Figures of Fictionality: Keywords of the Eighteenth-Century English Novel

Federica Perazzini, "La Sapienza" University of Rome

Abstract
Recent developments in the polyhedric field of Digital Humanities offer a desirable perspective for corpus-driven literary studies. This is mainly due to both the implementation of tools for the statistical treatment of textual data, as well as the rapid expansion of the Internet in terms of online availability of archives and collections. Notwithstanding a series of contributions highlighting the mutual benefits derived from the combination of computational methods and literary scholarship, traditional criticism seems to ignore the epistemological continuum between qualitative and quantitative approaches to literature, treating them as two separate impermeable realities. In this article I will attempt to reconcile these approaches by presenting an exercise in computational criticism about the linguistic and ideological constructions at the basis of the rising genre of Augustan England: the novel. The aim is to examine the keywords at the core of the extensively theorised modern paradigm of empirical narratives so as to disclose which lexical units may be seen as the distinctive trait of fictionality as well as those which constitute the figure of the novelistic canon. In this way, the article provides an example of how the application of quantitative methods in literary and cultural scholarship can enhance the quality of individual research in the pursuit of the validity of interpretation.

1. The Long Eighteenth Century: A Time of Ontological Instability

As the cultural site for the emergence of scientific rationality and liberal humanism that shaped the foundations of the Western contemporary world, the long eighteenth century is a fascinating object of study for literary scholars. The challenge of a proper historiographical periodisation of such a complex time is generally considered in light of the continuum of political, economic, and ideological changes that took place in Britain from the aftermath of the Glorious Revolution until the end of the Napoleonic wars.[1] These are the years that witnessed the

---

[1] As Black points out in his study *Eighteenth-Century Britain, 1688–1783*, the temporal interval between 1688 and 1783 which he calls "the workings out of the 'Glorious Revolution'" might seem more appropriate in the definition of the

corollary of well-known revolutions in the financial, agricultural, and industrial sectors, which in turn led to the establishment of the English hegemony worldwide. Hobsbawm's *The Age of Revolutions* (1962) investigates the whole process of modernisation that occurred in the nineteenth century introducing the thesis of the "twin revolutions". According to this line of argument, the political and ideological changes resulting from the collapse of the *Ancién Regime* found further reinforcement in the technological and economic transformations brought by the Industrial revolution. Hobsbawm also describes the phenomenon of the agricultural revolution as the actual condition of possibility for the development of self-sustained growth and the industrial revolution itself. Conversely, the notion of financial revolution has been discussed by North and Weingast as the set of economic reforms based on the Dutch financial model imported to Britain along with the enthronement of William of Orange. The need for a centralised system of public debt traded by a private bank— which was to be formed with the creation of the Bank of England in 1694—is undoubtedly one of the most important causes in the fall of James II in 1688 as well as the main trigger for the British transition into modernity (North and Weingast 1989).

In this perspective, the long eighteenth century is also regarded as the period that hosted the radical reconfiguration of the paradigms of intellectual life of the new culture of Enlightenment. Such an era affirmed itself as an optimistic "cornucopia of ideas" tied together by four major discourses: reason, science, humanism, and progress (Pinker 2018: 29). However, as a reflection of the Cartesian doubt and Kantian daring attitude to knowledge acquisition,[2] the culture of Enlightenment appears as a far more ambiguous ideological construction rooted in a profound sense of categorial instability affecting both the socio-economical and generic-ontological spheres (McKeon 1985).

In particular, the term *socio-economical instability* indicates the cultural crisis in the perception of social ranks and the moral state of its members (the so-called *question of virtue*). This is linked to the

---

Long Eighteenth Century. Broader historiographical periodisations are provided by Clark in his book *English Society, 1688–1832* (1985), and O'Gorman in *The Long Eighteenth Century: British Political and Social History 1688–1832* (2016).

[2] Kant summarises the very essence of the Enlightenment movement in the famous motto "Saper aude".

progressive ideology: the historical process of alteration of traditional status groups due to the loss of significance of rank-determined values that occurred in Britain from the late seventeenth century on. In fact, it is right within the cultural experience of the Enlightenment that social prestige began to be measured according to a whole new set of social criteria based on personal merit and virtues as opposed to the previous conception of worth exclusively related to genealogy (McKeon 1987).

