

constellation, the novel returns to be central to our understanding of modernity: not despite, but *because* of its pre-modern traits, which are not archaic residues, but functional articulations of ideological needs. To decipher the geological strata of consensus in the capitalist world—here is a worthy challenge, for the history and the theory of the novel.

Style, Inc.: Reflections on 7,000 Titles (British Novels, 1740–1850)

How does a literary historian end up writing on 7,000 titles?¹ In the presentation of 'The Slaughterhouse of Literature' I used the metaphor of the snowball, and that's really how it felt, during these twenty years; at the beginning, I had no idea I would end up doing quantitative research; all I was interested in was evolution, the great theory of how forms change in time; while I was studying evolution, however, Ernst Mayr's theory of speciation made me aware of the role of geography in the generation of new forms; so I turned to cartography, to make literary maps; but maps need homogeneous data, so I started to extract small series from novels (beginnings and endings in Austen, Balzac's young men in Paris), and, later, to make diagrams of library holdings and translation flows. Evolution, geography, maps, series, diagrams . . . One step led to the next; one step asked for the next. And one day I realized that the study of morphological evolution had itself morphed into the analysis of quantitative data.

¹ This article has been greatly improved by exchanges with Sam Bowles, David Brewer, Bob Folkenflik, Matthew Jockers, David Krakauer, and Michael Silverstein: to all of them, my thanks. A further, more abstract study (co-authored with Matthew Jockers and David Krakauer) should soon extend and refine the argument that follows.

But there was a problem. The more 'literary' series (Austen, Balzac, etc.) revealed interesting formal patterns, but had a weak quantitative foundation; the data from book history were more solid, but much further removed from formal analysis. I could find form—or I could find data; but, it seemed, not both of them together. Then, with 'Style, Inc.', the two sides finally clicked, in a full-blown instance of that 'formalism without close reading' that I mentioned in the presentation of 'Slaughterhouse'—or of the 'quantitative formalism' of the first pamphlet of the Literary Lab. But since quantification can only occur if its basic units are reasonably well-defined—and since the units of language are much better defined than those of plot—at this point it was form that split into two, making the formalism of 'Style, Inc.' focus entirely on the analysis of language, with virtually no reference to plot. It's another case, it seems, of the analytical one-sidedness I mentioned in the presentation to 'The End of the Beginning'—which returns one more time in the last piece of this collection, which errs unreservedly in the opposite direction: an analysis of Hamlet's plot, conducted as if Shakespeare's characters didn't utter a single word in the entire play. Clearly, I like to put on 'blinders', as Weber wrote in his lecture on 'Science as a Profession'. It must be the *longue durée* of my Marxist formation (the work of Della Volpe and Colletti), whose ferociously anti-dialectical stance urges me to devote all my energies to analysis, and none to synthesis. And, as I said, I like it.

One last thing that became clear, while working at 'Style, Inc.', was the enormous difference between the archive of the Great Unread, and the world of the canon. You enter the archive, and the usual coordinates disappear; all you can see are swarms of hybrids and oddities, for which the categories of literary taxonomy offer very little help. It's fascinating, to feel so lost in a universe one didn't even know existed; but it's hard to extract a rational picture from this Walpurgisnacht of discordant voices. And then, to make matters worse, there is the opposite problem, too: working with large quantities, the average becomes an inevitable presence—and the average means loss of distinction, slowness,

boredom . . . Too much polyphony, and too much monotony: it's the Scylla and Charybdis of digital humanities. The day we establish an intelligible relationship between these two, a new literary landscape will come into being.



The British novel, from 1740 to 1850. Peripheral, often despised at the beginning of the period, by its end the novel has moved very close to the core of the national culture. So, this is an important century, for this literary form. But, truth be told, the historical framework of this study has been largely dictated by an extrinsic reason: unlike earlier and later periods, from 1740 to 1850 we have very good bibliographies. Which is to say, good lists of titles; in a few years, we will have a digital archive with the full texts of (almost) all novels ever published; but for now, titles are still the best way to go beyond the 1 per cent of novels that make up the canon, and catch a glimpse of the literary field as a whole. And then, titles are not just a good research tool: they are important in themselves—Walter Scott's first word as a novelist, literally, was 'title' ('The title of this work has not been chosen without the grave and solid deliberation . . .')—and they are important because, as Claude Duchet has put it, they are 'a coded message—in a market situation'.² A code, in the market: half sign, half ad, the title is where the novel as language meets the novel as commodity, and their encounter can be extremely illuminating. In what follows, I focus on three moments of this history: first, I describe a major metamorphosis of eighteenth-century titles, and try to explain its causes; next, I suggest how a new type of title that emerged around 1800 may

2 Claude Duchet, "La fille abandonnée" et "La bête humaine"; éléments de titrologie romanesque', *Littérature* 12 (1973), p. 50.

have changed what readers expected of novels; and finally, I make a little attempt at quantitative stylistics, examining some strategies by which titles point to specific genres. Three sections; three pieces in the large puzzle of the literary field.

I

The major metamorphosis of eighteenth-century titles is simple: in the space of two generations, they become much, much shorter. In Figure 1, where their length is measured in the number of words, the median oscillates between ten and twenty words for the first twenty-five years; it drops quickly to ten, around 1770; then to six, by 1790; and it remains there (with minor ups and downs) until the mid nineteenth century.³ From fifteen or twenty words, to six. And titles don't just become shorter, in the course of these 110 years, they also become much more similar to each other: in Figure 2, the steep drop of the standard deviation (which measures the degree of variation within a system) indicates precisely how rapidly the

3 The chart in Figure 1 follows both mean and median length in order to offer a more complete picture of how titles vary: the mean (or average) provides information about the often extravagant length of some titles—while the median draws attention to the ‘central’ length of each given year (that is to say, that which has an equal number of results above and below it). The difference between the two forms of measurement becomes particularly clear in years such as 1780 (with the 346-word long *History of Miss Harriot Fairfax*), or 1784 (with the 273 words of *The Maid of the Farm; Or memoirs of Susannah James*): in these two instances, the mean balloons to 37.9 and 19.7 respectively, while the median (8.5 and 7) is hardly affected.

Unless otherwise indicated, the sources of all the figures are the following: 1740–49, Jerry C. Beasley, *Novels of the 1740s*, Georgia, 1982; 1750–69, James Raven, ed., *British Fiction 1750–1770: A Chronological Checklist of Prose Fiction Printed in Britain and Ireland*, Delaware, 1987; 1770–1829, Peter Garside, James Raven, and Rainer Schöwerling, eds, *The English Novel 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, vols. I and II, Oxford 2000; 1830–36, online database ‘The British Novel 1830–36’, directed by Peter Garside at Cardiff University; 1837–50, Andrew Block, ed., *The English Novel 1740–1850*, London 1968.

