxes*. 2nd. Vol. II. Longmans.

Downie, James Alan (2006). "Who Says She's a Bourgeois Writer? Reconsidering the Social and Political Contexts of Jane Austen's Novels". In: *Eighteenth-Century Studies* 40.1, pp. 69–84.

– (2010). "Rehabilitating Sir Thomas Bertram". In: *SEL Studies in English Literature 1500-1900* 50.4, pp. 739–758.

Drucker, Johanna (2011). "Humanities Approaches to Graphical Display". In: *Digital Humanities Quarterly* 5.1.

– (2017a). "Information visualization and / as enunciation". In: *Journal of Documentation* 73.5, pp. 903–916.

– (2017b). "Why Distant Reading Isn't". In: *PMLA* 132.3, pp. 628–635.

Du, Yunfei (2009). "Librarians' responses to "reading at risk": A Delphi Study". In: *Library and Information Science Research* 31.1, pp. 46–53.

DuBois, Andrew (2003). "Close Reading: An Introduction". In: *Close Reading: The Reader*. Ed. by Frank Lentricchia and Andrew DuBois. London: Duke University Press, pp. 1–40.

Eagleton, Terry (1995). *Heathcliff and the Great Hunger: Studies in Irish Culture*. London: Verso.

Edgeworth, Maria (1806). *Leonora*. Vol. II. London: J. Johnson and Co.

– (1814). *Patronage*. Vol. I. London: J. Johnson and Co.

– (1815). *Patronage*. 3rd. Vol. I. London: R. Hunter, Baldwin, Chadock, and Joy.

– (1833). "Madame De Fleury". In: *Tales and Novels vol. X*. London: Baldwin and Cradock, pp. 55–142.

– (2008a). *Belinda*. Ed. by Kathryn J. Kirkpatrick. Oxford Wor. Oxford: Oxford University Press.

– (2008b). *Castle Rackrent*. Ed. by George Watson. Oxford Wor. Oxford: Oxford University Press.

Edgeworth, Richard Lovell (1812). "Notice Respecting Miss Edgeworth's Works". In: *Belfast Monthly Magazine* 8.44, p. 220.

Elson, David K., Nicholas Dames, and Kathleen R. McKeown (2010). "Extracting social networks from literary fiction". In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 138–147.

Emsley, Sarah (1999). "Radical Marriage". In: *Eighteenth-Century Fiction* 11.4, pp. 477–498.

Erickson, Lee (2009). "The Romantic-Era Book Trade". In: *Concise Companion to the Romantic Age*. Ed. by Jon Klancher. Wiley, pp. 212–231.

Faludi, Susan (1993). *Backlash*. London: Vintage.

Featherstone, Coral and Etienne Van Der Poel (2017). "Human creativity in the data visualisation pipeline". In: *African Conference On Information Systems & Technology (ACIST)*, pp. 1–11.

Fergus, Jan (2007). *Provincial Readers in Eighteenth-Century England*. Oxford: Oxford University Press.

Ferrer-i-Cancho, Ramon and Brita Elvevag (2010). "Random Texts Do Not Exhibit the Real Zipf's Law-Like Rank Distribution". In: *PLoS ONE* 5.3, pp. 1–10.

Ferris, Ina (1996). "Narrating Cultural Encounter: Lady Morgan and the Irish National Tale". In: *Nineteenth-Century Fiction* 51.3, pp. 287–303.

– (2002). *The Romantic National Tale and the Question of Ireland*. Cambridge: Cambridge University Press.

Fessenbecker, Patrick (2011). "Jane Austen on Love and Pedagogical Power". In: *SEL Studies in English Literature 1500-1900* 51.4, pp. 747–763.

Firth, J. R. (1968). "A Synopsis of Linguistic Theory 1930-1955". In: *Selected Papers of J. R. Firth 1952-1959*. Ed. by F. R. Palmer. London: Longman, pp. 1–32.

Fischer-Starcke, Bettina (2009). "Keywords and Frequent Phrases of Jane Austen's Pride and Prejudice: A corpus-stylistic analysis". In: *International Journal of Corpus Linguistics* 14.4, pp. 492–523.

– (2010). *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*. London: Continuum.

Fish, Stanley (2012). *Mind Your P's and B's: The Digital Humanities and Interpretation*. URL: https://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation.

Flanders, Julia and Fotis Jannidis (2018). "Data modeling in a digital humanities context". In: *The Shape of Data in Digital Humanities*. Ed. by Julia Flanders and Fotis Jannidis. Abingdon, Oxon: Routledge. Chap. 1, pp. 3–25.

"Florence Macarthy by Lady Morgan" (1819). In: *The British Review and London Critical Journal* 13, pp. 482–487.

"France, by Lady Morgan" (1817). In: *The Quarterly Review* 17.April & July, 1817, pp. 260–283.

Friendly, Michael and Daniel J. Denis (2009). *Varieties of Data Visualization*. (Visited on 02/16/2019).

Garside, Peter and Rainer Schöwerling (2000). *The English Novel 1770-1829: A Bibliographical Survey of Prose Fiction Published in the British Isles. Vol. II: 1800-1829*. Oxford: Oxford University Press.

