

# THE INTERNATIONAL JOURNAL OF HUMANITIES & SOCIAL STUDIES

## Geography of a Stereotype: A Computational Study on the Italian Presence in the British Nineteenth Century Novel

Dr. Federica Perazzini

Adjunct Professor, Department of History, Culture, and Religious Studies,  
Sapienza University of Rome, Italy

### Abstract:

*Often chosen as the privileged setting for plays, poems, and novels, Italy has certainly been at the core of the English literary imagery from the Middle Ages to Modernism. However, more than an actual geographical space, the presence of Italian locations within the corpus of British literature configures itself as a distinctive discursive practice disclosing a variety of literary possibilities; an inexhaustible basin of themes and conventions capable of putting the English writer in what Roland Barthes called a situation d'écriture, a writing situation (Barthes 1970). But what are the features at the basis of the construction of Italy as a writing situation? What is the relationship between the use of certain Italian locations and the development of specific sub-genres? Using innovative tools of the computer-based macro analysis such as topic modelling and words-cohort correlation, in what follows I will give evidence of the occurrences and the transformations of the Italian stereotype throughout the British novel from 1780 to 1890. The aim of the project is to investigate the changes in the relative representation of different geographical areas in the fictional space of the novel thus testing, with new empirical tools and evidences, the assumption that "a new space . . . gives rise to a new form" (Moretti 1998: 197)*

**Keywords:** *Literary geography, computational analysis, Italy, British Novel*

### 1. Introduction: Italy as a "Situation D'écriture"

As a sort of perennial imaginative necessity to the northerner, the study of the role of the Bel Paese in the shaping of the English cultural imagery has always been at the core of both Italian and British scholarly tradition. In particular, Piero Boitani's *Chaucer and the Italian Trecento*, the essays collection by Michele Marrapodi titled *Shakespeare's Italy: Functions of Italian locations in Renaissance drama*, along with Massimiliano Demata's *Gothic Italy* are but few distinguished contributions about the matter. Likewise, Stephen Hebron, David Laven, Charles Brad, have highlighted the symbiotic relationship between the British cultural production and the perpetuation of over-simplifications and stereotypes about Italy since the Renaissance as well as the expansion of tourism in the eighteenth and nineteenth centuries.

Interestingly enough, what seems to powerfully emerge from such studies is the common idea of an almost entirely fictionalized nature of the Italian setting as perceived and used throughout British literature. In fact, more than an actual geographical space, Italy configures itself as a forge of literary possibilities wherein specific areas appear as embedded in an abundance of objects, images, and symbols that soon crystallize into forms and narratological conventions. This is what Roland Barthes—referring to Japan in his book *Empire of Signs*—defines as a situation d'écriture; the tangible correlative of the condition of estrangement and otherness at the root of the dynamics of aesthetic creation where the foreign country appears in the eyes of the writer as "a reserve of features whose manipulation — whose invented interplay — allows me to 'entertain' the idea of an unheard-of symbolic system, one altogether detached from our own." (Barthes, 1970: 3). But how could Italy as a writing situation be constructed? What were the tropes and the semantic traits connected to the Italian identity and how did these ascend, evolve and decline within the British literary system? Using the innovative tools of the computer-based macro analysis, in what follows I will give evidence of the occurrences and the transformations of the Italian stereotype throughout the British novel from 1780 to 1890. The article will be articulated in 3 parts: first I will present some of the theoretical premises at the basis of the methodology applied in this study. Then, I will proceed with a brief introduction about the origins of the project, describe the phases of the computational analysis, and present the results and conclusions on the possible meanings of such results.

### Methodology

#### 2.1. Theory and Practice of Computer-Based Macro Analysis

In his work on the evolution of literary forms, *Maps Graphs and Trees*, Franco Moretti introduces the pioneering concepts of "distant-reading". This can be resumed as the need of a new bird's eye perspective for the study of literary

phenomena as an alternative to the “close reading” of few canonical masterpieces. Nowadays, in this age of digital revolution we live in, the idea of a more comprehensive study of the literary history has been gaining concreteness thanks to the computer-based macro analysis. Computer-based macro analysis, in fact, is an experimental methodology that applies the theories and the statistical tools of computational linguistics to the analysis of huge corpora of digitized literary texts. Allowing the quantitative investigation of thousands of books at once, computer based-macro analysis expands the object of study of literary criticism as never before getting us one step closer to the formalist dream of a vision d’ensemble of the literary system as a whole. In the sole territory of the United States, there are currently more than 50 centers for Digital Humanities working on projects concerning the computer-based literary macro-analysis. One of these, the one I recently visited and cooperated with, is the Stanford Literary Lab: where this project and article began.

In fact, born as a final paper for Matthew Jockers’ class “Literature and Digital Libraries”, this study is part of a larger on-going experiment of the Stanford Literary Lab about the use of geographical locations in the nineteenth century British-Irish and American novel. The aim of the project is to investigate the changes in the relative representation of different areas in the fictional space of the novel thus testing, with new empirical tools and evidences, the relationship between geographical locations and the development of specific literary genres (Moretti 1998: 197).