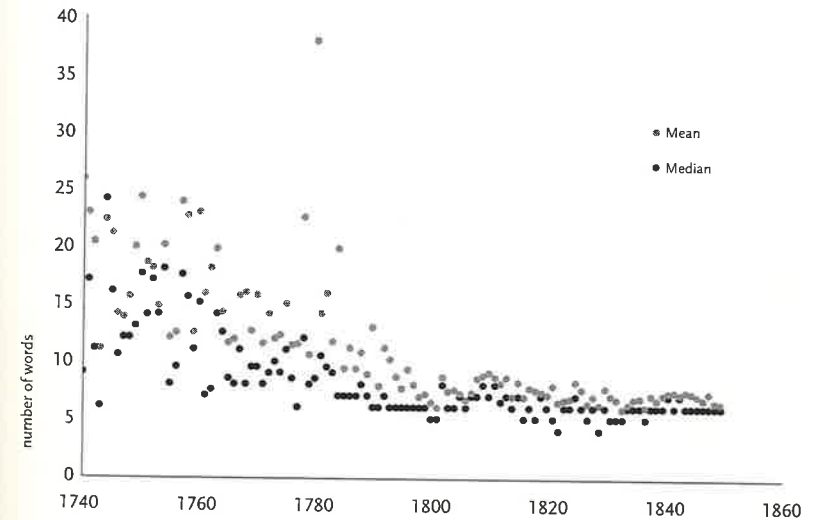


Figure 1: Length of Titles

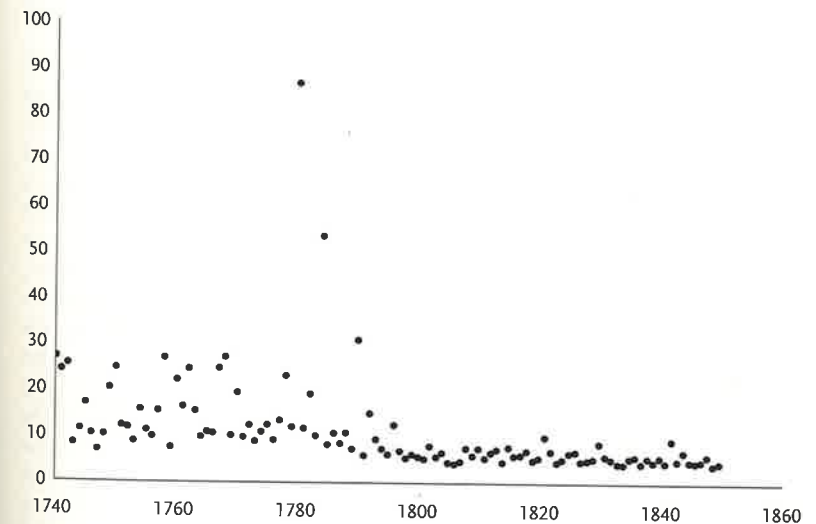


Figure 2: British novels 1740-1850, standard deviation from the mean

range of options is shrinking. To understand what this means, look at the distribution of mid-eighteenth-century titles in Figure 3: many of them are already quite short, with between one and ten words, but there is still a lot of variety, with plenty of titles which use fifteen words, twenty, twenty-five, thirty, forty, or more. A hundred years later (Figure 4) this 'tail' is gone, and long titles have virtually vanished.⁴ It's not just that all titles are becoming shorter, in other words; it's also that a certain type of title disappears altogether. How long is 'long' is of course an open question, but if we set the limit at fifteen or twenty words—which is quite long, for a title—then long titles were between 40 and 60 per cent of the total in the mid eighteenth century (Figure 5, overleaf); by 1800 their number had already dropped to 5 to 10 per cent; eventually, they just disappeared. Why?⁵

4 The thirty novels in the *New York Times* best-seller list in November 2006 used between one and six words; the forty in November 2008, between one and seven. In both cases, the mean was around 2.7 words—slightly higher than Austen's 2.0.

5 Counting the number of words in a title . . . But what exactly *is* a title? Among the novels for the year 1802, Peter Garside's masterful bibliography lists *Delaval. A Novel. In three Volumes*. But are expressions that point so explicitly to extra-textual realities like 'in three Volumes' (or 'dedicated to Her Royal Highness The Dutchess of York', 'from the French of M. Victor Hugo', and so on) really part of the title? In my opinion, no; and so, useful as such information is in other respects, I have removed it from the database, leaving the title in question as *Delaval. A Novel*. But what about 'A Novel' ('A Romance', 'A Tale', 'In a Series of Letters')? Here, the reference is not so much extra-, as meta-textual: all these markers designate a *class*, rather than a specific book: invaluable for the analysis of novelistic sub-genres, they have little or nothing to say about individual cases. As a consequence, I have preserved them the first few times they appear in a title (when they are presumably indicating something new and specific about the given book), and deleted them thereafter; making an exception for those bizarre cases where the wider class is evoked only in order to estrange it: 'A Rhapsodical Romance', 'A Dramatic Novel', 'A Neapolitan Tale'—and also, sure enough, 'A Novel Without a Hero'. This said, since my choices may strike some readers as not merely subjective but perverse, Figure 6 charts the length of titles as they appear in the bibliographical sources, without any intervention on my part. As a comparison with Figure 1 shows, the general trend does not

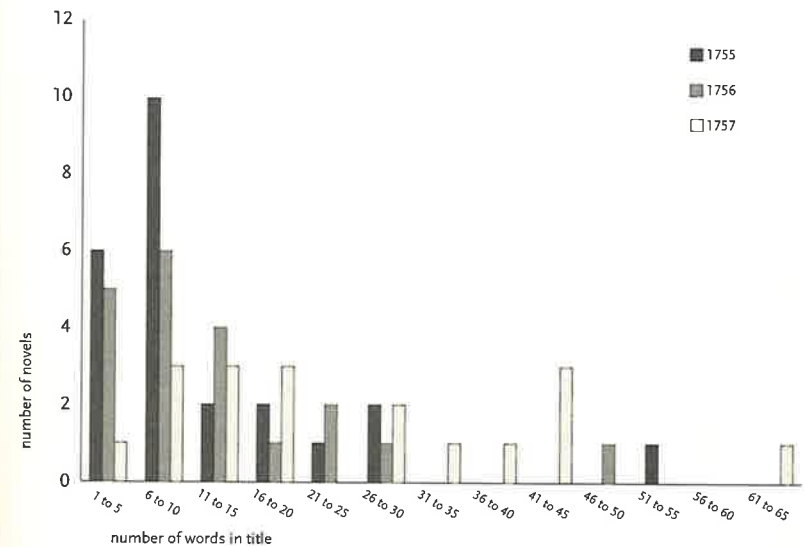


Figure 3: Mid-18th century titles

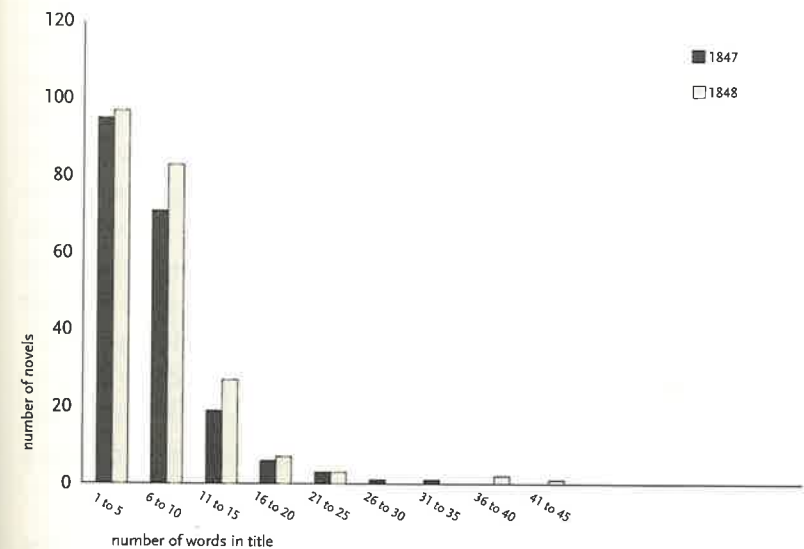


Figure 4: Mid-19th century titles

And, before coming to that, what were they like, those long titles—what did they *do*, with all those words? Usually, they provided a summary of the novel: *A letter from H—g—g, Esq; One of the Gentlemen of the Bedchamber to the Young Chevalier, And the Only Person of his Retinue that attended him from Avignon, in his late Journey through Germany, and elsewhere; Containing Many remarkable and Affecting Occurrences which happened to the P— during the course of his mysterious Progress. To a Particular Friend.* Today, this sounds odd; but actually, a summary at the beginning of a novel makes sense: a novel is a narrative, and the title—the title-page, here one sees why books needed a whole page for their title—the title-as-summary was a *shorter* narrative: it presented the main events of the story, the characters, the setting, the ending. It made sense.

But the cultural ecosystem was changing in a way that was incompatible with it: in the course of the eighteenth century, the publication of novels in Britain grew dramatically (Figure 7, overleaf), from a few books a year in the early decades, to twenty-five or so in mid-century, seventy to eighty around 1800, and about a hundred a year in early Victorian times. And as more novels circulated, two things happened. In the third, and even more so in the fourth quarter of the eighteenth century, the *Monthly* and other magazines started to publish reviews of many new novels, making title-page summaries somewhat superfluous: as the literary system grew, in other words, some of its functions became more specialized, ‘freeing’ titles from having to provide a detailed description. And then, as the number of new novels kept increasing, each of them had inevitably a much smaller ‘window’ of visibility on the market, and it became vital for a title to catch quickly and effectively the eye of the public. Summaries were not good at that. They were good at describing a

change much: the decline in length is slightly less dramatic (the median being significantly higher for the first forty years, and then stabilizing around seven or eight words rather than six), but equally evident.