Gay, Penny (1995). "Theatricality and the Theatricals in Mansfield Park". In: *Persuasions: The Jane Austen Journal* 17, pp. 121–129.

Geertz, Clifford (1994). "Thick description: Toward an interpretive theory of culture." In: *Readings in the Philosophy of Social Science*. Ed. by Michael Martin and Lee C. McIntyre. Cambridge, Massachusetts: The MIT Press. Chap. 14, pp. 213–231.

Gemeinböck, Iris (2015). "Containing chaos : compiling a corpus of eighteenth century prose fiction". In: *On-line Proceedings of the Annual Conference of the Poetics and Linguistics Association (PALA)*.

Geng, Li-Ping (2001). "The Loiterer and Jane Austen's Literary Identity". In: *Eighteenth-Century Fiction* 13.4, pp. 579–592.

Gilbert, Sandra M. and Susan Guber (2000). *The Madwoman in the Attic: The woman writer and the nineteenth-century literary imagination*. 2nd. New Haven: Yale UP.

Gilson, David (1982). *A Bibliography of Jane Austen*. Oxford: Clarendon Press.

– (1997). "Henry Austen's 'Memoir of Miss Austen'". In: *Persuasions: The Jane Austen Journal* 19, pp. 12–19.

Gittens, Alex, Dimitris Achlioptas, and Michael W. Mahoney (2017). "Skip-Gram – Zipf + Uniform = Vector Additivity". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Accociation for Computational Linguistics, pp. 69–76.

Goldstone, Andrew (2017). "The Doxa of Reading". In: *PMLA* 132.3, pp. 636–642.

Goldstone, Andrew and Ted Underwood (2014). "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us". In: *New Literary History: A Journal of Theory and Interpretation* 45.3, pp. 359–384.

Goodwyn, Andrew (2012). "One size fits all: The increasing standardisation of English teachers' work in England". In: *Teaching, English* 11.4, pp. 36–53.

Google (2017). *An Open Source Machine Learning Framework for Everyone*. URL: https://www.tensorflow.org (visited on 06/15/2017).

Grainger, Sam, Feng Mao, and Wouter Buytaert (2016). "Environmental data visualisation for non-scientific contexts: Literature review and design framework". In: *Environmental Modelling and Software* 85, pp. 299–318.

Grayson, Siobhán et al. (2016). "Novel2Vec: Characterising 19th Century Fiction via Word Embeddings". In: *24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS'16)*. CEUR-WS, pp. 68–79.

Grayson, S. et al. (2017). "Exploring the Role of Gender in 19th Century Fiction Through the Lens of Word Embeddings Exploring the Role of Gender in 19th Century Fiction Through the Lens of Word Embeddings". In: *International Conference on Language, Data and Knowledge*.

Great Britain Historical GIS Project (2017a). *Census Report 1801*. URL: http://www.visionofbritain.org.uk/census/1801 (visited on 2017).

– (2017b). *Census Report 1811*. URL: http://www.visionofbritain.org.uk/census/GB1811PRE/2 (visited on 10/17/2017).

Gries, Stefan Th. (2009). *Quantitative Corpus Linguistics with R*. London: Routledge.

– (2013). *Statistics for Linguistics with R*. 2nd. Berlin: De Gruyter Mouton.

Grün, Bettina and Kurt Hornik (2011). "topicmodels : An R Package for Fitting Topic Models". In: *Journal of Statistical Software* 40.13, pp. 1–30.

Grundy, Isobel (2008). "Jane Austen and Literary Traditions". In: *The Cambridge Companion to Jane Austen*. Ed. by Edward Copeland and Juliet McMaster. Cambridge, UK: Cambridge University Press. Chap. 11.

Guillory, John (2008). "How Scholars Read". In: *ADE Bulletin* 146.November, pp. 8–17.

Hammond, Adam, Julian Brooke, and Graeme Hirst (2013). "A Tale of Two Cultures: Bringing Literary Analysis and Computational Linguistics Together". In: *2nd Workshop on Computational Linguistics for Literature (CLfL 2013)*, pp. 1–8.

Handler, Richard and Daniel A. Segal (1985). "Hierarchies of Choice: The social construction of rank in Jane Austen". In: *American Ethnologist* 12.4, pp. 691–706.

Hart, Michael S. (2007). *Project Gutenberg Mission Statement*. (Visited on 11/14/2017).

Hartwell, Ronald Max (1981). "Taxation in England During the Industrial Revolution". In: *Cato Journal* 1.1, pp. 129–153.

Haslett, Moyra (2010). "Bluestocking Feminism Revisited: the Satirical Figure of the Bluestocking." In: *Women's writing : the Elizabethan to Victorian period* 17.3, pp. 432–451.

Havre, Susan, Beth Hetzler, and Lucy Nowell (2000). "ThemeRiver : Visualizing Theme Changes over Time". In: *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pp. 115–123.

Hayles, N. Katherine (2010). "How We Read: Close, Hyper, Machine". In: *ADE Bulletin* 150, pp. 62–79.

