

“Replete with instruction and rational amusement”?: Unexpected Features in the Register of British Didactic Novels, 1778–1814

Juliette Misset

University of Strasbourg, France

jmisset@unistra.fr

 0000-0002-0036-8919

Abstract

British didactic novels of the turn of the 19th century have been defined as works of fiction where instruction in moral codes of behavior rather than imaginative elements is the primary focus (Havens 2017, p. 5). My research aims to investigate the generic specificities of such novels by working with the open-source software TXM and AntConc to compare two corpora of novels published between 1778 and 1814 in Britain. These corpora were created using reviews from the *Monthly Review* and the *Critical Review*. Contrary to my hypothesis, a lexical comparison of the two corpora shows that the novels they contain do not materially differ in their use of lexis related to instruction and morality. This leads me to reassess the basis for the early reception of didacticism in these novels. Fruitful new hypotheses are generated using both corpus stylistics and close reading.

1 Introduction

The quote in my title, which comes from an early review of Frances Burney’s second novel, *Cecilia* (1782) found in the *Critical Review*, matches recent definitions of the didactic novel as a prevalent subgenre of fiction at the turn of the 19th century. According to Havens (2017), “while didactic novels were allowed imaginative elements, instruction had to remain the primary focus” of the narrative, and this “perpetuated strict moral codes” (pp. 5, 8). Similarly, Wood (2003) argues that the two decades immediately following the French Revolution were “perhaps the most tolerant of overt didacticism in the history of British fiction” and this was in part because of the fear of revolution; as such, the overt didacticism in these novels “coexists with or subsumes aestheticism” (p. 12). While Wood focuses her study on anti-revolutionary female novelists, she states that writers across the political spectrum wrote didactic fiction, which

was “constructed to avoid ambiguity, and to discourage personal and idiosyncratic exegesis” (pp. 64–65). These novels rely on plot “to inculcate particular morals” (p. 68), and the language they use, for example, in their embedded statements and value judgments “implicitly support[s] the text’s moral basis” and “indicate[s] the appropriate readerly response” (p. 66). These critics define turn-of-the-19th-century British didactic novels as works with a straightforward moral message that was delivered through the medium of narrative fiction; here instruction came first and amusement second. This is opposed to fiction that foregrounds the kinds of “narratorial ‘indirection’” that literary critics have increasingly come to value since the 19th century (Wood 2003, p. 16).

Both Havens and Wood consider didactic novels to have been of cultural—although not necessarily artistic—importance in Britain in the decades surrounding the French Revolution. Nevertheless, while adjacent categories of fiction of the period such as sentimental or domestic novels have been studied at length, the formal specificities of the didactic novel as a subgenre in this narrative landscape are yet to receive much critical attention (Havens 2017, p. 13). This paper aims to investigate elements of what may be called the constitutive register of didactic novels based on Biber’s definitions of this notion (2019, p. 16). I use a combination of computer-aided textual analysis and close reading to compare two corpora of novels published between 1778 and 1814 in Britain. As we will see, the former approach yielded very different results from the ones I expected, which forced an abrupt shift in my perspective on this research. I complement the original corpus-based methodology with a corpus-driven one, which leads to stimulating new approaches to the textual basis for these novels’ reception.

2 The Corpora

Contemporary reviews of novels of this period found in the *Monthly Review* and the *Critical Review* were used to create the two corpora.¹ This allowed for a systemic approach to the creation of the corpora since both these *Reviews* professed to address all new publications in their book reviews rather than “select[ing] drastically” as the *Edinburgh Review* did at the beginning of the 19th century (Butler 1993, p. 131; Christie 2018, p. 282). In addition, starting with these early reviews made it possible to study the evolution of the reception of didacticism over time and to compare this with the findings of my textual analysis. In order to qualify for inclusion in the didactic corpus, novels had to

1. have been praised by at least one reviewer for their ability to instruct as well as amuse or entertain readers,²
2. be set primarily in Britain in the period of their conception, and

¹ Partial reviews of all of the novels published between 1770 and 1799 are available in Raven et al. 2000, while all full reviews for the period 1800–1829 can be found in P. Garside et al. 2004.

² Both instruction and amusement were necessary for inclusion in keeping with the *utile et dulce* formula central to 18th-century conceptions of the value of fiction (Wood 2003, p. 15; Millet 2007, p. 43).

1778	<i>Evelina</i>	Frances Burney
1778	<i>Munster Village</i>	Lady Mary Hamilton
1782	<i>Cecilia</i>	Frances Burney
1788	<i>Mary, A Fiction</i>	Mary Wollstonecraft
1790	<i>Julia, A Novel</i>	Hannah Maria Williams
1796	<i>Hermsprong, or Man as He Is Not</i>	Robert Bage
1796	<i>Memoirs of Emma Courtney</i>	Mary Hays
1798	<i>Edgar: or, The Phantom of the Castle</i>	Richard Sicklemore
1798	<i>Maria; or, The Wrongs of Woman</i>	Mary Wollstonecraft
1801	<i>Belinda</i>	Maria Edgeworth
1801	<i>The Father and Daughter</i>	Amelia Opie
1805	<i>The Nobility of the Heart</i>	Elizabeth Spence
1808	<i>Cœlebs in Search of a Wife</i>	Hannah More
1810	<i>Romance Readers and Romance Writers</i>	Sarah Green
1811	<i>Sense and Sensibility</i>	Jane Austen
1811	<i>Self-Control</i>	Mary Brunton
1813	<i>Pride and Prejudice</i>	Jane Austen
1814	<i>Patronage</i>	Maria Edgeworth

Table 1: The didactic corpus

3. be available in electronic format.