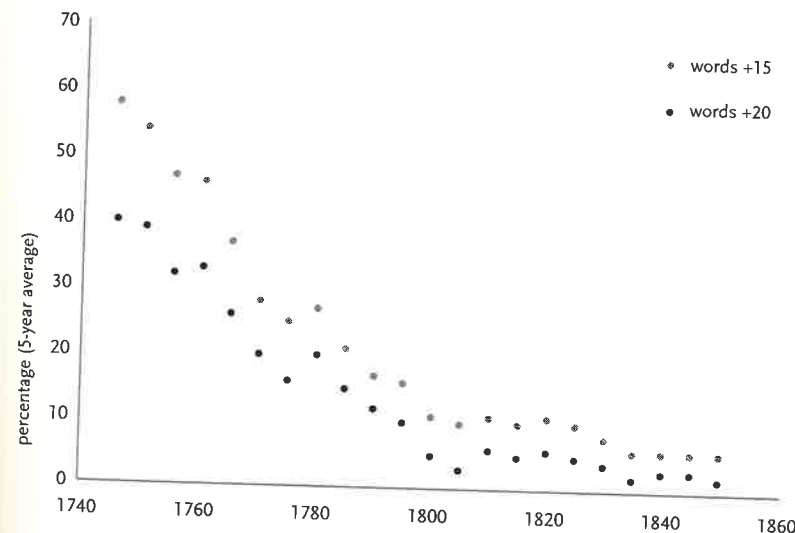


Figure 5: Novels with very long titles

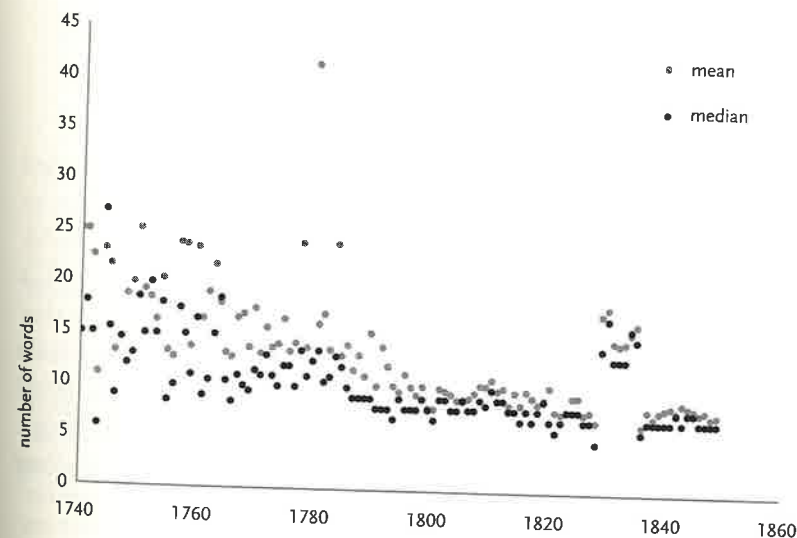


Figure 6: Length of titles (unedited)

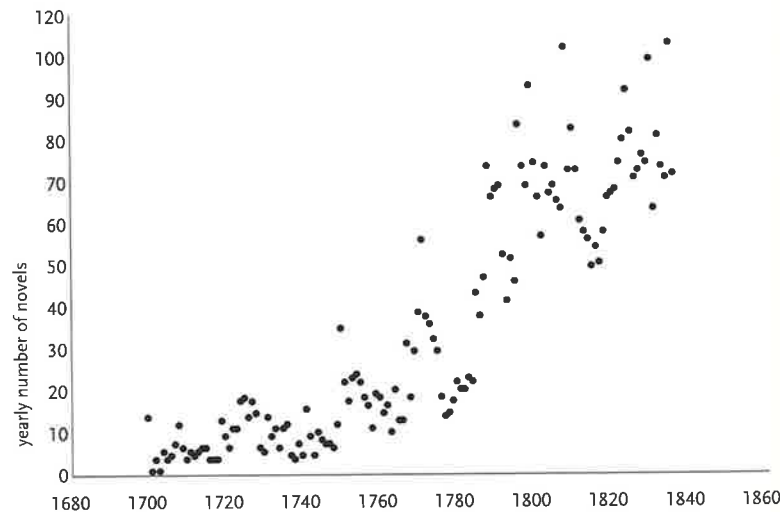


Figure 7: Publication of British novels, 1700–1836

Additional source: 1700–39, William H. McBurney, ed., *A Check List of English Prose Fiction, 1700–1739*, Harvard 1960. The chart stops in 1836 because it seems very likely that Block's bibliography (*The English Novel 1740–1850*) significantly overstates the number of novels published after that date.

book in isolation: but when it came to standing out in a crowded marketplace, short titles were better—much easier to remember, to begin with (but not only). That's why long titles disappeared: because between the size of the market, and the length of titles, a strong negative correlation emerged: as the one expanded, the other contracted. Nothing much had changed, in the length of titles, for a century and a half, as long as the production of novels had remained stable around five or ten per year; then, as soon as publishing took off in earnest, titles immediately shrank (Figure 8). By 1790, their 'quantitative' transformation was virtually complete.⁶

⁶ The other type of title that disappeared in the late eighteenth century was the 'title-compilation', like the 1772 *The Egg, Or the Memoirs of Gregory Giddy*,

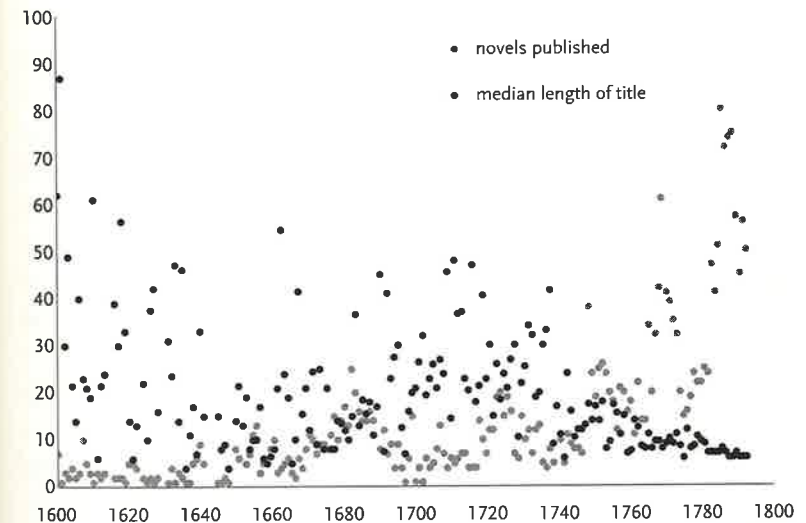


Figure 8: Size of the literary market, length of novelistic titles

Additional source: 1600–1700, Charles C. Mish, ed., *English Prose Fiction, 1600–1700: A Chronological Checklist*, Charlottesville, VA 1967.

As long as only a few novels per year are published, the median length of titles keeps oscillating between 10 and 40–50 words; after the first 'rise' of the 1720s and '30s, however, it drops below 20, and below 10 during the late-eighteenth century take-off first noticed by Clifford Siskin. Earlier on, more ephemeral publishing spurts (1655–60; the 1680s) had also coincided with a marked drop in the length of titles.

Esq: With the Lucubrations of Messrs. Francis Flimsy, Frederick Florid, and Ben Bombast. To which are Added Private Opinions of Patty Pout, Lucy Lucious, and Priscilla Positive. Also the Memoirs of a Right Honourable Puppy. Or Bon Ton Display'd: Together with Anecdotes of a Right Honourable Scoundrel. Conceived by a Celebrated Hen, and Laid Before the Public by a Famous Cock-feeder. Just as summaries drew the reader's attention to the multiplicity of episodes along the diachronic axis, compilations foregrounded a 'horizontal' proliferation of perspectives, characters, and locations—a naive, picaresque-like poetics of 'variety' (to use a keyword of the age), at the very moment when novelistic structure was becoming tighter and more homogeneous. Out of step with the times, summaries and compilations became even more unimaginable in the nineteenth century.

The market expands, and titles contract. Figure 8 showed the temporal correlation of the two processes; a closer look at the market adds a specific causal relationship. Because 'market', in the late eighteenth century—when readers almost never bought novels—really means: circulating libraries. Commercial enterprises, which disseminated the novel throughout Britain (and France, and Germany: one of Brecht's early plays, *In the Jungle of Cities*, opens in one of these libraries), and whose catalogues have frequently survived until today. Catalogues: lists of titles. But not quite the same titles we find in Raven and Garside. In Sander's library, in 1780s Derby, *Capacity and Extent of the Human Understanding; Exemplified in the Extraordinary Case of Automathes: A Young Nobleman; who was Accidentally left in his Infancy, upon a desolate Island, and continued Nineteen Years in that solitary State, separate from all Human Society. A Narrative abounding with many surprising Occurrences, both Useful and Entertaining to the Reader*, becomes: *History of Automathes, A Young Nobleman*. At Phorson's, in Berwick, in 1790, *Unfortunate Sensibility; or, the Life of Mrs L****. Written by Herself. In a Series of Sentimental Letters. Dedicated to Mr. Yorick, in the Elysian Fields* becomes *Unfortunate Sensibility*. At Sael's, in the Strand, in 1793, *Emmeline, the Orphan of the Castle* becomes *Emmeline*. And so on.

A coded message, in a market situation. And the key institution of the market takes the code, and *compresses* it: typically, to a proper name. Libraries couldn't waste space on a catalogue page; they didn't want any confusion between this novel and that; the spine of the book had only room for a few words anyway; and then, readers were getting used to novels, and needed less 'guidance' from titles.⁷

7 If everything was really pushing towards shorter titles—observed Sam Bowles during a discussion of this paper—shouldn't these be 'rewarded' by the cultural ecosystem, and be on average more successful than other types? Yes, they should; and since James Raven had already identified which of the 1,400 novels published between 1770 and 1799 had been reprinted at least five times by 1829 (see his 'Historical Introduction: the Novel Comes of Age', in *The English*

So, the average length decreased, long titles disappeared, and, at the opposite end of the spectrum, titles with only one, two, or three words multiplied rapidly (Figure 9, overleaf): they were 5 per cent in the 1740s and '50s, but by 1800 they were already around 20 to 30 per cent, and had completely traded places with the long titles which had been their predecessors (Figure 10). It's the same thing that happened in advertising a hundred years later, when the detailed descriptions of the nineteenth century were replaced by the evocative oblique brevity of today's ads; *literally* the same thing: title-pages with long summaries of novels were often used as flyers, and

Novel 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles, vol. I, Oxford 2000, p. 40), I compared the lengths of these sixty-five titles to the median for their years—fully expecting them to be significantly shorter. That, however, turned out not to be the case: thirty-two of the titles were indeed shorter than the median, but twenty-nine were longer (at times, *much* longer), and four were exactly the same length.