He, Hua, D Barbosa, and G Kondrak (2013). "Identification of Speakers in Novels." In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Vol. 1 (Long Pa. Association for Computational Linguistics, pp. 1312–1320.

Heller, Deborah (2011). "Bluestocking Studies : the State of the Field – and into the Future". In: *Literature Compass* 8.4, pp. 154–163.

Heuser, Ryan (2016a). *Word Vectors in the Eighteenth Century, Episode 3: From Fields to Vectors*.

– (2016b). *Word vectors in the eighteenth century, episode 4: Semantic networks*. URL: http://ryanheuser.org/word-vectors-4/.

Heuser, Ryan and Long Le-Khac (2011). "Learning to Read Data: Bringing out the Humanistic in the Digital Humanities". In: *Victorian Studies* 54.1, pp. 79–86.

– (2012). "A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method".

Hill, Felix et al. (2014). "Not All Neural Embeddings are Born Equal". In: *CoRR* abs/1410.0, pp. 1–5.

Hill, Jonathan E. (2006). "Minerva at Aberdeen: A. K. Newman and Books in Boards". In: *Romantic Textualities: Literature and Print Culture, 1780-1840* Summer 200.16, pp. 21–39.

Hoaglin, David C. (2003). "John W. Tukey and Data Analysis". In: *Statistical Science* 18.3, pp. 311–318.

Hoover, David L. (2004a). "Delta Prime?" In: *Literary and Linguistic Computing* 19.4, pp. 477–495.

– (2004b). "Testing Burrows's Delta". In: *Literary and Linguistic Computing* 19.4, pp. 453–475.

– (2010). "Teasing out Authorship and Style with t-tests and Zeta". In: *Digital Humanities 2010*, pp. 1–3.

Hopkins, Lisa (1994). "Jane Austen and Money". In: *The Wordsworth Circle* 25.2, pp. 76–78.

Horsmann, Tobias, Nicolai Erbs, and Torsten Zesch (2015). "Fast or Accurate ? – A Comparative Evaluation of PoS Tagging Models". In: *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL-2015)* 2009, pp. 22–30.

Hubbard, E.H. (2012). “Conversation, characterisation and corpus linguistics: Dialogue in Jane Austen’s Sense and Sensibility”. In: *Literator* 23.2, pp. 67–85.

Jänicke, Stefan (2016). *Valuable Research for Visualization and Digital Humanities : A Balancing Act*.

Jänicke, S et al. (2015). “On Close and Distant Reading in Digital Humanities : A Survey and Future Challenges”. In: *Eurographics Conference on Visualization (EuroVis) (2015)*, pp. 1–21.

Jenkins, Joyce (2006). “The Puzzle of Fanny Price”. In: *Philosophy and Literature* 30.2, pp. 346–360.

Jockers, Matthew (2013a). *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press, p. 328.

– (2013b). *"Secret" Recipe for Topic Modelling Themes*. (Visited on 07/13/2018).

– (2014). *Text Analysis with R for Students of Literature*. London: Springer.

Johns, Alessa (2014). “Jane Austen the Stoic : Channeling Elizabeth Carter and the Bluestocking Ethos”. In: *Women’s Writing* 21.4, pp. 444–463.

Johnson, Claudia L. (2012). *Jane Austen’s Cults and Cultures*. Chicago: University of Chicago Press.

Jones, Darryl (2004). *Jane Austen*. Basingstoke: Palgrave Macmillan.

Jones, Vivien (1990). *Women in the Eighteenth Century: Constructions of femininity*. London: Routledge.

Kaplan, Deborah (1994). *Jane Austen Among Women*. JHU Press.

Kelleher, Margaret and James H. Murphy (1997). *Gender Perspectives in Nineteenth Century Ireland: Public and Private Spheres*. Dublin: Irish Academic Press.

Kirkham, Margaret (1997). *Jane Austen, Feminism and Fiction*. London: Athlone Press.

Kirkpatrick, Kathryn J. (2008). "Introduction". In: *The Wild Irish Girl: A national tale*. Oxford: Oxford University Press, pp. vii–xviii.

Kirschenbaum, Matthew (2007). "The Remaking of Reading: Data mining and the digital humanities". In: *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*. Baltimore.

– (2010). "What Is Digital Humanities and What's It Doing in English Departments?" In: *ADE Bulletin* 150, pp. 55–61.

– (2014). "What is "Digital Humanities," and Why Are They Saying Such Terrible Things About It?" In: *differences* 25.1, pp. 46–63.

Klein, Lawrence E (1995). "Gender and the Public/Private Distinction in the Eighteenth Century: Some Questions about Evidence and Analytic Procedure". In: *Eighteenth-Century Studies* 29.1, pp. 97–109.

Knapp, S and W. B. Michaels (1982). "Against Theory". In: *Critical Inquiry* 8.4, pp. 723–742.

Knox, Vicesimus (1990). "'On the Insensibility of the Men to the Charms of a Female Mind Cultivated with Polite and Solid Literature. In a Letter'". In: *Women in the Eighteenth Century: Constructions of femininity*. Ed. by Vivien Jones. London: Routledge.