The earliest work in the didactic corpus is Frances Burney’s first published novel, *Evelina* (1778), which was chosen as a chronological starting point based on Burney’s importance as a novelist in the last two decades of the 18th century (Havens 2017, p. 8).³ Eighteen novels fit all the criteria for inclusion; of these, nine were published between 1778 and 1799 and nine between 1800 and 1814 (see Table 1).

A reference corpus was built to provide a representative sample of fiction of the same period to which the didactic corpus could be compared. Here I relied on Mahlberg’s (2013) claim that “corpus work is essentially comparative: a text or text extract is compared to an appropriate reference corpus providing a relative norm” (2013, p. 24). The reference corpus is comprised of 18 novels that were noted for their instructive effect in the *Monthly Review* or the *Critical Review* but otherwise have the same characteristics of being set primarily in contemporary Britain and now available in digital form. The reference corpus also features nine novels published between 1788 and 1799 and another nine that appeared between 1800 and 1814 (see Table 2).

Based on Havens’ and Wood’s claims that didactic novels of the turn of the 19th century predominantly rely on straightforward language to perpetuate moral norms and codes of behavior, the two corpora were compared using computer-aided textual analysis. In particular, this analysis sought to confirm

³ Interestingly, while Havens excludes *Evelina* from her discussion because of its “pervasive satire,” a critic from the *Critical Review* pronounced the book full of “lessons” leading “to improvement and to virtue” thanks in part to its “useful humour and diverting satire” (*CR* 1778, vol. 46: 203).

1778	<i>Learning at a Loss; or, The Amours of Mr. Pedant and Miss Hartley</i>	Gregory Lewis Way
1788	<i>Emmeline; or, The Orphan of the Castle</i>	Charlotte Smith
1791	<i>A Simple Story</i>	Elizabeth Inchbald
1792	<i>Anna St. Ives</i>	Thomas Holcroft
1794	<i>Caleb Williams</i>	William Godwin
1795	<i>Henry</i>	Richard Cumberland
1796	<i>Nature and Art</i>	Elizabeth Inchbald
1798	<i>Rosamund Gray</i>	Charles Lamb
1799	<i>The Vagabond</i>	George Walker
1804	<i>Adeline Mowbray</i>	Amelia Opie
1805	<i>Fleetwood; or, The New Man of Feeling</i>	Caleb Williams
1806	<i>Leonora</i>	Maria Edgeworth
1806	<i>The Wild Irish Girl</i>	Sydney Owenson
1812	<i>The Son of a Genius</i>	Barbara Hofland
1813	<i>The Heroine</i>	Eaton Stannard Barrett
1814	<i>Mansfield Park</i>	Jane Austen
1814	<i>Discipline</i>	Mary Brunton
1814	<i>The Wanderer</i>	Frances Burney

Table 2: The reference corpus

whether the topic of morality and instruction is a salient marker of the register of the didactic corpus and can be established as the primary reason why these novels were received as didactic upon first publication. In fact, my results completely negated this hypothesis, which led at first to considerable frustration. Eventually, however, it prompted an extremely fruitful reappraisal of the complex links between reception and textual elements, as evidenced by digital tools.

3 Testing the Original Hypothesis

From the outset of my research, I endeavored to trace the themes of morality and instruction in the didactic corpus as an element of textual register (Biber et al. 2019, p. 40). Given Biber's claim that "the words used in a text are to a large extent determined by the topic of the text" (p. 40), keyword analysis was chosen as a means to determine whether these themes were particularly salient in the didactic corpus in comparison to the reference corpus. These keywords were, thus, posited as a potential aspect of the self-evident register of what has been termed "overt didacticism" in novels of the period (Butler 1972, p. 449; Wood 2003, p. 12).⁴ My hypothesis was that novels received as didactic would be likely to engage with questions of morality and instruction in

⁴ Word frequency is often used to attribute authorship in corpus linguistics and stylistics (Jockers 2013, p. 70; Burrows 2018, p. 724; Szudarski 2018, p. 25), and it has also been used to classify novels in terms of genre (Allison et al. 2011, p. 5).

a direct manner perceivable in terms of vocabulary frequency. This stemmed from the overall critical consensus regarding the porosity between conduct books and novels, and particularly between conduct books aimed at women and novels written by women (Bilger 1998, p. 21; Spencer 1986, p. 142). This seemed particularly pertinent given that 16 of the 18 novels in my didactic corpus were written by women. Based on concepts from 18th-century moral philosophy and definitions from the *Oxford English Dictionary*,⁵ I consequently devised a list of words related to morality and instruction to be quantified using the TXM software. The latter allows for the creation of complex Corpus Query Language (CQL) queries, making it possible to combine several terms into a concept or theme to be studied.