What these results seem to suggest is that—although a crowded market does exert a strong *negative* pressure against long titles—it remains relatively neutral once a certain length has been reached: it prohibits at one end of the spectrum, but it does not prescribe at the opposite one. Comparative work in other European traditions should provide additional evidence on this matter; meanwhile, and more anecdotally, a look at some canonical British novelists is as inconclusive as the wider bibliographical investigation. If Edgeworth and Austen use much shorter titles than their contemporaries, and Fielding, Smollett, and Burney remain slightly below the median, Richardson and Radcliffe behave in an average way, while Scott and Galt and Dickens often enjoy playing with extremely long titles (which, by their time, are a quaintly obsolete choice): *Tales of My Landlord, Collected and Arranged by Jedediah Cleishbotham, Schoolmaster and Parish-Clerk of Gandercleugh; The Annals of the Parish; or, The Chronicle of Dalmailing; During the Ministry of the Rev. Micah Balwhidder. Written by himself; Dealings with the Firm of Dombey and Son, Wholesale, Retail and for Exportation*.

Now, if neither 'successful' nor 'canonical' novelists took the lead in shortening titles, then, inevitably, *someone else* must have done so: writers who—as we will see in the next section—were neither particularly popular, nor especially good. Perhaps, once the literary system had started moving in a certain direction, some developments were so inevitable that they didn't require any special talent. Or perhaps—as suggested in footnote 11, below—in this case the key variable was not literary, but political.

pasted around to advertise a book. But short titles, as we will see, were not just better titles—they were better *ads*, too.

Titles allow us to see a larger literary field, I said at the beginning of this study; and the first thing we see in this larger field, at this moment in history, is *the force of the market*: how its growth creates a major constraint on the presentation of novels. This of course doesn't mean that all titles gave the same answer to the pressure of the market; but it does mean that they all had to face the same question: How could one shorten a message—without losing information? There was a lot of information in summaries: what happened to it? Was it—gone? reformulated? replaced by something else? I will return to this in a moment; now let me close this first section by acknowledging a limit of this investigation: I began by showing the average length of titles, but I then shifted to very long and very short titles—and I did so, because these trends are much more dramatic than the slow decline of the average, and thus also much easier to talk about. Which is not exactly *wrong* (after all, those trends are real!), but, even aside from a question of completeness—of the 7,000 titles in the study, around 900 are 'long', 1,600 'short', and 4,500 somewhere in between—the focus on extremes misses a decisive aspect of quantitative work: what really counts, here, are not a few major and rapid changes, but many small and slow ones.

But the trouble is, we literary historians don't really know how to think about what is frequent and small and slow; that's what makes it so hard to study the literary field as a whole: we must learn to find meaning in small changes and slow processes—and it's difficult. Especially so, in the case of titles: which are by definition the most public part of a book, hence the most subject to censorship: what we find in titles reflects the 'legitimate irradiation' of existing ideas, wrote Jean-Louis Flandrin, and it's true, titles are so

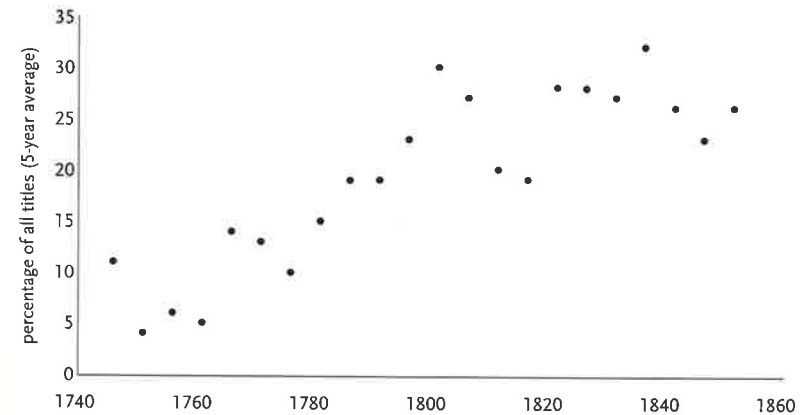


Figure 9: novels with very short titles

'On the twenty-seventh evening, *Nanine*, by M. de Voltaire, was performed. *Nanine*? asked so-called critics when this piece first appeared in 1749: what sort of a title is that? what idea does that give us? Nothing more and nothing less than a title should. A title must be no bill of fare. The less it betrays of the contents, the better it is. It is better for both poet and spectator. The ancients rarely gave to their comedies any other than insignificant titles.' Lessing, *Hamburg Dramaturgy*

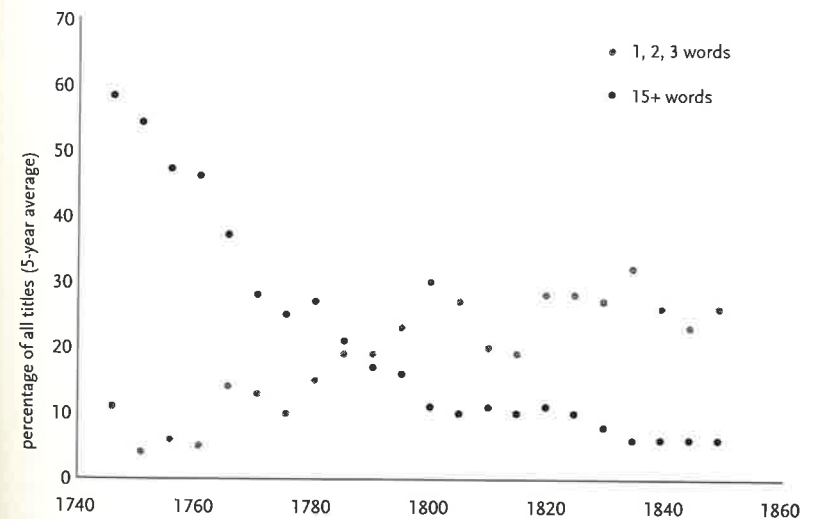


Figure 10: long titles, short titles

'respectable'; and again, how do you make respectable messages interesting . . . ?⁸

II

Very short titles: one, two, or three words. Where the question that interests me is, How can a couple of words stand in for hundreds of pages? What does it mean, that they should do so? For summaries, it's clear: they are scaled-down versions of the whole story. Two words? So, I started looking at these short titles, and found three main clusters within the group: proper names (*Octavia*; *George Barnwell*), which make up around one-third of the total; the article-noun (*The Steam-Boat*; *The Smuggler*) and article-adjective-noun combinations (*The Tuscan Vase*; *The Invisible Gentleman*) just below 30 per cent; and conceptual abstractions (*Fatality*; *Enthusiasm not Religion*) around 10 per cent. 'A large change in size inevitably carries with it a change of form', wrote J. B. S. Haldane, and here one sees how right he was: a title

8 Jean-Louis Flandrin, 'Sentiments et civilisation. Sondage au niveau des titres d'ouvrages', *Annales*, September–October 1965, p. 939. Later I hope to study the 'average title' of these 110 years, taking as a starting point the formula in 'or' (*Pamela*, or *Virtue Rewarded*; *Vensenshon*; or, *Love's Mazes*; *Manfrone*; or, *The One-Handed Monk*). There are over 2,000 such titles in the database, most of which use between three and fifteen words, thus occupying exactly the middle of the field. To get a sense of the morbid diffusion of 'or' in eighteenth-century titles, let me just say that it is the fourth-most-frequent word of the database, following 'the', 'of', and 'a' (and preceding 'and!'); by contrast, in Gaskell's *North and South*, 'or' is the forty-fourth-most-frequent word; in *Our Mutual Friend*, the fifty-fifth.

Aside from quantitative reasons, the formula in 'or' is important because it codified the form of the 'double' title, where the second (on the right of the 'or') is an explication of the first: *Waverley*, that is to say, events of sixty years ago; *Pamela*, a story in which virtue is rewarded. Here, we are clearly beyond the title as summary, though not quite yet in the world of *Belinda* or *Persuasion*: as if the 'or' were a sort of afterthought—a hiccup: maybe one word is not really enough for a title, let's add something else, just to be sure. A compromise formation that coexisted first with summaries, then with short titles, the formula in 'or' thus mediated between explanatory and intuitive strategies; but as readers became more comfortable with allusion, it lost its *raison d'être*. By 1900, it had become a thing of the past.

with twenty words and one with two are not the same creature, one larger and one smaller; they are different animals altogether. Different styles. There is a 'less is more' elegance to short titles—*Persuasion*; *Emma*; *Mansfield Park*—that was unthinkable in summaries; there, the aim was to squeeze as many things as possible into the front page—more is more, as it were—and if the title turned out to be a mess, so be it: *Robinson Crusoe*'s mentioned an episode that doesn't even appear in the novel (*An Account how he was at last as strangely deliver'd by PYRATES: pyrates? what pyrates?*)—it didn't really matter. But a short title is a delicate structure, sensitive to every small change. Consider the article-noun, and article-adjective-noun combinations: similar forms, similar semantic horizon—*The Monk*, 1796; *The New Monk*, 1798—and so at first I assumed that the adjective wouldn't change much: the monk and the new monk: big deal; the adjective would specify the noun, as adjectives do, but no more than that.