On the other hand, the term generic instability refers to the notion of epistemological crisis of the concept of truth and its relationship with the emergence of realistic narrative instances (the so-called *question of truth*). Such a phenomenon finds its philosophical precedent in the Baconian praxis of nominalisation that aimed to restore the "commerce between the Mind and the Things" in order to replace the old axioms of Aristotelian ontological realism (Robertson 2013: 240). The same quest for the proper correspondence between words and things was later reformulated by Locke when addressing the dangers of rhetorical language. He distinguished two types of narratives according to their purposes: the first one—rooted in the meticulous observation of the particular—aims to "inform and improve", whereas the second one—originating from the ungoverned use of imagination—serves the pursuit of "pleasure and delight" (Maioli 2017: 9). In this framework of empiricist fervour, McKeon points out how narratives had to reproduce the state of the world in terms of observable peculiarities without any a priori conclusions in order to gain cognitive value. Such a predominance of the factual dimension over the fictional one began to pervade every aspect of the early eighteenth-century literary production thus generating a whole new craze for historical reports and authentic documentation in a variety of genres, including travel narratives, spiritual or criminal biographies, and—most importantly—in the new literary forms referred to as periodicals and novels.

While only recent critical contributions have acknowledged the relationship of mutual dialogue established among the above-mentioned genres, the role of journalistic writing in the rise of the novel has been a widely discussed topic. Watt, for example, argued that the periodical essay "did much in forming a taste that the novel, too, could cater for" (Watt 1957: 49), whereas Cowan showed how "the style of early modern periodical prose-writings resembled two of the dominant conventions found in the emergent English novel: epistolarity and verbatim accounts

of conversational dialogues between various characters" (Cowan 2005: 66).

Similarly, the problematization of the novel's status as credible fiction has been thoroughly explored from a variety of critical perspectives including stylistics, corpus linguistics, and narratology (McKeon 1987; Davis 1983; Hunter 1990). McKeon provides an aggressive picture of the novelistic genre as "the newcomer that arrives upon a scene already articulated into conventional generic categories and proceeds to cannibalize and incorporate bits of other forms" (McKeon 1987: 11), thus reflecting the novel's dynamic structure and lack of internal rules. In addition to that, Catherine Gallagher's study on the origin of the novel convincingly isolates fictionality as the "new rules" for its determination as a genre (Gallagher 2006: 313). But how are these new rules of novelistic writing configured from a lexical point of view? Is fictionality computationally and linguistically detectable?

By assembling a digital corpus of eighteenth-century works of fiction and periodical essays from 1688 until 1815, the article offers an experiment in computational criticism with the aim of identifying those keywords subtending the discursive practices and social imageries of the novelistic genre. The research takes inspiration from Williams's *Keywords* (1976), which features a preselected list of terms he considered as constitutive of modernity along with an anecdotal commentary on the evolution of their meaning and cultural significance. However, in contrast to Williams's study, which uses a wordlist chosen by the author, the following article pursues an inductive approach with no a priori selection in order to discover if and which words characterise eighteenth-century fictionality. The corpus-assembling procedure along with the methodological choices at the core of the analysis will be discussed in Section 2.

The experiment will then proceed with a definition of the concept of canon in computational terms and its linguistic construction within the corpus of reference. By classifying each of the novels in the corpus into different categorial subsets, namely canonic or non-canonic fiction, an endogenous exploration of the distinctive features of such subsets will be carried out through the calculation of specificity analysis. In particular, Section 3 will be devoted to the description of the different operations performed, while Section 4 will present the conclusions and a possible

interpretation of the results with a brief reflection on the nature of literary evidences and their hermeneutical potential.

2. Method: Corpus Composition and Variables

The assembled corpus of reference at the core of this research is composed of fifteen novels published between 1688 and 1815: of these, nine belong to the canon, while the remaining six can be ascribable to the domain of non-canonic fiction. Additionally, it includes the complete issues of five of the most representative periodicals of the time, such as *The Tatler* or *The Spectator* (see Appendix 1). The selection derives from the bibliographies and the lists of cited works identified as the typical expression of Augustan literature by some of the most authoritative studies in literary criticism (Watt 1957, McKeon 1987, Davis 1983). In terms of sample size, the EC corpus counts almost 4 million words, all of which were analysed through T-LAB: an Italian text-mining software chosen to perform the statistical calculations.[3] Its great potential and operational flexibility comprises a customisable phase of variable attribution which enables the user to assign each of the corpus items a label so as to generate different categorial subsets to compare.