Given the influence of Adam Smith's system of moral philosophy on 18th-century thought (Howell 1971, p. 447), I included Smith's "cardinal virtues," namely prudence, benevolence, justice, self-command, and sympathy, in my list of moral terms (Haakonssen 2002, pp. viii, xiii, xx). I added the term "modesty" since it is particularly associated with women in John Locke's *Some Thoughts Concerning Education* (1902, p. 164).⁶ "Honor" and "courage" were also added because critics mentioned them in the novels' reviews. Similarly I included "sensibility" and "delicacy" as Hugh Blair considered them the grounds for a "superior moral life" (quoted in Van Sant 1993, p. 5). "Reason" and "passion" made it onto the list as the main contentious forces at play in the exercise of virtue according to Mary Wollstonecraft (2004, pp. 30–31). "Propriety" was included based on Jane Spencer's claims about its growing importance throughout the 18th century and its links to morality and modesty, particularly for women (1986, p. 75). I added "duty" and "conduct" to complement ideas about the behavioral norms that helped define "morality" along with "propriety." The list of terms related to instruction was built using definitions and synonyms from the *Oxford English Dictionary* online.

In order to compare the values yielded by TXM, I used a per-mill approach as well as log-likelihood (LL), a "test [which] helps you determine whether differences in the frequency of words are reflective of the actual variation in language or whether they result from chance occurrences" (Szudarski 2018, p. 27). According to Rayson, Berridge, et al. (2004), "one million words gives sufficient evidence for mid- to high-frequency words" in corpus linguistics studies (p. 1). Since my corpora were respectively 2, 532, 943 and 2,683, 379 words long based on a count by TXM, the log-likelihood test was well-placed to deliver valid results.

The threshold for statistical significance commonly used for statistical measures such as log-likelihood is 5%, which amounts to a critical value of 3.84.⁷ In recent years, however, critics have questioned the pertinence of the log-likelihood test if used on its own. It has been argued, for instance, that the test claims to detect too many significant differences when comparing two corpora

⁵ *OED Online*. Oxford University Press, June 2021 [accessed 1 July 2021.]

⁶ John Locke is directly quoted in several novels from both corpora: Lady Mary Hamilton's *Munster Village* (1778), Hannah More's *Cælebs in Search of a Wife* (1808), and Maria Edgeworth's *Patronage* (1814) in the didactic corpus, and George Walker's *The Vagabond* (1799), Amelia Opie's *Adeline Mowbray* (1804), and Sydney Owenson's *The Wild Irish Girl* (1806) in the reference corpus.

⁷ See, for example, the online calculator <https://www.korpus.cz/calculator/>.

(Bestgen 2017, p. 37). Moreover, log-likelihood is a measure of statistical significance, and “does not by itself inform us of whether the difference between the frequencies [...] carries any descriptive value” (Fidler et al. 2015, p. 227). Effect size statistics may be used to complement research as they “focus [...] on how large the difference between the two frequencies of a word is” (Pojanapunya et al. 2018, p. 145). One example of an effect size metric is the Log Ratio (LR); this is included in Rayson’s online calculator, the tool used to process the data set presented in Tables 3 and 4.⁸

Each word category from Tables 3 and 4 includes all the grammatical forms of the lemma that pertain to the central notion. For example, the category “instruction” includes the nominal and verbal lemmata “instruction” and “instruct”. Where a grammatical category changed the fundamental meaning of a word, that category was not included in the table, and concordance lines were used to select relevant occurrences of polysemous words based on the context in which they were used. An example is the verb “to conduct,” which among other things may mean “to behave” or “to lead.”

These tables suggest that the two corpora do not differ materially when it comes to the explicit presence of the topics of morality and instruction, the central elements of the concept of didacticism in fiction at the time (Havens 2017, p. 5).

The results presented in Tables 3 and 4 invalidate the hypothesis tested by this corpus-based approach. Although the differences in the frequency of several terms in Table 3 are shown to be statistically significant, the log-likelihood values remain rather low. It therefore becomes difficult to make any reliable claim about the greater engagement with the topic of morality and instruction of the didactic corpus when compared with the reference corpus. The LR measure corroborates this: if a word has the same relative frequency across the corpora, its LR value is 0; if it is twice as common in the analyzed corpus, its LR value is 1, and every additional point represents a doubling of the ratio (Collins et al. 2020). In Table 3, “prudence” stands out as the term with the highest LL and LR values, but the overall picture suggests that both corpora include the topic of morality to similar degrees. This is also true for Table 4, which shows little difference in the presence of the topic of instruction in the two corpora, as seen in the list of terms. Furthermore, two of the three lemmata with the highest LL and LR values, “edify” and “tutor” are actually over-represented in the reference corpus.