And instead, it turns out that the adjective does not specify the semantic field; it transforms it. In the article-noun combination, half of the titles describing a social type evoke an exotic-transgressive field—*The Fakeer*, *The Vampyre*, *The Fire-eater*, *The Pirate*, *The Sabbath-Breaker*, *The Spectre*, *The Rebel*, *The Epicurean*, *The Mussulman*, *The Libertine*, *The Parricide* . . .—and only a small minority evokes the idea of the 'familiar' (wife, brother, father, daughter, etc.). But when an adjective is added to the title, the ratio is exactly reversed (Figure 11, overleaf): fakeers and libertines drop from 50 to 20 per cent, while wives and daughters rise from 16 to 40 per cent: *The Unfashionable Wife*, *The Discarded Daughter*, *The Infidel Father*, *The Rival Brothers*, *The Posthumous Daughter*, *The False Friend*, *The Maniac Father* . . . Without adjectives, we are in a world of adventures; with adjectives, in a destabilized domesticity. The adjective is the only change, but it changes everything. And of course, once you think about it, it makes sense: if all that is in the title is a noun, then that noun must guarantee an interesting story all by itself, and vampires and parricides are a very good choice; but if an adjective is present, then even the most familiar figures can be

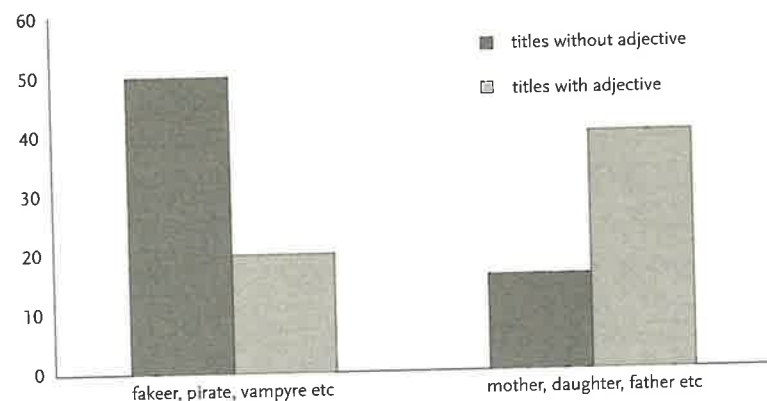


Figure 11: Semantic fields in very short titles: the role of adjectives

estranged into infidel fathers and posthumous daughters. The adjective relocates narrative from substance to accident, as it were. And again, it makes sense: the adjective introduces *predication* within the title, and predication is the germ of storytelling. 'The wife' is a stable quantity; the *unfashionable* wife is a question mark: why is she unfashionable? what does her husband think? her daughters? This is why short titles are so interesting: they are on the border: between two and three words lies the invisible barrier that separates storytelling from—something else, which we'll see in a minute.

Common nouns are frequent, in short titles, but proper names are even more frequent, especially at the turn of the century (Figure 12), when one title in twelve (1786–90), then one in ten (1791–95), then almost one in seven (1796–1800), consists of a proper name, and nothing else: *Emily*; *Henry*; *Georgina*. The growth of the market forced titles to become shorter, and, as we saw in circulating libraries' catalogues, proper names were a great way to do so: one word, and a novel was immediately singled out from the rest.⁹ Singled out, by pointing to

⁹ Needless to say, different names—Evelina, Mary and Moll; Edward, Tom and Dick—evoked very different semantic associations: a great topic for further study.

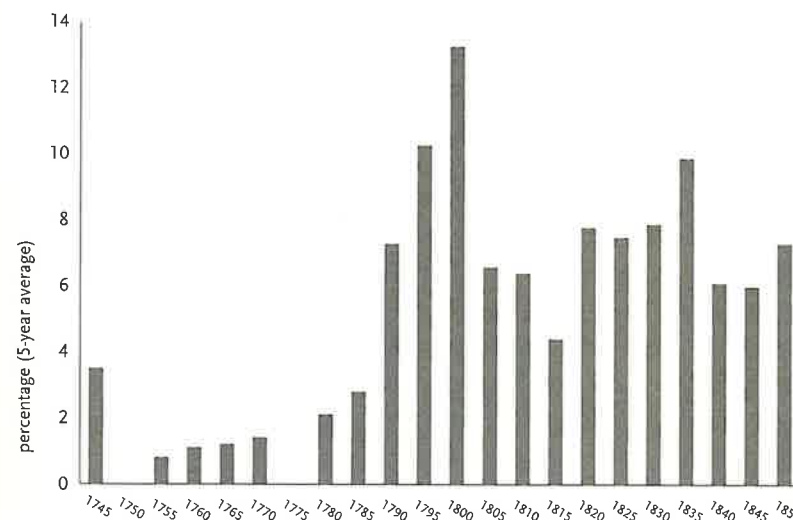


Figure 12: Short titles consisting only of a proper name

'If the Name [...] is a sign, it is a voluminous sign, a sign always pregnant and crammed full of meanings that no use can reduce or flatten [...] It is immune from any kind of selective restriction, and the syntagm in which it is located is a matter of indifference to it. In a certain sense, the name is thus a semantic monstrosity.' Barthes, 'Proust and Names'.

its protagonist; a choice which was not inevitable—of the 'six master-pieces' of the Chinese canon, for instance, none was ever shortened to a proper name, because none had a name in the title to begin with—but which has been typical of European narrative since Greek and medieval times (probably, because in our tradition the central character has always played a greater role). And of course, in the late eighteenth century, protagonist mostly means *female* protagonist (Figures 13 and 14, overleaf): a woman's name, and often just a *first* name (Figures 15 and 16): *Lucy*, *Caroline*, *Belinda*, *Emma* . . . Heroines who lack a last name: a very simple, very crude hint, typical of the British marriage plot (which reaches its apex in these decades): they lack a husband. But the wider field charted in Figures 13 through 16