Such a simple act of labelling actually bears a deep ontological significance, since it marks the passage of electronic texts from their concrete status of real objects to the abstract one of measurable models computers can operate on (Moretti 2013). In the case of this research, I proceeded with an attribution scheme based on two basic variables (genre and typology), and four attributes (novels, periodicals, canonic, and non-canonic fiction) which partitioned the EC corpus into the four specific subsets at the core of the computational analysis. Table 1 shows a synthesis of the variable attribution scheme:

Table 1. EC corpus variable attribution scheme and subsets

Variables	Attributes (Subsets)
Genre	Novels / Periodicals
Typology	Canonic Fiction / Non-Canonic Fiction

---

[3] T-LAB 2008, https://tlab.it/it/book.php created by Franco Lancia.

The criteria used to distinguish canonic from non-canonic fiction was human-supervised and based on the publishing history of each text. In particular, using Peter Garside's bibliographical surveys (Garside, Raven, Schöwerling 2000), I reconstructed the number of yearly editions, reprints, and translations into French or German from each book's first appearance on the market up until 1815—a symbolic date identified by Prewitt Brown as the end of the first phase of the novel's stabilisation and canonisation as a genre (Prewitt Brown 1979). Reprints and translations should be regarded as valuable data sources as they can explicitly quantify the appeal of novels to the general audience and therefore help determine their popularity and institutionalisation in the literary market. The need to apply an unbiased and measurable theoretical framework for canonic fiction led me to Moretti's pamphlet *Canon/Archive* (2016) where he draws inspiration from Bourdieu's reflections on taste, *Distinction: A Social Critique of the Judgment of Taste* (1979). According to the French philosopher, mechanisms of canon formation in the cultural field of literature rely on the recognition of the artistic quality and the assignation of "social values" to certain literary works. At the same time, they also imply the acceptance of certain writers and genres as part of the mainstream culture through processes of cultural familiarization.

Subsequent stages of corpus pre-processing include tokenization, disambiguation, and lemmatization, which serve to classify every lexical unit according to specific dictionaries. Once this procedure was completed, I chose the most suitable T-lab functions for the purposes of the experiment which, in my case, were two: keywords and specificity analysis.

The study of keyness has long been a field of interest because it offers a path towards textual investigations capable of combining corpus linguistics and cultural analytics. Indeed, in terms of cultural history, Firth originally considered keywords as "pivotal" words whose distribution and use in context point to cultural values (1935: 40–41). On the other hand, from a sheer corpus linguistic perspective, keywords not only provide an insight into the interpretation of cultural trends, but also to the analysis of parts of speech and typical co-occurrences within certain lexico-semantic contexts. As words whose high frequency—or low frequency—is statistically significant, keywords become focal

elements of phrases and key clusters, thus opening possibilities for the study of collocations, phraseology, and semantic preferences.

Specificity analysis further extends the examination of lexical keyness by allowing the scholar to isolate the words that are typical or exclusive of a selected categorial partition, or corpus subset. This implies an act of endogenous information extraction based on data ascribable to the analysed core-corpus which does not take into account any exogenous resource or external model of reference. An extensive body of research in corpus linguistics related to keywords investigation (Culpeper 2009, Mahlberg and Smith 2010, Hunston 2008) offers a variety of valuable examples of experimental approaches to discourse analysis and stylistics. However, this study differs from such a line of investigation as it is more aligned to Moretti's recent contributions to computational criticism (Moretti 2017): a syncretic modus operandi that rejects the close-analysis of concordances and single textual features in favour of the macro socio-cultural patterns underpinning the literary system.

In this perspective, keywords and specificity analysis constitute two excellent resources for the present experiment which responds to the challenge of mining the lexical configuration of Augustan prose embodied by the new genres of novels and periodicals.

3. A Study of Variables: Figures of the Canon

In this section, I will present the keywords and specificity analysis aiming to explore the features of the novelistic subset of the sample corpus in order to examine the specific traits and word classes that might have determined the development of canon fictionality. As previously stated, the dangers of any biased categorisation were carefully avoided thanks to a labelling process based on measurable and emic criteria, that is to say that only those novels counting at least 5 printed editions per year from the moment of their first appearance until 1815 were ingested and categorised in our corpus as canonic fiction. Table 2 features the genre-specific keywords of the novelistic subset which emerged from the comparison with the periodical subset. The measure of over- or underuse of specific lexical units is given by the value of the Chi-square test: a statistical test that checks if frequency values are different from an expected threshold, where the p-value indicates the probability that the resulted Chi-square was not obtained by pure chance. In this way, the

closest to zero a p-value is, the more the results must be interpreted as statistically significant. As T-LAB applies the Chi-square test to 2 x 2 tables, the threshold value is 6.64 (df = 1; p. 0.01).