To complement these results, I completed a qualitative review of the endings of all of the novels. My aim was to determine whether language about the moral conclusions to be drawn from the narrative was more prevalent and/or less ambiguous in the didactic novels than in the reference novels. In both corpora, 15 of the 18 novels mention a vice punished and a virtue rewarded in their closing paragraphs; this mirrors the quantitative results in Tables 3 and 4. Even more strikingly, the endings of five of the 18 novels in the didactic corpus are morally ambiguous in some respect. This may be seen, for example, in the last words of Frances Burney’s *Cecilia* (1782) about the spendthrift Mrs. Harrel, who does not learn from her first husband’s financial ruin and subsequent

⁸ Rayson’s log-likelihood calculator can be found at <http://ucrel.lancs.ac.uk/llwizard.html>.

	Didactic corpus 2,532,943 tokens	%	Reference corpus 2,683,379 tokens	%	Log-likelihood	Log Ratio
benevolence	270	0.11	307	0.11	-0.72	-0.10
conduct	669	0.26	522	0.19	27.66	0.44
delicacy	479	0.19	406	0.15	10.98	0.32
duty	535	0.21	584	0.22	-0.25	-0.04
hono(u)r	1075	0.42	998	0.37	9.03	0.19
justice	318	0.13	457	0.17	-17.69	-0.44
modest	182	0.07	211	0.08	-0.8	-0.13
moral	264	0.10	291	0.11	-0.22	-0.06
passion	806	0.32	922	0.34	-2.54	-0.11
proper	565	0.22	580	0.22	-0.28	0.05
prudence	365	0.14	226	0.08	41.51	0.77
reason	1314	0.52	1225	0.46	10.37	0.18
self-command	32	0.01	36	0.01	-0.06	-0.09
sensibility	254	0.10	280	0.10	-0.21	-0.06
sympathy	170	0.07	153	0.06	2.14	0.24
vice	203	0.08	212	0.08	-0.02	0.02
virtue	728	0.29	816	0.30	-1.23	-0.08
TOTAL	8229	3.25	8226	3.07	13.86	0.08

Table 3: Lemmata related to morals in both corpora

	Didactic corpus 2,532,943 tokens	%	Reference corpus 2,683,379 tokens	%	Log-likelihood	Log Ratio
advise	183	0.072	206	0.077	0.36	-0.09
edify	11	0.004	44	0.016	-19.34	-1.92
educate	338	0.133	262	0.098	14.54	0.45
enlighten	45	0.018	62	0.023	-1.82	-0.38
explain	476	0.188	398	0.148	12.2	0.34
improve	310	0.122	270	0.101	5.55	0.28
inculcate	25	0.009	12	0.004	5.45	1.14
influence	304	0.120	353	0.132	-1.38	-0.13
inform	804	0.317	735	0.274	8.36	0.21
instruct	199	0.079	254	0.095	-3.9	-0.27
learn	498	0.197	625	0.233	-8	-0.24
lesson	94	0.037	139	0.052	-6.34	-0.48
study	360	0.142	367	0.137	0.27	0.06
teach	349	0.138	407	0.152	-1.74	-0.14
tutor	19	0.008	49	0.018	-12.03	-1.28
urge	180	0.071	263	0.098	-11.22	-0.46
TOTAL	4195	1.656	4446	1.657	0.00	0.00

Table 4: Lemmata related to instruction in both corpora

suicide, and far from being punished, “married very soon a man of fortune in the neighbourhood, and quickly forgetting all the past, thoughtlessly began the world again, with new hopes, new connections,—new equipages and new engagements!” (Book 10, chapter 10). In contrast, only one novel from the reference corpus includes a morally ambiguous ending: the cautionary tale promised at the beginning of William Godwin’s *Caleb Williams* (1794) turns out to simply be a profession of frankness by the autodiegetic narrator, without any consequences for his perceived vices.

Yet while the tables and analysis of endings show that overall and unexpectedly the novels of both corpora engage with the topics of morality and instruction to similar degrees, a divergence based on gender emerges for three terms from Table 3. In these cases, the LL and LR values respectively show a statistically significant difference in frequency and comparatively greater differences in frequencies within the data. The terms “conduct” and “prudence,” which are both over-represented in the didactic corpus, evoke values prevalently attached to femininity in the period.⁹ In contrast, “justice” is a term linked to the traditionally and historically male-dominated world of legal power as the basis of jurisprudence in Adam Smith’s moral philosophy (Haakonssen 2002, p. xx). The didactic corpus features works predominantly written by women (16 of the 18 novels) whereas the reference corpus is comprised of 10 works by female authors and eight by male authors, a rough reflection of the gendered distribution of authorship in the period (Mandal 2007, pp. 13, 27). The ratio of female to male protagonists in the novels in each corpus mirrors these proportions, with 15 novels from the didactic corpus and 10 from the reference corpus featuring a female main character. At the same time, some of the male-authored novels in both corpora have a female protagonist, and vice versa. It is therefore not surprising to find that gender plays a role in the differences in vocabulary use within the corpora, especially given that didacticism has often been linked to female authorship (Towsey 2015, p. 33; Havens 2017, p. 13).