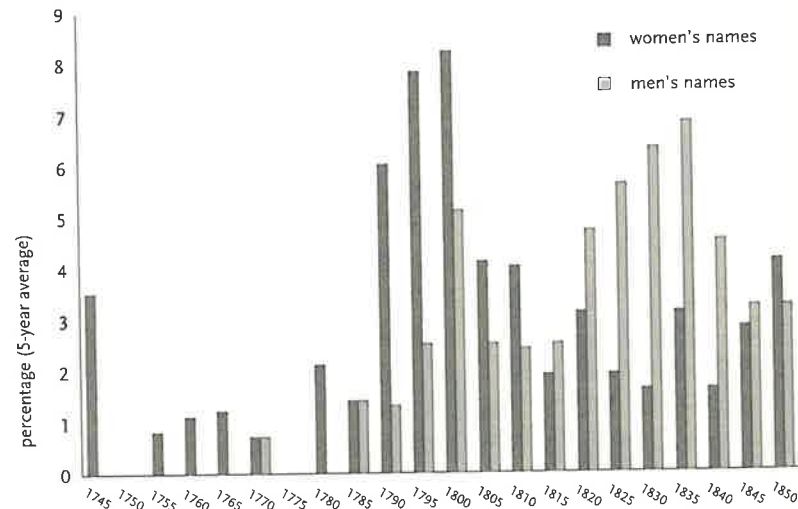


Figure 13: Short titles consisting only of a proper name

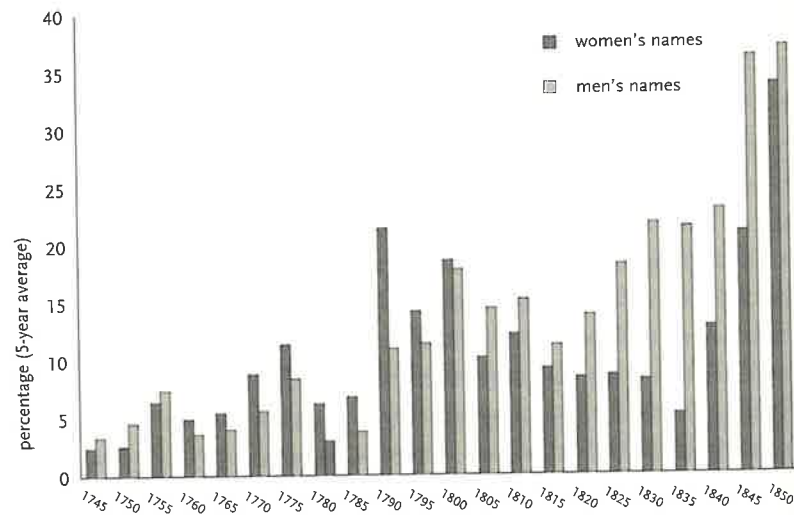


Figure 14: All titles including a proper name

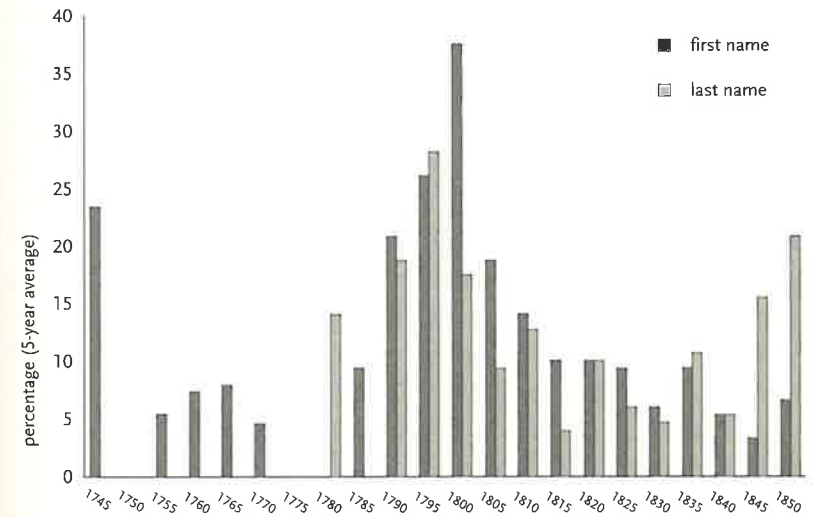


Figure 15: Short titles including only a woman's name

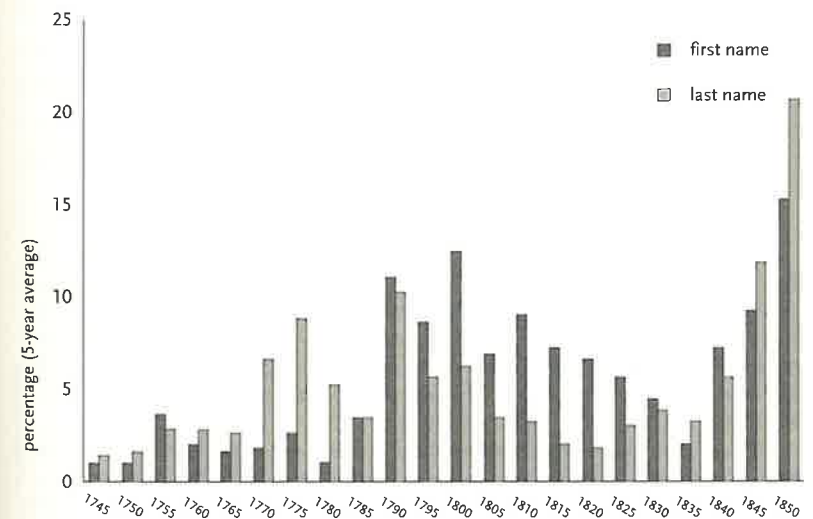


Figure 16: All titles including a woman's proper name

also shows how quickly the gender asymmetry was reversed in the 1820s and '30s, and how frequent the heroine's *last* name could actually be.¹⁰ In both cases, the main changes were almost certainly caused by shifts in the system of genres: the historical novel's rise to prominence after 1815, for instance, with its mostly male heroes;¹¹ or the marriage plot becoming embedded within genres like the *Bildungsroman* and the industrial novel, where the heroine acquired a public life, which was promptly mirrored in titles like *Jane Eyre* or *Mary Barton*. Again, see how much can be done with how little, in short titles: one word, and the image of the heroine rotates 180 degrees: from private, to public. Short titles were a constraint imposed by the market, yes, but the constraint could also be a fantastic opportunity for the literary imagination: the art of allusion, of condensation: the title as trope, ultimately. Odd twist: the market promoting—style.

The market expanded, titles contracted; by 1790, as we have seen, the issue of length had been settled, and didn't really change for at least sixty years. But something else *did* change, between 1790 and 1850, and the last type of title I will discuss in this section—abstractions—will help us to understand what. Abstractions were usually a single word (*Generosity*, *Indiscretion*, *Independence*, *Delusion*), or a conceptual pair (*Liberalism and Prejudice*, *Jesuitism and Methodism*),

10 That female protagonists are more often indicated by their full name than by their first name is one of the surprises of this study. In the century under investigation, the old aristocratic (and often French) form of the proper name—*Rosa de Montmorien*, *Eloise de Montblanc*—found a newer, 'bourgeois' (and British) incarnation in *Alice Lemington*, or *Margaret Graham*; the golden age of the marriage plot, and of titles mentioning first names only, falls in between these two alternative typologies.

11 If one looks at the entire period in question, men's names outnumber women's by about 10 percent, probably because many novelistic subgenres—most travel narratives, nautical tales, later 'Irish' novels, war stories, Newgate novels—were quite unlikely to have female protagonists. That the 'Emma' type of title should immediately come to mind when thinking of proper names in titles, is a sign of its exceptional power of allusion, and of the centrality of the marriage plot in the modern English novel.

and although they were never very frequent, in the first quarter of the century, and especially in the 1820s, they were not insignificant, thanks largely to the tireless Barbara Hofland, who in the five years from 1823 to 1827 published, one after the other, *Integrity*, *Decision*, *Patience*, *Moderation*, *Reflection*, and *Self-Denial*. And as you read these titles you realize that abstractions here really means—ethics. Nineteenth-century ethics; previously, abstractions had often emphasized moral violations (*Disobedience*, *Indiscretion*, *Fatality*, *Retribution*, *False Gratitude*, *The Relapse*, *Conscious Duplicity* . . .), but after 1800 it is the *construction* of the ethical that is highlighted (Figure 17): *Self-Control*, *Conduct*, *Discipline*, *Correction*, *Decision*, *Reformation*. Morality not as purity, but as *work*: one takes one's own self and transforms it, in a process that is both spiritual and pragmatic. Hofland's *Moderation*, wrote the *Monthly Review* in 1825, is 'fabricated . . . to . . . strongly enforce a precept in morals of great practical utility'—and that moral precept fabricated for practical utility is really the dawn of Victorianism.

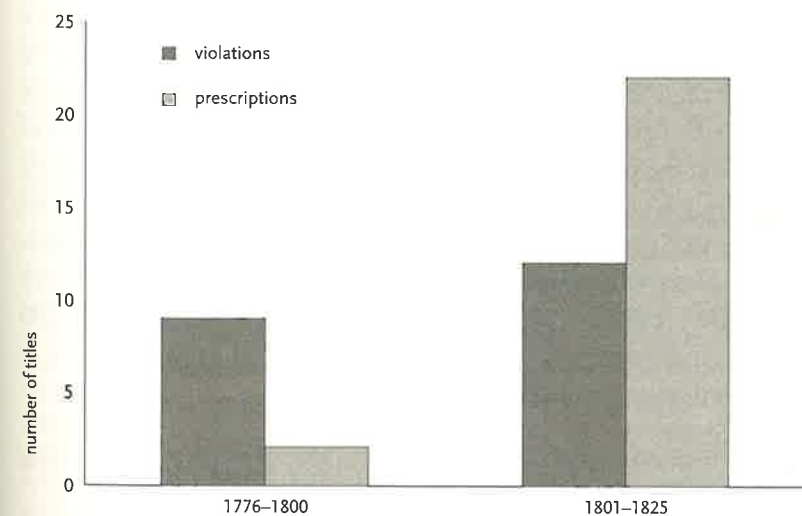


Figure 17: Ethical content of abstractions in short titles

When titles were summaries, they of course used verbs (*The misfortunes in which this young woman has been cruelly involved* etc. etc.); but once summaries disappear, so do verbs (aside from the occasional *Says She to Her Neighbour, What?*), and titles like *Patience* or *Moderation* are the logical endpoint of the process: titles that sound more and more like nominal sentences. A grammatical form that 'places the utterance beyond all temporal or modal localization, and beyond the subjectivity of the speaker', wrote Benveniste in his classic analysis of this type of sentence: beyond subjectivity, beyond temporal localization . . . the telos of nominal sentences is *the abolition of contingency*: 'they don't describe a situation', Benveniste again, 'they posit an absolute'.¹² *Self-Control; Patience; Integrity*: they don't describe a situation, not even in the minimal way of maniac fathers and unfashionable wives; they don't allude to what happens in the novel, or to where and when it takes place; they posit an absolute, and that absolute is of course *the meaning of the novel*. This was the great historical achievement of abstractions: they made titles meaning-ful: nothing *but* meaning, as if the essence of the novel had been distilled and purified of all narrative contingency. And readers, faced with this type of title, have to change their expectations: the first thing they are told about the novel asks them to imagine, not so much a story, but *the point* of the story: the point of the story as a single, unifying concept. And this is important. That titles became short is interesting, yes, but in the end, So what? That by becoming short they adopted a signifying strategy that made readers look for a unity in the narrative structure—this is a perceptual shift which has persisted for 200 years. And mediocre conservative writers did more to make it happen than anyone else.¹³

12 'The Nominal Sentence', in *Problems in General Linguistics*, 1966, Miami 1971, pp. 138, 142.