Table 2. Lexical specificities of the novelistic subset

Word	Subset	Total	Percent	CHI²
answer	2,057	2,789	73.75	819.50
ask	1,002	1,529	65.53	217.20
assure	739	1,050	70.38	235.94
aunt	287	383	74.93	122.34
bed	554	797	69.51	166.11
believe	1,458	2,388	61.06	196.97
brother	824	1,076	76.58	385.22
chamber	307	425	72.24	111.01
consent	407	537	75.79	182.14
cry	1,772	2,072	85.52	1253.99
daughter	634	993	63.85	116.83
door	625	841	74.32	257.09
earnest	184	231	79.65	100.54
father	1,417	2,269	62.45	225.39
feeling	216	254	85.04	149.70
forgive	340	454	74.89	144.57
girl	384	566	67.84	101.32

Word	Subset	Total	Percent	CHI²
greatly	279	371	75.20	120.77
hastily	228	270	84.44	154.23
hear	1,948	3,449	56.48	131.82
home	815	1,195	68.20	221.31
hope	1,780	2,989	59.55	197.57
house	1,479	2,269	65.18	310.61
instantly	145	158	91.77	128.73
know	4376	7580	57.73	370.17
lady	3,032	4,875	62.19	469.66
leave	1,221	2,000	61.05	164.79
letter	1,460	2,355	62.00	220.61
marry	760	1,253	60.65	97.55
minute	306	431	71.00	101.92
moment	842	1,223	68.85	240.35
money	631	961	65.66	138.31
mother	762	1,188	64.14	144.65
nephew	159	181	87.85	122.87
no_doubt	174	218	79.82	95.83
oblige	119	153	77.78	59.22
poor	1,657	2,171	76.32	764.68

Word	Subset	Total	Percent	CHI²
pray	501	759	66.01	113.26
presently	417	475	87.79	321.65
promise	733	1,150	63.74	133.62
reply	717	931	77.01	343.03
resolve	638	1,025	62.24	99.06
return	1,453	2,180	66.65	347.82
room	878	1,369	64.13	166.58
satisfy	417	628	66.40	97.57
saw	1,193	1,833	65.08	248.17
send	1,176	1,899	61.93	176.23
sister	758	955	79.37	408.83
suppose	669	1,093	61.21	91.99
sure	1,188	1,656	71.74	416.31
tell	3,257	5,233	62.24	507.23
think	4,610	8,569	53.80	172.74
violent	304	428	71.03	101.46
wish	1,314	1,995	65.86	293.60

The data emerging from the novelistic subset suggest the development of a specific discourse of bourgeois domesticity. McKeon (2005) has extensively pointed out the propaedeutical role of narratives of domestication and privatisation for the rise of the novel as a genre, and

the data resulting from this analysis lends support to this view. The lexical configuration of such narratives is characterized by nouns denoting a vast gallery of characters subjected to vertical and horizontal family relations (*father*, *mother*, *daughter*, but also *brother*, *sister*, *cousin*, *aunt* or *nephew*), along with material correlatives of the private sphere conveyed by the prevalence of household interiors (*house*, *home*, *room*, *chamber*, *bed*, *door*). It is the *oikos,* the multi-layered term connoting family, the family's property, and the house itself, that stands at the core of the discursive field of prose fiction, unveiling the profound changes that occurred in the nature of domestic life between seventeenth and eighteenth-century England. Indeed, concrete historical evidences of steadily increasing expenditure on women and children's clothes, hooks, pets, jewellery, and toys, among the gentry and commercial classes are but hints of the conceptual and material re-evaluation of the nuclear family based on a more affectionate model of marriage as well as indulged children (Stone 1979).[4]

As far as the verbal class is concerned, the narrative possibilities delineated by the specificities of the novelistic action are ascribable to the field of mobility (*leave*, *return*), visual perception (*saw, hear, feeling*), sentimentalism (*cry*), and forms of pragmatic resolution (*assure*, *resolve*, *satisfy*). At the same time, the weight of epistolary formulas becomes tangible in verbs such as *send*, *answer* or *reply*, especially when associated with adverbs expressing the sense of immediacy and impending action so typical of the novels in letters (*presently, instantly, hastily).* Moreover, as can be seen in the following examples, verbs such as *oblige* and *consent* strengthen the contractual nature of the narrative transactions which occur in the novel in the form of marriage plots (*marry*, *money*).