4 Looking for Overt Didacticism...

Before pursuing the angle of gender, however, I was prompted by my initial results to look for more lexical and syntactical markers of what has been seen as overt didacticism (Butler 1972, p. 449; Wood 2003, p. 12). Here I moved beyond tracing moral didacticism as a topic to investigating the linguistic features of its register. Wood (2003) has opposed overt didacticism to the kinds of indirection found and valued, for instance, in the works of Jane Austen, and she notes the importance of an authoritative narrative voice in making overt didacticism effective (p. 66). Susan Lanser (1992) conceives of an overt authoriality involving narrative voices that engage in extra-representational acts such as “reflections, judgments, generalizations about the world ‘beyond’ the fiction, direct addresses to the narratee, comments on the narrative process, allusions to other writers and texts” (pp. 16–17). She also contrasts this explicit authoriality with forms of indirection such as “free indirect discourse, irony, ellipsis, nega-

⁹ See, for example, Butler 1987, p. 122; Mellor 1993, p. 40; Spencer 1986, p. 75.

tion, euphemism, [and] ambiguity,” which are characteristic of Austen’s novels (p. 62). Lesser-known female novelists of the 18th century, she suggests, tended to engage in this overt authoriality even though women novelists consistently received more praise when their authoriality was most covert (pp. 66, 78).

Given that my didactic corpus consists mostly of works which have not entered the literary canon, I was led to investigate whether overt authoriality is a marker of overt didacticism as the register of these novels. These findings were then compared with those for the reference corpus. Again, this did not yield the expected results but instead showed that an authoritative tone is not a prevalent feature of moral didacticism as this was received in novels at the turn of the 19th century. My investigation, for example, of direct addresses to readers (DAR) in the prefaces and main texts of the novels in both corpora found that this device is by no means specific to the novels of the didactic corpus. Moreover, rather than being a strategy to ensure readers’ ideological assent, in both the corpora, DAR marks an attempt to negotiate the places of the author, the reader, and the critic in relation to one another at a time when the novel was in the process of becoming a legitimate genre (Misset [forthcoming]).

My study of DAR also showed that the quintessential didactic novel of the period, Hannah More’s *Cælebs in Search of a Wife* (1808) is far from representative of didactic fiction in general.¹⁰ While the addresses to readers in the other novels of both corpora are voiced by the narrator to extradiegetic readers, 10 of the 11 DAR in *Cælebs* take place through intradiegetic dialogue, as shown in Table 5.

All the characters featured in Table 5 are firmly established as sound moral authorities in the novel, and eight of the 11 occurrences directly address the question of moral improvement through various kinds of reading material. This was a unique finding in the context of the DAR in the other novels in both corpora. Instead of confirming moral didacticism as a unifying trait of DAR, I therefore found their use as vehicles for overt didacticism to be an exception.

5 ...and Finding Gender and Class Bias

The results of the corpus-based approach proved crucial for reappraising ideas about overt didacticism as a fictional register. Nevertheless more information was needed in order to move the focus from what didacticism is *not* according to these corpora and, thus, to try to determine what it is. This led me to adjust my method from one that was corpus-based to one that was corpus-driven and so would generate rather than verify hypotheses (Cornby et al. 2016, p. 7). I consequently switched analytic tools from TXM to AntConc, a concordance tool which, as its creator writes, generates keyword lists that show “which words are unusually frequent (or infrequent) in the corpus in comparison with the words in a reference corpus. This allows you to identify characteristic words in the corpus, for example, as part of a genre [...] study” (Anthony 2019b, p. 7).

¹⁰ *Cælebs* is known for being an early 19th-century bestseller (Stott 2003, pp. 277, 281) and often cited as an example of didactic fiction of the period (Kelly 2018, p. 198; Wood 2003, p. 66; Mandal 2007, p. 95).

Cœlebs	however, will inevitably dazzle the feeling	reader	, till it produce the common effect of
Mr S	simplicity, as far removed from the careless	reader	of a common story, as from the declamation
	Lady Belfield, who, though not new to the	reader	or the writer, were new at Stanley Grove. ^a
Mrs S	the omnipotence of love, that the young	reader	was almost systematically taught an
Cœlebs	it may and does tire the patience of the	reader	, yet it never leaves him ignorant; and of
Mr S	human nature is corrupt; that the young	reader	is helpless, and wants assistance; that he is
Sir J	fancy, nor to extinguish a taste for them in	readers	. “” Show me any one instance of good that
Cœlebs	elevation of fancy led Milton, or Milton his	readers	? Into what immoralities did it involve
Cœlebs	of the living minstrel of the LAY? What	reader	has Mason corrupted, or what reader has
Cœlebs	What reader has Mason corrupted, or what	reader	has Cowper not benefitted? Milton was an
Sir J	communicated, a hundred thousand	readers	caught, the infection. Sentimentality was

Table 5: DAR in *Cœlebs in Search of a Wife*^b

^a This is the only occurrence which is not reported speech.