13 Why them? Perhaps, because there was much in common between the conservative reaction to the French Revolution (for which basic social values had to be preserved from historical transformation), and the type of titles I am

I have discussed abstractions next to proper names, because they both make for very short titles; but, clearly, their relationship to plot is completely different: proper names are a *part* of the story, whereas abstractions are an *interpretation* of it. It would be tempting to say that names have a metonymic relationship to the novel, and abstractions a metaphoric one; but if characters' names (and the—rarer—place names like *Minerva Castle* or *Mansfield Park*) are indeed metonymies of the plot, abstractions are not quite metaphors,¹⁴ and in fact it is curious how *few* metaphors are there, in these 7,000 titles.¹⁵ By the end of the century they are everywhere (*The Belly of Paris, The Doll, Ghosts, The Octopus, Heart of Darkness, The Beast in the Jungle*), so they must have taken root sometime in the third quarter of the nineteenth century, and the glimpses one gets suggest a lot of hesitation on the part of writers: Gaskell shifting at the last minute from *Margaret Hale* to *North and South* (proper name to metaphor); Dickens doing the opposite, from *Nobody's Fault* to *Little Dorrit*. Announcing a story with a metaphor must have seemed strange—and it is strange: if abstractions are removed from the plot, then metaphors are twice removed: interpretations that *require an interpretation*, as it were. But it is precisely this 'difficulty' of metaphors that holds the secret of the title-as-ad.

discussing here (for which fundamental ethical absolutes had to be freed from narrative relativization).

14 Nor are they allegories or personifications: Hofland's *Moderation* is not meant to come alive and be part of a story like its homonym in the 1669 *History of Moderation*; or, *The Life, Death and Resurrection of Moderation: together with her Nativity, Country, Pedigree, Kindred, Character, Friends, and also her Enemies*.

15 It's only at the very end of the period that they begin to appear: *Loss and Gain* (1848), *Rough and Smooth* (1849), *Shadows and Sunshine, Flies in Amber* and *The Swan's Egg* (1850). In general, if the years between 1790 and 1830 see the establishment of metonymies and abstractions, no further novelties seem to emerge between 1830 and 1850: instead of looking for new forms of brevity, writers seem to devote their best energies to the second title, as if that were the key to the problem: *Helen Halsey. A Tale of the Borders. A Romance of Deep Interest; The Slave Captain; A Legend of Liverpool; Goals and Guerdons: Or, The chronicles of a life. By a very old lady; Rebecca and Rowena. A Romance Upon Romance*.

Eighteenth-century summaries told readers a lot of things about the novel, yes; but they never really engaged their intelligence. And instead, by puzzling and challenging readers, metaphors induced them to take *an active interest in the novel* from the very first word. If you are trying to sell a product, that's exactly what you want.

Summaries, adjectives, proper names, nominal sentences, metonymies, metaphors . . . In a minute I will turn to articles (and am thinking of sections on conjunctions and participles). This is a quantitative study: but its units are linguistic and rhetorical. And the reason is simple: for me, formal analysis is the great accomplishment of literary study, and is therefore also what any new approach—quantitative, digital, evolutionary, whatever—must prove itself against: prove that it can do formal analysis, better than we already do. Or at least: equally well, in a different key. Otherwise, what is the point?

III

As the market expands, titles contract; as they do that, they learn to compress meaning; and as they do *that*, they develop special 'signals' to place books in the right market niche. 'Had I, for example, announced in my frontispiece, "Waverley, a Tale of other Days", must not every novel-reader have anticipated a castle scarce less than that of Udolpho . . . a "Sentimental Tale" would . . . have been a sufficient presage of a heroine with a profusion of auburn hair . . . "a Tale of the Times" [of] a dashing sketch of the fashionable world' . . . *Tale of other Days*, *Sentimental Tale*, *Tale of the Times*: that these words would make readers think of specific genres is of course true—and trivial: it's obvious. The code may be in the market, but it remains transparent. And instead, the interesting cases are the opaque ones: where the signal works, and we somehow know what kind of a novel we have in our hands, but we don't know why we know it, because it is all

conveyed by traits that escape our attention; 'subliminal', as we used to say.

Let me illustrate this point with two genres—the so-called anti-jacobin and 'new woman' novels—that are separated by a hundred years:¹⁶ two explicitly ideological genres, which rely heavily on contemporary politics, and whose titles have thus a lot in common—except for one detail. Among anti-jacobin titles, 36 per cent begin with the definite article (*The Banished Man*, *The Medallion*, *The Parisian*, *The Democrat*) and 3 per cent with the indefinite; a result which is perfectly aligned with the rest of the field, since at the time the overall frequencies are 38 and 2 per cent.¹⁷ New woman titles, no; the definite article is obviously still present, in 24 per cent of the cases, but the use of the indefinite leaps from 2, or 3, to 30 per cent of the cases (Figure 18, overleaf). Now, this is odd, not only because it is completely out of scale with anything else I have found, but because in many other ways the conventions of the two genres are quite similar. *The democrat*; *A blue-stocking*: two well-known figures of the contemporary political scene; why is the article different? *The infidel father*; *A hard woman*; the same grammar, the same dissonance between adjective and noun; why is the article different? What do the articles *do*, that they need to be different? An essay by Harald Weinrich offers an answer; for Weinrich, the starting point to understand linguistic categories is always the text, and, since all texts are linear, 'there are always two main directions in which the attention of the reader may be directed': backwards, or forwards: backwards, towards what we already know from the text, and forwards, towards what we don't.¹⁸ And the simplest way to alert

16 See the bibliographies included in M. O. Grenby, *The Anti-Jacobin Novel*, Cambridge 2001, and Ann Ardis, *New Women, New Novels*, New Brunswick 1990.

17 In the *New York Times* best-seller list of November 2008, 38 per cent of the titles began with the definite article, and 6 per cent with the indefinite: not that different from two centuries earlier.

18 Harald Weinrich, 'The Textual Function of the French Article', in Seymour Chatman, ed., *Literary Style: A Symposium*, Oxford 1971, p. 226.

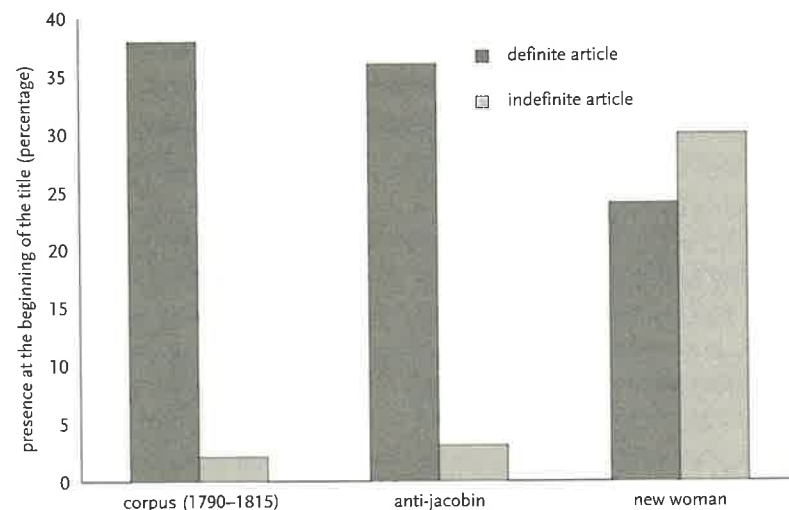


Figure 18: The role of articles as genre signals

the reader's attention is—articles: the definite article announcing a noun as something that we already know (thus directing our attention backwards); and the indefinite suggesting the opposite: Take heed, here comes something that you haven't encountered yet. The first time the wolf appears in *Little Red Riding Hood* it is 'a' wolf; afterwards, 'the' wolf, forever. So: *A Girton Girl*, *A Hard Woman*, *A Mummer's Wife*, *A Domestic Experiment*, *A Daughter of Today*, *A Semi-detached Marriage*: what the article 'says' is that we are encountering all these figures *for the first time*; we think we know what daughters and wives are, but we actually don't, and must understand them afresh. The article announces the novel as a challenge to received knowledge. And instead, 'the' democrat, 'the' parisian, 'the' infidel father . . . we know these people! Anti-jacobin titles don't want to change received ideas, they want to *use* them: the French Revolution has multiplied your enemies—beware.