Knox-Shaw, Peter (2009). *Jane Austen and the Enlightenment*. Cambridge: Cambridge University Press.

Koppel, Moshe, Shlomo Argamon, and Anat Rachel Shimoni (2002). "Automatically Categorizing Written Texts by Author Gender". In: *Literary and Linguistic Computing* 17.4, pp. 401–412.

Krijthe, Jesse H. (2015). *Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation*. URL: https://github.com/jkrijthe/Rtsne.

Lacaita, Francesca (2001). "The Journey of the Encounter: The Politics of the National Tale in Sydney Owenson's Wild Irish Girl and Maria Edgeworth's Ennui". In: *Critical Ireland: New Essays in Literature and Culture*. Ed. by Alan A. Gillis and Aaron Kelly. Dublin: Four Courts Press, pp. 148–154.

Lamont, Claire (2003). "Jane Austen and the Old". In: *The Review of English Studies* 54.217, pp. 661–674.

Lauritsen, John (2007). *The Man Who Wrote Frankenstien*. Pagan Press.

Leech, Geoffrey and M. H. Short (2007). *Style in fiction : a linguistic introduction to English fictional prose*. Longman.

"Letter to Times Newspaper, 6 June, 1815" (1815). In: *Cobbett's Weekly Political Register* 27.23, pp. 705–706.

Lin, Dekang and Patrick Pantel (2001). "DIRT- Discovery of Inference Rules from Text." In: *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2001)*. San Francisco, CA: ACM SIGKDD, pp. 323–328.

Lodge, David (1975). *Changing Places: A Tale of Two Campuses*. London: Secker and Warburg.

Loeber, Rolf, Magda Stouthamer-Loeber, and Joep Leerssen (2004). "Early calls for an Irish national literature". In: *Writing Irishness in 19th Century British Culture*. Ed. by Neil MaCaw. Aldershot: Ashgate. Chap. 2.

Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing High-Dimensional Data Using t-SNE". In: *Journal of Machine Learning Research* 1, pp. 1–48.

Macaulay, Catherine (1990). "Letters on Education". In: *Women in the Eighteenth Century: Constructions of femininity*. Ed. by Vivien Jones. London: Routledge.

Mahlberg, Michaela (2010). "Corpus Linguistics and the Study of Nineteenth-Century Fiction". In: *Journal of Victorian Culture* 15.2, pp. 292–298.

Mandal, Anthony (2015). "Fiction". In: *The Cambridge Companion to Women's Writing in the Romantic Period:* ed. by Devoney Looser. Cambridge: Cambridge University Press. Chap. 2, pp. 16–31.

Manning, Christopher D. (2011). "Part-of-speech tagging from 97% to 100%: Is it time for some linguistics?" In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6608 LNCS.PART 1, pp. 171–189.

Marche, Stephen (2012). *Literature is not Data: Against Digital Humanities*. URL: http://lareviewofbooks.org/essay/literature-is-not-data-against-digital-humanities (visited on 02/17/2015).

Marcus, Mitchell P. et al. (1999). *Treebank-3 LDC99T42*. Philadelphia.

Martin, Graham (1998). *Austen and class*.

Mazzeno, Laurence (2011). *Jane Austen: two centuries of criticism*. Rochester: Camden House.

McAllister, Marie E. (2004). “"Only to Sink Deeper": Venereal Disease in Sense and Sensibility”. In: *Eighteenth-Century Fiction* 17.1, pp. 87–110.

McAuley, Jenny (2016). “Introduction”. In: *Florence Macarthy: An Irish Tale*. Ed. by Jenny McAuley. Chawton Ho. Routledge, pp. vii–xx.

McCallum, Andrew Kachites (2002). *MALLET: A Machine Learning for Language Toolkit*.

McCaw, Neil (2004). *Writing Irishness in 19th Century British Culture*. Aldershot: Ashgate.

McClintock, John and James Strong (1872). *The Cyclopedia of Biblical, Theological, and Ecclesiastical Literature*. Vol. IV - H, I, New York: Harper and Brothers, p. 389.

Meaney, Gerardine, Mary O’Dowd, and Bernadette Whelan (2013). *Reading the Irishwoman: Case Studies in Cultural Enounters and Exchange, 1714-1960*. Liverpool University Press.

Midgley, C (2007). *Feminism and Empire: Women activists in Imperial Britain 1790-1865*. Abingdon, Oxon: Routledge.

Mikolov, Tomas, Greg Corrado, et al. (2013). “Efficient Estimation of Word Representations in Vector Space”. In: *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, pp. 1–12.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (2013). “Linguistic regularities in continuous space word representations”. In: *Proceedings of NAACL-HLT* June, pp. 746–751.

Mimno, David (2013). *mallet: A wrapper around the Java machine learning tool MALLET*.

"Miss Edgeworth's Popular Tales" (1804). In: *The Edinburgh Review v.4* 4, pp. 329–337.

Miss Edgeworth's Works (1809).