---

[4] Stone claims that the shift in ideas about marriage was profoundly influenced by the rise of fiction. Even though the actual extent of the impact of the popularity of the novel in changing the common views on marriage can hardly be determined, literature did certainly reflect the clash between older and younger generations about the matter. The emphasis on self-expression and free will expressed in early eighteenth-century novels such as Daniel Defoe's *Roxana*, or the importance of personal feelings presented in novels such as Samuel Richardson's *Pamela*, were blamed for undermining the custom of arranged marriages while fuelling expectations of romantic love.

Ex. 1: If I call'd, I should be waited upon *instantly*; and so left me to ruminate on my sad Condition, and to read my Letter, which I was not able to do *presently*. After I had a little come to myself, I found it to contain these Words: "Dear PAMELA, THE Passion I have for you, and your Obstinacy, have constrain'd me to act by you in a manner that I know will occasion you great Trouble and Fatigue, both of Mind and Body."[5] (Richardson 2001: 135)

Ex. 2: "They seem very comfortable as they are, and if she were to take any pains to *marry* him, she would probably repent it. Six years hence, if he could meet with a good sort of young woman in the same rank as his own, with a little *money*, it might be very desirable."
"Six years hence! Dear Miss Woodhouse, he would be thirty years old!"
"Well, and that is as early as most men can afford to *marry*, who are not born to an independence. Mr. Martin, I imagine, has his fortune entirely to make—cannot be at all beforehand with the world. Whatever *money* he might come into when his father died, whatever his share of the family property, it is, I dare say, all afloat, all employed in his stock, and so forth." (Austen 2007: 788)

On the other hand, novelistic writing differentiates itself from the periodical essay for its exploration of characters' cognitive dimension (*think* and *know)* and, most importantly, for its conjectural activities evoked in verbs such as *hope, believe, suppose, wish.* By designing the empirical and predictive models of knowledge acquisition typical of the Enlightenment culture, this latter set of words can also be found at the core of the linguistic construction of the novelistic canon. Table 3 illustrates the typological specificities of canonic fiction compared to the non-canonic subset. Here too, the measure of over- or underuse of specific lexical units is given by the Chi-square value.

Table 3. Lexical specificities of the canonic fiction subset

Word	Subset	Total	Percent	CHI²
afraid	306	591	51.78	163.22
aunt	246	383	64.23	246.61
believe	925	1,458	63.44	115.50

---

[5] In this, as in all further quotations, the italics are mine so as to highlight the stated point.

Word	Subset	Total	Percent	CHI²
comfort	312	569	54.83	200.42
cousin	364	472	77.12	559.86
daughter	523	634	52.68	122.06
hope	1,211	1,780	68.03	195.37
instantly	105	145	72.41	112.87
know	2,622	4,376	59.91	130.26
landlord	101	163	61.96	92.16
nephew	141	181	77.90	221.55
niece	112	148	75.68	165.31
no_doubt	142	174	81.60	144.12
presently	251	417	60.19	139.35
sorry	248	470	52.77	140.88
suppose	487	669	72.79	129.78
sure	906	1,188	76.26	574.11
surely	150	266	56.39	104.93
think	2,980	4,610	64.64	165.83
wish	809	1,314	61.56	140.09
opinion	465	649	71.64	188.80

Once the whole series of patronymics of major and minor characters are removed,[6] the narrative horizon of canonic fiction compared to its non-canonic counterpart confirms its articulation in the discourse of familiar horizontal connections. This latter re-emerges in nouns such as *nephew, cousin, niece*, and *aunt,* enriched by a patrimonial connotation of household management evoked in the word *landlord.* Moreover, the crucial importance of *opinion* as a noun reflecting a subject's capacity for appraising people and situations according to a certain interpretation of the surrounding world embodies the new epistemological framework of social imagination of the Enlightenment culture's civil society (Taylor 2004). Likewise, the predominance of a verbal class oriented to characters' cognitive and counterfactual activities can be considered, once again, an example of how canonic fiction exploits a speculative use of knowledge in order to unfold all the potential of the narrative of possible worlds. In particular, as a verb employed in the construction of hypothesis mostly based on trust, *believe* implies a lesser degree of certainty than *know,* thus contributing to maintaining a certain emotional tension in the text by either emphasising the exceptionality of the events narrated, or introducing the counterfactual dimension of characters' opinions and conjectures about events that could be interpreted differently:

> Ex. 3: Betty, who was just returned from her charitable office, answered, she believed he was a gentleman, for she never saw a finer skin in her life.
> "Pox on his skin!" replied Mrs Tow Wouse, "I suppose that is all we are like to have for the reckoning. I desire no such gentlemen should ever call at the Dragon." (Fielding 2014: 54)

So predictions, but also, and especially, expectations appear to be the distinctive markers of fictionality. In fact, the overuse of the verbs *hope* and *wish* (as well as *hope* and *wish* in the noun class) in the novelistic subset suggests that only those works displaying a certain tension toward the future—internally developed as propositional attitudes of the characters towards desires—ended up constituting the canon.

Of course, the expression of hope is not an eighteenth-century discovery, and in the works of the past it is significantly more connected

---

[6] Characters' names have been deliberately ignored and removed from the list for the purpose of this analysis.

to an affective rather than cognitive dimension. However, it is only with the emergence of the new probabilistic matrix of modern empirical thought that novelists seemed to understand more thoroughly the counterfactual potential of the rational desire. In this perspective, as illustrated by example 4, *wish* is much higher in the canonic novel as an anticipatory and programmatic statement of purpose.

> Ex. 4: Now I wanted nothing but a boat to furnish myself with many things which I foresaw would be very necessary to me. It was in vain to sit still and wish for what was not to be had; and this extremity roused my application. . . . So I went to work, and with a carpenter's saw I cut a spare topmast into three lengths, and added them to my raft, with a great deal of labour and pains. But the hope of furnishing myself with necessaries encouraged me to go beyond what I should have been able to have done upon another occasion. (Defoe 2001: 55)

Robinson's hope is basically a reflection of the absence of some objects upon which his well-being depends; such an acknowledgment sets the conditions for the material attempt to procure them while the counterfactual nature of his desire, its articulation in a future prediction that motivates action, becomes explicit the moment Robinson's desire "roused [his] application". In this way, associating the verb *wish* to the practical application such a desire implies, leads to the construction of a narrative program based on the achievement of a specific goal after a series of logically identifiable actions.

As far as the study of modifiers is concerned, the adverbs of certainty *sure*, *surely*, and the locution *no-doubt,* constitute important indicators of epistemic and evidential modality aiming to signal when an utterance presents a stronger argument than an alternative one. Besides revealing the dialogical potential of the novelistic genre, these structures disclose the cognitive complexity of novel fictionality where characters express their assumptions on the pondered outcomes of specific events. So, for example, this occurs when after collecting enough data and drawing the possible conclusions, Pamela imagines the ill-fate that awaits her:

> Ex. 5: "For to be sure, now it is too plain, that all your cautions were well grounded. O my dear mother! I am miserable, really miserable! — But still, do not be frightened, I'm honest! — God, of his goodness, keep me so!" (Richardson 2001: 16).

Or also when Mr Lovelace in Richardson's *Clarissa* (1748) constructs a possible narrative scenario based on his projection on Clarissa's behaviour:

> Ex. 6: "Now, Jack, what can a man make of all this? My intelligence as to the continuance of her family's implacableness is not to be doubted; and yet when I read her letter, what can one say? Surely, the dear little rogue will not lie! I never knew her dispense with her word, but once." (Richardson 2014: 606).

By embodying the "what if" reasoning condition through which the narrator or the characters evaluate situations and ponder the future, counterfactuality carries out the models of knowledge of the Enlightenment culture while unveiling its connection with the cognitive systems of the empirical world. In this respect, the specificity of the novelistic genre emerges in its dual character of fictional and ideological construction of cognitive narratives: subjectivised and emotional representations of private experiences realised through the act of knowing and imagining possible worlds.

4. Conclusions: Bringing Quality to the Quantitative

As illustrated so far, this article represents an innovative contribution to the fields of literary studies and computational criticism in its attempt to translate the theories and practices of textual hermeneutics into empirically verifiable assumptions. By constituting a replicable practice in modelling literary history, it offers the chance to make the qualitative quantifiable while bringing quality to the quantitative.