Here is a modest example of what quantitative stylistics could do: take those units of language that are so frequent we hardly notice them

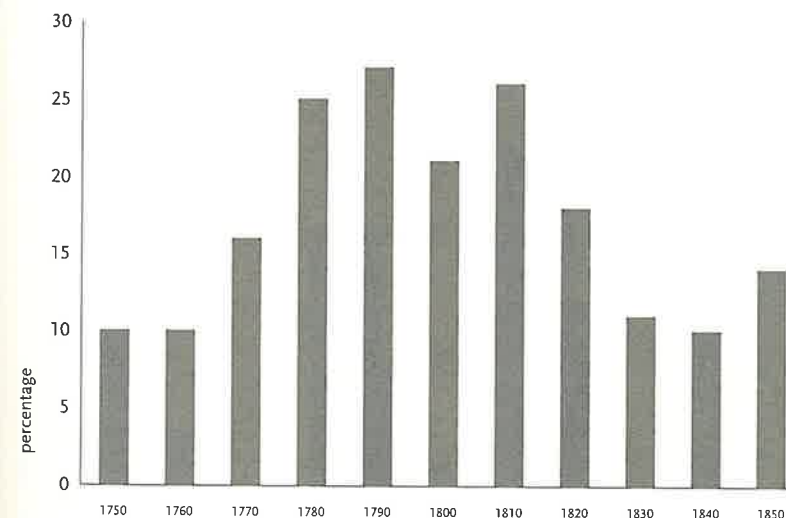


Figure 19: A fortunate formula: 'the x of y'

and show how powerfully they contribute to the construction of meaning.¹⁹ Which is also the point of my last example: a formula that, at first sight, looks as flat and uninspiring as could be: *The Duchess of York*, *The Novice of Corpus Domini*, *The Heir of Montgomery Castle*: let's call it 'the x of y'. As Figure 19 shows, the formula has always been quite frequent in titles, never dropping below 10 per cent of the total; but around 1800 its frequency increases, and if we look more closely at those decades, we find that the surge does not occur evenly everywhere, but is almost entirely concentrated in a single genre, which is the gothic. There, 'the x of y' appears three times more often than in the rest of the corpus (Figure 20), which is too big a difference to be the product of chance, especially since something very similar also occurs within the gothic itself: we all know that the word 'castle' was the shibboleth of the genre's imagination, from *The Castle of*

¹⁹ The model here remains John Burrows's analysis of Austen's characters' styles in *Computation into Criticism*; that he did it twenty years ago, without the help of today's technology, puts us all to shame.

Otranto onwards; well, in gothic titles 'the x of y' occurs *three times more often* than 'castle'.

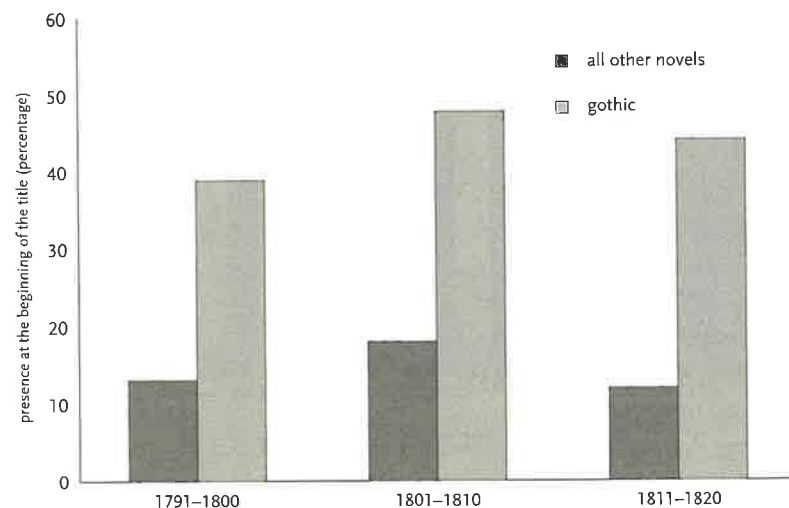


Figure 20: Distribution of 'the x of y', 1791-1820

But why? Castles in gothic titles, it makes sense. The x of y? Here semantics helps; if we look at the 'x' in the formula, we find that 'romance' appears in 7 per cent of the cases (*The Romance of the Pyrenees*), a cluster of genre indicators like mysteries, horrors, secrets, adventures in 13 per cent (*The Horrors of Oakendale Abbey*), personal nouns in 34 per cent (*Emmeline, or the Orphan of the Castle*), and space nouns in 41 per cent of the total (Figure 21): from *The Castle of Otranto* in 1764 to *The Mines of Wieliczka* and *The Rock of Glotzen* a half century later. So, in three-quarters of the cases 'the x of y' specifies an 'x' which is either a person or a space. And when we move from the subject of the formula to its predicate—from the x to the y—what we find is so striking that I don't even need a graph to point it out: *The Romance of the Pyrenees*, *The Horrors of Oakendale Abbey*, *The Orphan of the Castle*, *The Castle of Otranto*, *The Mines of Wieliczka*, *The Rock of Glotzen* . . . in 82 per

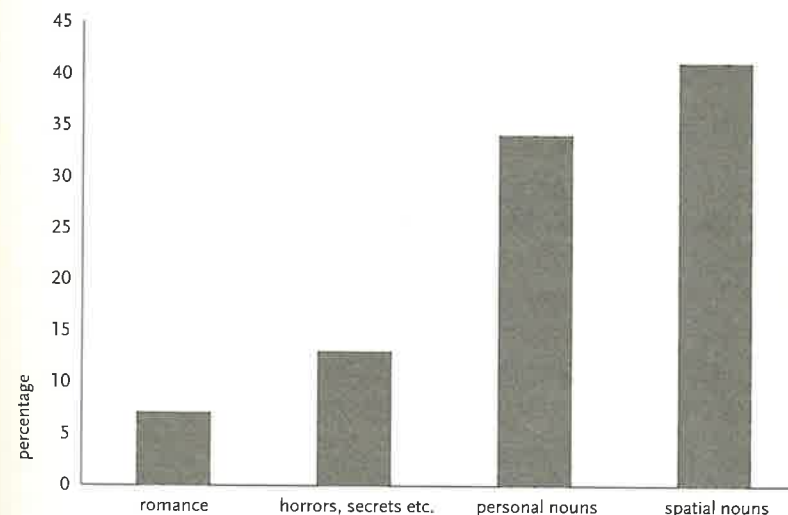


Figure 21: The x in 'the x of y'

cent of the cases, the 'y' is a space: a person defined by a space, or, most frequent of all, a space defined by another space. *The Castle of Otranto*: a spatial noun specified by a place name.

There are many intriguing traits to gothic titles—this is the genre that discovers that readers like villains, for instance, and shamelessly parades them in titles—but space is really the cornerstone of the convention: place names are much more frequent than human proper names; spatial nouns like castle, abbey, forest, cave, etc., show up in 50 per cent of the cases; and there are even other kinds of geographical signals, like *A Sicilian Romance* or *The Danish Massacre*. Nothing is as typical of gothic titles as this fixation with space; and of course this is true not just of titles, but of gothic novels: where space is dark, labyrinthine, cold; it imprisons, it terrifies, it kills . . . 'The x of y' takes this power of space, and activates it at two scales at once: human, and geographical. *The Castle of Otranto*: there is a building; there is a town; they are

both gothic. Escape from the castle, you're still in southern Italy. There is no way out.

'Literature is the fragment of fragments', wrote Goethe in *Wilhelm Meister's Years of Wandering*, the great sad novel of his old age: 'the least part of all that ever happened and was spoken was written down, and of what was written only the least part has survived . . .'. 'Of this history we possess the last volume alone', wrote Darwin in *The Origin of Species*: 'of this volume, only here and there a short chapter has been preserved, and of each page, only here and there a few lines'. There are differences, of course, between the history of nature and that of culture: the 'fossils' of literary evolution are often not lost, but carefully preserved in some great library, like most of those 7,000 novels whose titles I have discussed here; but for the purposes of our knowledge, it's as if they too had crumbled into dust, because we have never really tried to read the entire volume of the literary past. Studying titles is a small step in that direction.

Network Theory, Plot Analysis

'Style, Inc.' had sketched out some hypotheses for quantitative stylistics; *'Network Theory, Plot Analysis'* was an attempt to do the same for plot, thus providing an essential—and still missing—piece to the computational analysis of literature. Once I started working in earnest, however, I soon realized that the tools for a large-scale gathering of data were not (yet) available to me, and the essay retreated from quantification into a qualitative analysis of plot: space and time, network regions, central characters, periphery, and so on. And then a second shift followed: though in a couple of cases—the pages on clustering, and on guanxi in Chinese novels—the essay did engage the categories of network theory, its main claims were independent from its conceptual architecture. Did I really need network theory to discuss *Horatio and the State*, or 'symmetry' in *Dickens*?

No, I didn't need the theory; but I needed the networks. Though *Horatio* is an old fixation of mine, I had never fully understood his role in *Hamlet* until I looked at the play's network structure. The keyword, here, is 'looked'; what I took from network theory was its basic form of visualization: the idea that the temporal flow of a dramatic plot can be turned into a set of two-dimensional signs—vertices (or nodes) and edges—that can be grasped at a single glance. 'We construct, and construct, and yet, intuition remains a good thing', Klee once wrote,