Mokyr, Joel and John V C Nye (2007). "Distributional Coalitions , the Industrial Revolution , and the Origins of Economic Growth in Britain". In: *Southern Economic Journal* 74.1, pp. 50–70.

Moretti, Franco (2000). "The Slaughterhouse of Literature". In: *Modern Language Quarterly* 61.1, pp. 207–228.

– (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.

– (2013). *Distant Reading*. London: Verso.

– (2016). *Literature, Measured*.

– (2017). "Franco Moretti : A Response". In: *PMLA* 132.3, pp. 686–689.

Morgan, Sydney (1851). *Letter to Cardinal Wiseman, in Answer to His "Remarks on Lady Morgan's Statements Regarding St. Peter's Chair"*. 3rd. London: Charles Westerton.

Morgan, Sydney and William Hepworth Dixon (1863a). *Lady Morgan's Memoirs: Autobiography, Diaries and Correspondence*. 2nd. Vol. I. London: Wm. H Allen & Co.

– (1863b). *Lady Morgan's Memoirs: Autobiography, Diaries and Correspondence*. 2nd. Vol. II. London: Wm. H Allen & Co.

Morris, Eileen (1982). "Jane Austen's Hidden Message". In: *Persuasions: The Jane Austen Journal* 4, pp. 16–17.

Mosteller, Frederick and David Wallace (1964). *Inference and Disputed Authorship: The Federalist*. Reading, M.A.: Addison-Wesley Publishing Company.

Mueller, Martin (2012). *Scalable Reading*. URL: https://scalablereading.northwestern.edu/scalable-reading/ (visited on 10/20/2014).

Murakami, A. et al. (2017). "'What is this corpus about?' Using topic modeling to explore a specialized corpus". In: *Corpora* 12.2, pp. 243–277.

Murzintcev, Nikita (2016). *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*.

Narain, Mona (1998). "A Prescription of Letters : Maria Edgeworth ' s " Letters for Literary Ladies " and the Ideologies of the Public Sphere". In: *Journal of Narrative Theory* 28.3, pp. 266–286.

Nash, P. W. (2001). "The Abandonment of the Long s in Britain in 1800". In: *Journal of the Printing Historical Society* 3, pp. 3–19.

National Endowment for the Arts (2005). *Reading at Risk: A Survey of Literary Reading in America*. Tech. rep. Washington, D.C.

Nersessian, Anahid (2011). "Empire and Attachment: A Transnational Tale". In: *European Romantic Review* 22.3, pp. 339–346.

Newton, Judith Lowder (1978). "Pride and Prejudice: Power, fantasy, and subversion in Jane Austen". In: *Feminist Studies* 4.1, pp. 27–42.

Nicholas, David and Eti Herman (2009). *Assessing Information Needs in the Age of the Digital Consumer*. 3rd. London: Routledge.

Novels Online (2019). URL: https://chawtonhouse.org/the-library/library-collections/womens-writing-in-english/novels-online/ (visited on 04/14/2019).

Ó Gallchoir, Clíona (2005). "Revolution and Memory in 'Madame De Fleury', 'Emilie De Coulanges' and 'Ennui'". In: *Maria Edgeworth: Women, Enlightenment and Nation*. Dublin: University College Dublin, pp. 77–102.

O'Ciosain, Niall (1997). *Print and Popular Culture in Ireland, 1750-1850*. Basingstoke: Macmillan.

O'Halloran, Kieran (2007). "Corpus-assisted Literary Evaluation". In: *Corpora* 2.1, pp. 33–63.

"On the Female Literature of the Present Age" (1820). In: *The New Monthly Magazine and Universal Register*. Vol. 13. London: Henry Colburn, 271–275 and 633–638.

Owenson, Sydney (1806a). *The Novice of Saint Dominick*. Vol. I. London: Richard Phillips.

– (1806b). *The Novice of Saint Dominick*. Vol. IV. London: Richard Phillips.

– (1814a). *O'Donnel: A National Tale*. Vol. I. London: Henry Colburn.

– (1814b). *O'Donnel: A National Tale*. Vol. III. London: Henry Colburn.

– (2008). *The Wild Irish Girl*. Ed. by Kathryn J. Kirkpatrick. Oxford Wor. Oxford: Oxford University Press.

– (2012). *Florence Macarthy*. Ed. by Jenny McAuley. Abingdon, Oxon: Routledge.

Paley, W. Bradford (2002). *TextArc : Showing Word Frequency and Distribution in Text*.

Paranyushkin, Dmitry (2011). *Identifying the Pathways for Meaning Circulation using Text Network Analysis*. URL: http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis/.

Park, Deokgun et al. (2018). "ConceptVector : Text Visual Analytics via Interactive Lexicon Building using Word Embedding". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1, pp. 361–370.

"Patronage by Miss Edgeworth" (1814). In: *The British Critic and Quarterly Theological Review* 1, pp. 159–173.

Pawlicka, Urszula (2017). "Data, Collaboration, Laboratory : Bringing Concepts from Science into Humanities Practice". In: *English Studies* 98.5, pp. 526–541.

Pearson, Jacqueline (1999). *Women's Reading in Britain, 1750-1835: A Dangerous Recreation*. Cambridge University Press.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1532–1543.