Inscribed in the methodological framework of keywords extraction and specificity analysis, the analyses here proposed have given an account of the linguistic units constitutive of the ideological structure of Enlightenment culture. In particular, as a cross-discursive practice of political, economic, social, and institutional statements on which we still depend in large part today, Enlightenment culture stands as the cradle that hosted the formation of the modern subject. Whereas its philosophical roots are to be found in the domain of empiricist perception, the most distinctive feature of Enlightenment culture is what Gallagher identifies as the peculiar experience of cognitive provisionality. Deeply intertwined with the practice of reading fiction, cognitive provisionality corresponds to the detached and ironic disposition toward incredulity that soon

became the necessary condition of modern civil life (Gallagher 2006). A general public's competence in investing in temporary credit is indeed the basic pre-requisite that the rising capitalist society implied and exported in a variety of human relations and intercourses. For example, as entrepreneurs and insurers had long learned to employ a certain degree of imaginative power to ponder the risks and collateral damages of their investments, so also women involved in the new market of companionate marriages would naturally speculate on how it would be like to love a particular man before committing themselves. In this way, novels became the literary form designated to describe the ascent of ironic credulity—later formalised by Coleridge as "willing suspension of disbelief" (*Biografia Literaria,* 1907, chapter XIV)—at the basis of the modern experience.

At the end of this long exercise in computational criticism, a journey through statistical calculations, linguistic evidences, and literary history, the question about how much quantitative methods in literary scholarship can enhance the quality of individual research probably remains. Does computational criticism really represent the revolution that would lead the humanities out of their crisis by bringing new knowledge and extraordinary results in the field? Maybe not, or simply not yet. Computational criticism is not *the* revolution but certainly requires one: a revolution that is not about new operational protocols or tools but one related to the nature of evidence in literary criticism. Indeed, as Stephen Ramsay points out:

> Literary criticism operates within a hermeneutical framework in which the specifically scientific meaning of fact, metric, verification, and evidence simply do not apply. […] "Evidence" stands as a metaphor for the delicate building blocks of rhetorical persuasion. We "measure" (as in prosody) only to establish webs of interrelation and influence. "Verification" occurs in a social community of scholars whose agreement or disagreement is almost never put forth without qualification. […] The scientist is right to say that the plural of anecdote is not data, but in literary criticism, an abundance of anecdote is precisely what allows discussion and debate to move forward. (Ramsay 2011: 7–8)

References
Austen, Jane. 2007. *The Complete Novels of Jane Austen*. Ware: Wordsworth Editions Limited.

Black, Jeremy. 2001. *Eighteenth-Century Britain, 1688–1783*. New York: Palgrave Macmillan.

Bourdieu, Pierre. 1979. *Distinction: A Social Critique of the Judgement of Taste* (transl. 1984). Harvard University Press.

Clark, Jonathan. 1985. *English Society, 1688–1832.* Cambridge University Press.

Coleridge, Samuel Taylor, and John Shawcross. 1907. *Biographia Literaria*. Oxford: Clarendon Press.

Cowan, Beverly. 2005. *The Social Life of Coffee*. New Haven & London: Yale University Press.

Culpeper, Jonathan. 2009. "Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet." *International Journal of Corpus Linguistics* 14: 1. pp. 29–59.

Davis, Lennard J. 1996. *Factual Fictions: The Origins of the English Novel*. University of Pennsylvania Press.

Defoe, Daniel. 2001. *Life and Adventures of Robinson Crusoe*. London: Penguin Classics.

Hobsbawm, Eric. 1962. *The Age of Revolutions*. London: Weidenfeld & Nicolson.

Hunston, Susan. 2008. "Starting with the small words: patterns, lexis and semantic sequences." *International Journal of Corpus Linguistics* 13: 3. pp. 271–295.

Hunter, J. Paul. 1990. *Before Novels: The Cultural Contexts of Eighteenth-Century English Fiction*. New York: W.W. Norton and Company.

Italia, Iona. 2005. *The Rise of Literary Journalism in the Eighteenth Century.* New York: Routledge.

Fielding, Henry. 2014. *Joseph Andrews*. Mineola: Dover Thrift Editions

Gallagher, Catherine. 2006. "The Rise of Fictionality" in *The Novel: History, Geography and Culture*, Vol. 1. Franco Moretti (ed.). Princeton University Press.

Garside, Peter, Raven, Schöwerling. 2000. *The English Novel 1770–1829, A Bibliographical Survey of Prose Fiction Published in the British Isles*. Oxford University Press.