^b This table was created using concordance lines generated by TXM

AntConc's keyword list ranks words according to their "keyness," which is measured using log-likelihood—the higher a word's keyness score, the more characteristic it is of the corpus as compared to the reference corpus. At the same time, the tool takes into account the difference in the size of the corpora it compares.¹¹ The keyword lists may also be made to incorporate Log Ratio as a measure of effect size.

In order to compare the novels in the didactic and the reference corpora using AntConc, the texts were lemmatized manually using TreeTagger. This merged all the different inflections of the same lemma into one term. The results of this method corroborate those of the corpus-based approach insofar as the terms found to be statistically overused in one corpus or the other using the LL calculation website and the TXM frequency values appear as keywords. "Prudence," for instance, (rank 481, LL 26.09, LR 0.7905) appears in the keyword list when the didactic corpus is set as the primary corpus, while "justice" (rank 265, LL 28.95, LR 0.5977) occurs in the list when the reference corpus is set as the primary corpus.¹²

The most striking feature of the keyword lists generated through AntConc's comparison of didactic and reference corpora is the gendered divide. This is clear from the frequency of female pronouns and nouns, and it corroborates the findings of the corpus-based study using TXM. When the didactic corpus is compared to the reference corpus in AntConc, the highest-ranking keywords that are not characters' names are "lady" (rank 12, LL 968.62, LR 0.8949), "she" (rank 25, LL 708.62, LR 0.3463), "her" (rank 32, LL 555.65, LR 0.2492), "ladyship" (rank 64, LL 332.43, LR 1.6632), and "daughter" (rank 163, LL 106.28, LR 0.6425). This suggests that the novels in the didactic corpus focus on female characters more than the novels in the reference corpus do. "Lady" and "ladyship" had the highest LR values among these didactic novels, which points to the setting of the novels of the didactic corpus in genteel society more often than the novels of the reference corpus.

Turning to the distribution of "lady" across the different novels in the didactic corpus, we find that the five works with the lowest concentration of the term are Richard Sicklemore's *Edgar, or The Phantom of the Castle* (1798, 6 hits; this is also the only novel in the corpus to focus almost exclusively on male characters); Mary Wollstonecraft's *Maria, or The Wrongs of Woman* (1798, 13 hits) and *Mary*,

¹¹ At the outset, I was directed to TXM as a tool for my research based on its ability to accommodate complex CQL queries. This appeared particularly useful at the start of my project, which originally aimed to trace the topics of morality and instruction. Although the two tools are not fundamentally different, AntConc was suggested for the subsequent exploratory phase of my research given the ease with which it can compare keyword use between two corpora.

¹² The LL values calculated based on the TXM data and the ones produced by AntConc do not quite coincide. However, this is a common phenomenon when using different tools to study the same corpora (Anthony 2013, p. 149). TXM and AntConc are complex software tools that are programmed differently, and they may make calculations in slightly different ways. This is a drawback of using ready-made software where the user cannot easily access all the settings (Gries 2009, p. 2). More specifically, AntConc can only count one lemma at a time, but I could combine different lemmata derived from the same notion (eg. "moral" and "morality"), as applicable, when working with the TXM data. I could also clean the data when faced with polysemous terms such as "conduct, n." and "conduct, v." Notwithstanding these issues, the different LL values calculated using the two sets of tools do not materially impact the overall results, and the findings correlate.

A Fiction (1788, 13 hits); Mary Hay's *Memoirs of Emma Courtney* (1796, 37 hits); and Amelia Opie's *The Father and Daughter* (1801, 38 hits). The four latter works all feature female protagonists who are gentry women in stories that do not, however, revolve around genteel social life. At the other end of the spectrum, the frequency of "lady" is highest in Maria Edgeworth's *Belinda* (1801) where it records 1554 hits, and most of the 12 remaining novels contain between 200 and 600 hits.

In the reference corpus, "lady" has the highest concentration in novels set among genteel society. These works overwhelmingly feature a female protagonist with only one, Richard Cumberland's *Henry* (1795, 476 hits) concentrating on a male protagonist. On the other hand, the seven novels with the lowest concentration of the term all focus on male protagonists, and in four cases, these characters come from lower walks of life than their counterparts in other novels in the corpora: they are poorer and have to earn their daily bread, whether as subordinates living in a wealthy household (William Godwin's *Caleb Williams*, 1794, 37 hits) or as traders (Elizabeth Inchald's *Nature and Art*, 1796, 61 hits; Barbara Hofland's *The Son of a Genius*, 1812, 20 hits; and George Walker's *The Vagabond*, 1799, 10 hits). The reference corpus, thus, includes novels with a greater variety of social settings and gendered perspectives than the didactic corpus. Female genteel experience may therefore be seen as a defining feature of the didactic corpus and one that is quantifiable through lexical frequency.