Perone, C. S. (2013). *Machine Learning: Cosine Similarity for Vector Space Models (Part III)*.

Pfeiffer, Peter C. and Laura Garcia-Moreno (1996). *Text and Nation: Cross-Disciplinary Essays on Cultural and National Identities*. Columbia, SC: Camden House.

Piatetsky, Gregory (2017). "New Leader, Trends, and Surprises in Analytics, Data Science, Machine Learning Software Poll". In: URL: https://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html.

Plane, D A, C J Henrie, and M J Perry (2005). "Migration up and down the urban hierarchy and". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.43, pp. 15313–15318.

Poovey, Mary (1984). *The Proper Lady and the Woman Writer*. Chicago: University of Chicago Press.

Prendergast, Amy (2011). "' The drooping genius of our Isle to raise ': the Moira House salon and its role in Gaelic cultural revival". In: *Eighteenth-Century Ireland / Iris an dá chultúr* 26, pp. 95–114.

Relationships (n.d.). (Visited on).

Resnik, Philip and Eric Hardisty (2010). *Gibbs Sampling for the Uninitiated*.

Rhody, Lisa Marie (2017). "Beyond Darwinian Distance : Situating Distant Reading in a Feminist Ut Pictura Poesis Tradition". In: *PMLA* 132.3, pp. 659–667.

Rockwell, Geoffrey and Stéfan Sinclair (2014). *About TAPoR 3.0*. URL: http://tapor.ca/home.

Rohrbach, Emily (2004). "Austen's Later Subjects". In: *SEL Studies in English Literature 1500-1900* 44.4, pp. 737–752.

Römer, Ute (2006). "Where the Computer Meets Language , Literature , and Pedagogy : Corpus Analysis in English Studies Introduction : Corpora and English Studies Corpus Analysis : Some Basic Techniques". In: *How Globalization Affect the Teaching of English: Studying Culture Through Texts*. Ed. by Andrea Gerbig and Anja Muller-Wood. Lampeter: E. Mellen Press, pp. 81–109.

Ross, Slaney Chadwick (2014). “Maria Edgeworth’s The Rose, Thistle, and Shamrock: Symbolic Unification, Women’s Education, and the Marriage Plot”. In: *The Eighteenth Century* 55.4, pp. 377–390.

Runge, Laura (2009). “Momentary Fame: Female Novelists in Eighteenth-Century Book Reviews”. In: *A Companion to the Eighteenth-Century Novel and Culture*. Ed. by Paula R. Backscheider and Catherine Ingrassia. Oxford: Blackwell. Chap. 12, pp. 276–298.

Russell, Gillian (2010). “"A hint of it, with initials": adultery, textuality and publicity in Jane Austen’s Lady Susan.” In: *Women’s writing* 17.3, pp. 469–86.

Ryder, Sean (2005). “Literature in English”. In: *Nineteenth-Century Ireland: A guide to recent research*. Ed. by Laurence Geary and Margaret Kelleher. Dublin: UCD Press, pp. 119–123.

Sabor, Peter (1991). “"Finished up to nature": Walter Scott’s Review of Emma”. In: *Persuasions: The Jane Austen Journal* 13, pp. 88–99.

Sahlgren, Magnus (2006). “The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces”. PhD thesis. Stockholm University, p. 156.

Saklofske, Jon, Estelle Clements, and Richard Cunningham (2012). “On the Digital Future of Humanities”. In: *Digital Humanities Pedagogy: Practices, Principles and Politics*. Ed. by Brett D. Hirsch. OpenBook Publishers. Chap. 13, pp. 311–330.

Salton, G, A Wong, and C S Yang (1975). “A Vector Space Model for Automatic Indexing”. In: *Information Retrieval and Language Processing* 18.11, pp. 613–620.

Schmidt, Ben (2015). *Vector Space Models for the Digital Humanities*. URL: http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html.

Schmidt, Ben and Jian Li (2015). *wordVectors: Tools for creating and analyzing vector-space models of texts*. URL: http://github.com/bmschmidt/wordVectors.

Schnabel, Tobias et al. (2015). "Evaluation methods for unsupervised word embeddings". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Accociation for Computational Linguistics, pp. 298–307.

Schneider, Gerold, Marianne Hundt, and Rahel Oppliger (2016). "Part-of-speech in historical corpora: tagger evaluation and ensemble systems on ARCHER". In: *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)* Konvens, pp. 256–264.

Schofield, Alexandra, Måns Magnusson, and David Mimno (2017). "Pulling Out the Stops: Rethinking Stopword Removal for Topic Models". In: *Eacl 2017*, pp. 432–436.

Schofield, Mary Anne and Cecilia Macheski (1986). *Fetter'd or Free?: British women novelists, 1670-1815*. Ohio and London: Ohio University Press.

Scrivner, Olga and Jefferson Davis (2017). "Interactive Text Mining Suite : Data Visualization for Literary Studies". In: *Proceedings of the Workshop on Corpora in the Digital Humanities (CDH 2017)*. CEUR, pp. 29–38.