Maioli, Roger. 2017. *Empiricism and the Early Theory of the Novel, Fielding to Austen*. New York: Palgrave Macmillan.

Mahlberg, Michaela and Catherine Smith. 2010. "Corpus approaches to prose fiction: civility and body language in Pride and Prejudice" in

Language and Style. Beatrix Busse and Dan McIntyre (eds.). Basingstoke: Palgrave, pp. 449–467.

McKeon, Michael. 1985. "Generic transformation and social change: rethinking the Rise of the Novel." *Cultural Critique* (1), 382–399.

McKeon, Michael. 1987. *The Origins of the English Novel: 1600–1740*. Baltimore: The John Hopkins University Press.

McKeon, Michael. 2005. *The Secret History of Domesticity*. Baltimore: The John Hopkins University Press.

Moretti, Franco. 2013. "Operationalizing, Or, the Function of Measurement in Literary Theory." *New Left Review*, 84, Nov./Dec.

Moretti, Franco. 2016. *Canon Archive. Large-scale Dynamics in the Literary Field*, Pamphlets of the Stanford Literary Lab.

Moretti, Franco. 2017. *Canon Archive. Studies in Quantitative Formalism*, New York: n+1 Foundation.

North, Douglass C. and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History*, 49, pp. 803–832.

O'Gorman, Frank. 2016. *The Long Eighteenth Century: British Political and Social History 1688–1832*. London: Bloomsbury Academic.

Pinker, Stephen. 2018. *Enlightenment Now. The Case for Reason, Science, Humanism, and Progress.* London: Penguin Random House.

Prewitt Brown, Julia. 1979. *Austen's Novels: Social Change and Literary Form*. Cambridge Mass.: Harvard University Press.

Ramsay, Stephen. 2011. Algorithmic Criticism, *Companion of Digital Literary Studies,* http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405148641/9781405148641.xml&chunk.id=ss1-6-7 (last consulted Jul 2019).

Richardson, Samuel. 2001. *Pamela, or, Virtue Rewarded*. Oxford: Oxford World Classics.

Richardson, Samuel. 2014. *Clarissa Harlowe, or The History of a Young Lady*. New York: Simon and Schuster.

Robertson, John M. 2013. *The Philosophical Works of Francis Bacon.* London: Routledge.

Stone, Lawrence. 1979. *The Family, Sex and Marriage in England 1500–1800* (abridged ed.). London: Penguin.

T-LAB 2008. https://tlab.it/it/book.php.

Taylor, Charles. 2004. *Modern Social Imaginaries*. Durham and London: Duke University Press.

Watt, Ian. 1957. *The Rise of the Novel*. Berkeley and Los Angeles: University of California Press.

Williams, Raymond. 1976. *Keywords, A Vocabulary of Culture and Society*. New York: Oxford University Press.

Appendix 1: Novel Corpus Subset:
AUBIN, Penelope. 1723. *The Life of Charlotta Du Pont*.
AUSTEN, Jane. 1815. *Emma*.
BEHN, Aphra. 1688. *Oronooko: or, the Royal Slave*.
CHETWOOD, William. 1726. *The Voyages, Dangerous Adventures, and imminent Escapes of Captain Richard Falconer*.
DEFOE, Daniel. 1719. *The Life and Strange Surprizing Adventures of Robinson Crusoe*.
DEFOE, Daniel. 1722. *The Fortunes and Misfortunes of Moll Flanders*.
FANNY BURNEY, Frances. 1778. *Evelina*.
FIELDING, Henry. 1742. *The History of the Adventures of Joseph Andrews and of his Friend Mr. Abraham Adams*.
FIELDING, Henry. 1749. *The History of Tom Jones, a Foundling*.
HAYWOOD, Eliza. 1719. *Love in Excess*.
HAYWOOD, Eliza. 1723. *Idalia or the Unfortunate Mistress*.
HAYWOOD, Eliza. 1751. *The History of Miss Betsy Thoughtless*.
JOHNSON, Samuel. 1759. *The History of Rasselas, Prince of Abissinia*.
RICHARDSON, Samuel. 1748. *Clarissa, or, the History of a Young Lady*.
RICHARDSON, Samuel. 1749. *Pamela, or Virtue Rewarded*.

Periodical Corpus Subset:
The Tatler, 1709–11
The Spectator, 1711–12
The Female Spectator, 1744–46
The Rambler, 1750–52
The Covent Garden Journal, 1752