Finally, a comparison of the novels of the five authors who appear in both corpora (Jane Austen, Mary Brunton, Frances Burney, Maria Edgeworth, and Amelia Opie) confirms the tendency of the novels received as didactic by *Monthly Review* and *Critical Review* critics to be set in genteel society. At the same time, it highlights that these narratives had to follow a certain pattern. All of the novels by these authors in the corpora center on a genteel female protagonist who is navigating moral questions, but the novelists' word use distinguishes the works in the didactic corpus from the ones in the reference corpus. Terms such as "say," "conversation," "company," "behaviour," "please," and "manner" are over-represented in the novels of the didactic corpus, which illustrates their central focus on the social behavior of genteel life. In contrast, the over-representation in the reference corpus of "my," "self," and "feeling," all of which are fairly evenly distributed among these novels, suggests a greater focus on the personal experiences of their female protagonists.

6 Conclusion

My research on these two corpora is still in progress. Nevertheless this use of quantitative data obtained through TXM and AntConc has shown the value of interrogating the reception of texts in light of their linguistic features: this method can help determine the part ideology has played in previous readings and categorizations. In the current case, the novels that critics initially received as instructive for readers are the ones that largely support and reinforce the specific norms of behavior and social hierarchies that would become central to the Victorian ethos. This is most notable in the foregrounding of genteel do-

mesticity as a feminine ideal, evocative of the “Angel in the House” (Bilger 1998, p. 85). Crucially, we find that the elements of register specific to didactic novels as a fictional subgenre hinge not on engagement with the topics of morality and instruction or on any tonal authority but rather on the presence of specific kinds of characters—genteel women—whose narrative and moral trajectories center their social interactions rather than their personal experiences and development. These findings have completely redirected my research and form the foundation for my ongoing qualitative work on the narrative trajectories that distinguish the novels of the didactic corpus from those of the reference corpus. As such, they allow for new claims to be made and substantiated about the register and reception of these novels.

References

- Allison, Sarah, Ryan Heuser, Matthew Jockers, Franco Moretti, and Michael Witmore (2011). “Quantitative Formalism: An Experiment”. In: Stanford Literary Lab. Chap. Pamphlet 1. URL: <https://litlab.stanford.edu/pamphlets/>.
- Anthony, Laurence (2013). “A Critical Look at Software Tools in Corpus Linguistics”. In: *Linguistic Research* 30.2, pp. 141–161.
- Anthony, Laurence (2019a). *AntConc*. Version 3.5.8. Tokyo. URL: <https://www.laurenceanthony.net/>.
- Anthony, Laurence (2019b). *AntConc Help Manual*. Tokyo: Center for English Language Education in Science, Engineering, School of Science, and Engineering, Waseda University.
- Bestgen, Yves (2017). “Getting Rid of the Chi-Square and Log-Likelihood Test for Analysing Vocabulary Differences between Corpora”. In: *Quaderns de Filologia: Estudis Lingüístics* 22, pp. 33–56. DOI: [10.7203/qf.22.11299](https://doi.org/10.7203/qf.22.11299).
- Biber, Douglas and Susan Conrad (2019). *Register, Genre, and Style. Second Edition*. Cambridge: Cambridge UP.
- Bilger, Audrey (1998). *Laughing Feminism: Subversive Comedy in Frances Burney, Maria Edgeworth, and Jane Austen*. Detroit, Michigan: Wayne State UP.
- Burrows, John (2018). “Rho-grams and rho-sets: Significant links in the web of words”. In: *Digital Scholarship in the Humanities* 33.4, pp. 724–747. DOI: [10.1093/llc/fqy004](https://doi.org/10.1093/llc/fqy004).
- Butler, Marilyn (1972). *Maria Edgeworth: A Literary Biography*. Oxford: Clarendon Press.
- Butler, Marilyn (1987). *Jane Austen and the War of Ideas*. Oxford: Clarendon Press.
- Butler, Marilyn (1993). “Culture’s Medium: The Role of the Review”. In: *The Cambridge Companion to British Romanticism*. Ed. by Stuart Curran. Cambridge: Cambridge UP.
- Christie, William (2018). “Critical Judgment and the Reviewing Profession”. In: *The Oxford Handbook of British Romanticism*. Ed. by David Duff. Oxford: Oxford UP.
- Collins, Luke C., Elena Semino, Zsófia Demjén, Andrew Hardie, Peter Moseley, Angela Woods, and Ben Alderson-Day (2020). “A linguistic approach to the