Sculley, D. and Bradley M. Pasanek (2008). "Meaning and mining: The impact of implicit assumptions in data mining for the humanities". In: *Literary and Linguistic Computing* 23.4, pp. 409–424.

Seeber, Barbara K. (2000). *General Consent in Jane Austen: A Study of Dialogism*. McGill-Queen's Press - MQUP.

Selivanov, Dmitriy and Qing Wang (2018). *text2vec: Modern Text Mining Framework for R*.

"Sense and Sensibility by A Lady" (1812). In: *The British Critic* 39, p. 527.

Silge, Julia (2017). *janeaustenr: Jane Austen's Complete Novels*.

Silge, Julia and David Robinson (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R". In: *The Journal of Open Source Software* 1.3.

– (2017). *Text Mining with R: A Tidy Approach*. O'Reilly.

Sinclair, Stéfan (2003). "Computer-Assisted Reading: Reconceiving Text Analysis". In: *Literary and Linguistic Computing* 18.2, pp. 175–184.

Sinclair, Stéfan and Geoffrey Rockwell (2007). "Reading Tools, or Text Analysis Tools as Objects of Interpretation". In: *Digital Humanities 2007*. Urbana-Champaign: Graduate School of Library and Information Science, University of Illinois, pp. 199–200.

– (2019). *Voyant Tools*. URL: https://voyant-tools.org (visited on 02/15/2019).

Skelton-Foord, Christopher (1998). "To buy or to borrow ? Circulating libraries and novel reading in Britain , 1778-1828". In: *Library Review* 47.7, pp. 348–354.

Slowikowski, Kamil (2016). *ggrepel: Repulsive Text and Label Geoms for 'ggplot2'*. URL: https://cran.r-project.org/package=ggrepel.

Smith, Barbara Herrnstein (2016). "What Was "Close Reading"?: A Century of Method in Literary Studies". In: *Minnesota Review* 87, pp. 57–75.

Smith, Peter W. H. and W. Aldridge (2011). “Improving Authorship Attribution: Optimizing Burrows’ Delta Method*”. In: *Journal of Quantitative Linguistics* 18.1, pp. 63–88.

Southam, Brian (2004). ““Manoeuvring” in jane austen”. In: *Women’s Writing* 11.3, pp. 463–476.

– (2008). “Jane Austen’s Englishness: Emma as National Tale”. In: *Persuasions: The Jane Austen Journal* 30, pp. 187–201.

Starcke, Bettina (2006). “The phraseology of Jane Austen’s Persuasion: Phraseological units as carriers of meaning”’. In: *ICAME journal* 30, pp. 87–104.

Steinbach, Susie (2012). “Can We still Use ‘Separate Spheres’? British History 25 Years After Family Fortunes”. In: *History Compass* 11, pp. 826–837.

Swinger, Nathaniel et al. (2018). “What are the biases in my word embedding?” In: *CoRR* abs/1812.0.

Tauchert, Ashley (2003). “Mary Wollstonecraft and Jane Austen: ’Rape’ and ’Love’ as (Feminist) Social Realism and Romance”. In: *Women: A Cultural Review* 14.2, pp. 144–158.

Terras, Melissa, Julianne Nyhan, and Edward Vanhoutte, eds. (2013). *Defining Digital Humanities: A Reader*. Farnham: Ashgate Publishing.

Thackeray, William Makepeace (1848). *Vanity Fair*. London: Bradbury and Evans.

The Historical Thesaurus of English (2018). Glasgow.

“The Wild Irish Girl, By Miss Owenson” (1806). In: *The Critical Review; or Annals of Literature* 9, pp. 327–328.

Thursday 17th February 1814 (1814). URL: https://www.britishnewspaperarchive.co.uk/viewer/bl/0000082/18140217/013/0004.

Tobin, Beth Fowkes (1990). "The Moral and Political Economy of Property in Austen's Emma". In: *Eighteenth-Century Fiction* 2.3, pp. 229–254.

Todd, Janet (2005). *Jane Austen in Context*. Cambridge University Press.

Tomalin, Claire (1998). *Jane Austen: A Life*. London: Penguin.

Topliss, Iain (1981). "Mary Wollstonecraft and Maria Edgeworth's Modern Ladies". In: *Etudes Irlandaises* 6, pp. 13–31.

Tracy, Robert (1985). "Maria Edgeworth and Lady Morgan: Legality versus Legitimacy". In: *Nineteenth-Century Fiction* 40.1, pp. 1–22.

Tracy, Thomas (2004). "The Mild Irish Girl: Domesticating the National Tale". In: *Éire-Ireland* 39.1, pp. 81–109.

Trumpener, Katie (1993). "National Character , Nationalist Plots : National Tale and Historical Novel in the Age of Waverley, 1806-1830". In: *ELH* 60.3, pp. 685–731.

– (2009). "Paratext and Genre System : A Response to Franco Moretti". In: *Critical Inquiry* 36.1, pp. 159–171.

Turney, Peter D. and Patrick Pantel (2010). "From frequency to meaning: Vector space models of semantics". In: *Journal of Artificial Intelligence Research* 37, pp. 141–188.