- psychosis continuum: (dis)similarities and (dis)continuities in how clinical and non-clinical voice-hearers talk about their voices”. In: *Cognitive Neuropsychiatry* 25.6, pp. 447–465. doi: [10.1080/13546805.2020.1842727](https://doi.org/10.1080/13546805.2020.1842727).
- Cornby, Émeline, Yannick Mosset, and Stéphanie de Carrara (2016). *Corpus de textes : composer, mesurer, interpréter*. Lyon: ENS Éditions.
- Fidler, Masako and Cvrček Václav (2015). “A Data-Driven Analysis of Reader Viewpoints: Reconstructing the Historical Reader Using Keyword Analysis”. In: *Journal of Slavic Linguistics* 23.2, pp. 197–239.
- Garside, Peter, Jaqueline Belanger, and Sharon Ragaz (2004). *British Fiction, 1800–1829: A Database of Production, Circulation & Reception*. Cardiff University. URL: <http://www.british-fiction.cf.ac.uk/>.
- Gries, Stefan (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction*. London, New York: Routledge.
- Haakonsen, Knud (2002). “Introduction”. In: *Quantitative Corpus Linguistics with R: A Practical Introduction*. Ed. by Adam Smith. Cambridge: Cambridge UP.
- Hallowell, Thomas Jewett, Tobias Smollett, and Laurence Hutton (1756–1817). *The Critical Review, or, Annals of Literature*. London: W. Simpkin and R. Marshall. URL: <https://catalog.hathitrust.org/Record/008890497/Home>.
- Havens, Hilary (2017). “Introduction”. In: *Didactic Novels and British Women’s Writing, 1790–1820*. Ed. by Hilary Havens. New York, London: Routledge.
- Heiden, Serge (2010). “The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme”. In: *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*. Tohoku University, Sendai, Japan: Institute of Digital Enhancement of Cognitive Processing, Waseda University, pp. 389–398. URL: <https://aclanthology.org/Y10-1044>.
- Howell, Wilbur Samuel (1971). *Eighteenth-Century British Logic and Rhetoric*. Princeton: Princeton UP.
- Jockers, Matthew (2013). *Macroanalysis: Digital Methods and Literary History*. Urbana, Chicago, and Springfield: University of Illinois Press.
- Kelly, Gary (2018). “The Spectrum of Fiction”. In: *The Oxford Handbook of British Romanticism*. Ed. by David Duff. Oxford: Oxford UP.
- Lanser, Susan (1992). *Fictions of Authority: Women Writers and Narrative Voice*. Ithaca, NY: Cornell University Press.
- Locke, John (1902). *Some Thoughts Concerning Education*. Cambridge: Cambridge UP.
- Mahlberg, Michaela (2013). *Corpus Stylistics and Dickens’s Fiction*. New York, London: Routledge.
- Mandal, Anthony (2007). *Jane Austen and the Popular Novel: The Determined Author*. Basingstoke: Palgrave Macmillan.
- Mellor, Anne (1993). *Romanticism & Gender*. New York: Routledge.
- Millet, Baudoin (2007). ‘Ceci n’est pas un roman’: *l’évolution du statut de la fiction en Angleterre de 1652 à 1754*. Louvain, Paris, Dudley, MA: Editions Peeters.
- Misset, Juliette ([forthcoming]). ‘I hope I shall please my readers’: *Negotiating the Author-Reader Relationship in Two Corpora of British Novels, 1778–1814*.

- Pojanapunya, Punjaporn and Richard Watson Todd (2018). “Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis”. In: *Corpus Linguistics and Linguistic Theory* 14.1, pp. 133–167. DOI: [doi:10.1515/cllt-2015-0030](https://doi.org/10.1515/cllt-2015-0030).
- Raven, James and Antonia Forster (2000). *The English Novel, 1770-1829: A Bibliographical Survey of Prose Fiction Published in the British Isles. Volume I, 1770–1799*. Oxford: Oxford UP.
- Rayson, Paul, Damon Berridge, and Brian Francis (2004). “Extending the Cochran rule for the comparison of word frequencies between corpora”. In: *JADT*, pp. 1–12.
- Rayson, Paul and Roger Garside (2000). “Comparing Corpora Using Frequency Profiling”. In: *Proceedings of the Workshop on Comparing Corpora - Volume 9*. Hong Kong: Association for Computational Linguistics, pp. 1–6. DOI: [10.3115/1117729.1117730](https://doi.org/10.3115/1117729.1117730).
- Spencer, Jane (1986). *The Rise of the Woman Novelist: From Aphra Behn to Jane Austen*. Oxford, UK and Cambridge, Mass.: Blackwell.
- Stott, Anne (2003). *Hannah More: The First Victorian*. Oxford: Oxford UP.
- Szudarski, Pawel (2018). *Corpus Linguistics for Vocabulary: A Guide for Research*. London, New York: Routledge.
- Towsey, Mark (2015). “Women as Readers and Writers”. In: *The Cambridge Companion to Women’s Writing in Britain, 1660-1789*. Ed. by Catherine Ingrassia. Cambridge: Cambridge UP.
- Van Sant, Ann Jessie (1993). *Eighteenth-Century Sensibility and the Novel: The Senses in Social Context*. Cambridge: Cambridge UP.
- Wollstonecraft, Mary (2004). *A Vindication of the Rights of Woman*. London: Penguin Books.
- Wood, Lisa (2003). *Modes of Discipline: Women, Conservatism, and the Novel after the French Revolution*. Lewisburg: Bucknell UP.