Underwood, Ted (2012). *What kinds of "topics" does topic modeling actually produce?*

– (2018). "Algorithmic modeling". In: *The Shape of Data in Digital Humanities*. Ed. by Julia Flanders and Fotis Jannidis. Abingdon, Oxon: Routledge. Chap. 13, pp. 250–263.

University of Glasgow (2015). *The Historical Thesaurus of English, version 4.2*. (Visited on 08/18/2015).

Urbanek, Simon (2018). *rJava: Low-Level R to Java Interface*. URL: https://cran.r-project.org/package=rJava.

Van Acker, Wouter and Pieter Uyttenhove (2012). "Analogous Spaces: An Introduction to Spatial Metaphors for the Organization of Knowledge". In: *Library Trends* 61.2, pp. 259–270.

Van Der Maaten, Laurens (2014). "Accelerating t-sne using tree-based algorithms". In: *Journal of Machine Learning Research* 15.1, pp. 3221–3245.

Velleman, P. F. and David C. Hoaglin (2004). *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: The Internet-First University Press.

Venables, W. N., D. M. Smith, and R Core Team (2017). *An Introduction to R*. R Core Team. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf.

Vickery, Amanda (1993). "Golden Age to Separate Spheres? A Review of the Categories and Chronology of English Women' s History". In: *The Historical Journal* 36.2, pp. 383–414.

Vogrinčič, Ana (2008). "The Novel-Reading Panic in 18th-Century in England: An Outline of an Early Moral Media Panic." In: *Medijska Istrazivanja/Media Research* 31.410, pp. 103–124.

Wakefield, Priscilla (1990). "Reflections on the Present Condition of the Female Sex; with Suggestions for its Improvement, 1798". In: *Women in the Eighteenth Century: Constructions of femininity*, pp. 121–130.

Watson, Nicola J. (1994). *Revolution and the Form of the British Novel 1790-1825*. Oxford: Clarendon Press.

Wayner, Peter (2017). "Python vs. R: The battle for data scientist mind share". In: *InfoWorld*.

Weiss, Deborah (2007). "The Extraordinary Ordinary Belinda: Maria Edgeworth's Female Philosopher". In: *Eighteenth Century Fiction* 19.4, pp. 441–461.

White, Matthew (2015). *Necrometrics*. (Visited on 10/17/2017).

Whitfield, Henry (1804). *A Picture From Life: or, The History of Emma Tankerville and Sir Henry Moreton*. London: S Highley.

Wickham, Hadley (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer. URL: http://ggplot2.org.

Wijitsopon, Raksangob (2013). "A Corpus-Based Study of the Style in Jane Austen's Novels". In: *MANUSYA: Journal of Humanities Regular* 16.1, pp. 41–64.

Wilkens, Matthew (2000). "Canons, Close Reading, and the Evolution of Method". In: *Debates in the Digital Humanities*. Ed. by Matthew K. Gold. University of Minnesota Press. Chap. 14, pp. 249–258.

Wiltshire, John (1997). "Mansfield Park, Emma, Persuasion". In: *The Cambridge Companion to Jane Austen*. Ed. by Edward Copeland and Juliet McMaster. Cambridge, UK: Cambridge University Press. Chap. 4, pp. 58–83.

Winter, Thomas Nelson (1973). "Catullus Purified: A Brief History of Carmen 16". In: *Arethusa* 6.2, pp. 257–265.

Wohlgemut, Esther (1999). "Maria Edgeworth and the Question of National Identity". In: *SEL Studies in English Literature 1500-1900* 39.4, pp. 645–658.

Wolf, Amy (2004). “Epistolarity, Narrative, and the Fallen Woman in Mansfield Park”. In: *Eighteenth-Century Fiction* 16.2, pp. 265–285.

Wollstonecraft, Mary (2008). *A Vindication of the Rights of Woman and A Vindication of the Rights of Man*. Oxford: Oxford University Press.

“Woman: or , Ida of Athens” (1809). In: *The Quarterly Review* 1, pp. 50–52.

“Woman : Or Ida of Athens (Continued) by Miss Owenson” (1809). In: *Belfast Monthly Magazine* 2.8, pp. 221–224.

Woolf, Virginia (1947). “The Leaning Tower”. In: *The Moment and Other Essays*. Chap. 18.

Wright, Nicole M. (2017). “Alt-Right Jane Austen”. In: *The Chronicle of Higher Education*.

Wyett, Jodi L (2015). “Female Quixotism Refashioned : Northanger Abbey , the Engaged Reader , and the Woman Writer”. In: *The Eighteenth Century* 56.2, pp. 261–276.

Yildirim, Ilker (2012). “Bayesian Inference : Gibbs Sampling”. In: 14627, pp. 1–6.

Yousafzai, Malala (2013). *Malala’s Speech (12th July 2013)*. New York. URL: http://www.un.org/News/dh/infocus/malala%7B%5C_%7Dspeach.pdf.

Zeender, Marie-Noelle (2008). “Resistance in The Wild Irish Girl”. In: *Cycnos* 19.1.