### 3. Analysis and Discussion of the Experiment

#### 3.1. Preliminary Phases

I began the actual experiment with the team of the Literary Lab by developing a tool called Stanford’s Name Identity Recognition (NER). NER, also known as entity identification, is an application that allows the unsupervised extraction and classification of single units of text according to a set of predefined categories. These categories can be proper nouns, organizations, quantities, percentages and, in the specific case of this research, locations. I therefore ran the NER on the Stanford Literary Lab digital corpus of eighteenth and nineteenth century novels. At the end of the process, all the 3,603 texts of the database had been tagged with “location” identifiers and subsequently, using a specific PHP script, a matrix could be extracted so to show for each of the texts of the corpus the exact number of occurrences of each “Location” name. After a series of other preliminary operations not necessary here, I calculated the mean of the presence of the single locations across the decades and normalized the data according to the expected rate of occurrences. Ultimately, I used the open-source program for statistical analysis R to convert the results into a more legible graphical output. Here are the charts about the occurrence of Italian locations in the 3,603 nineteenth century British Novels:

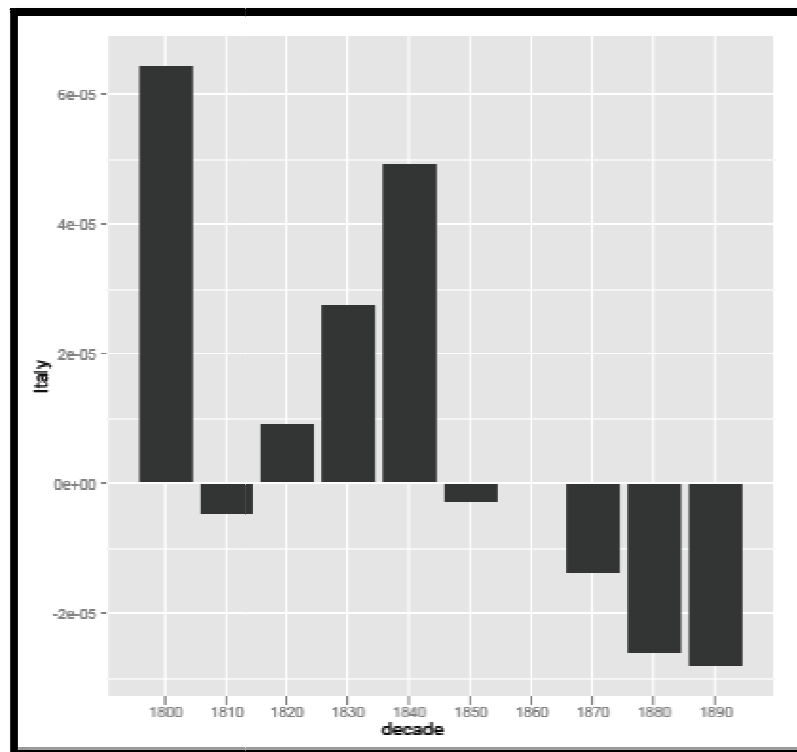


Figure 1: Corpus Presence of “Italy”

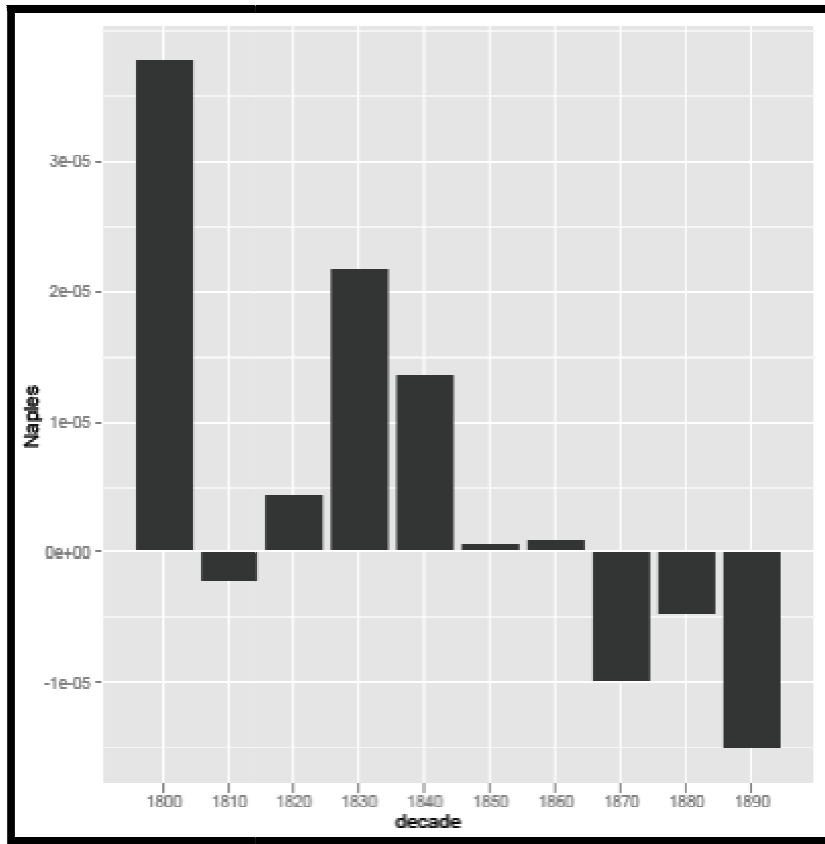


Figure 2: Corpus Presence of "Naples"

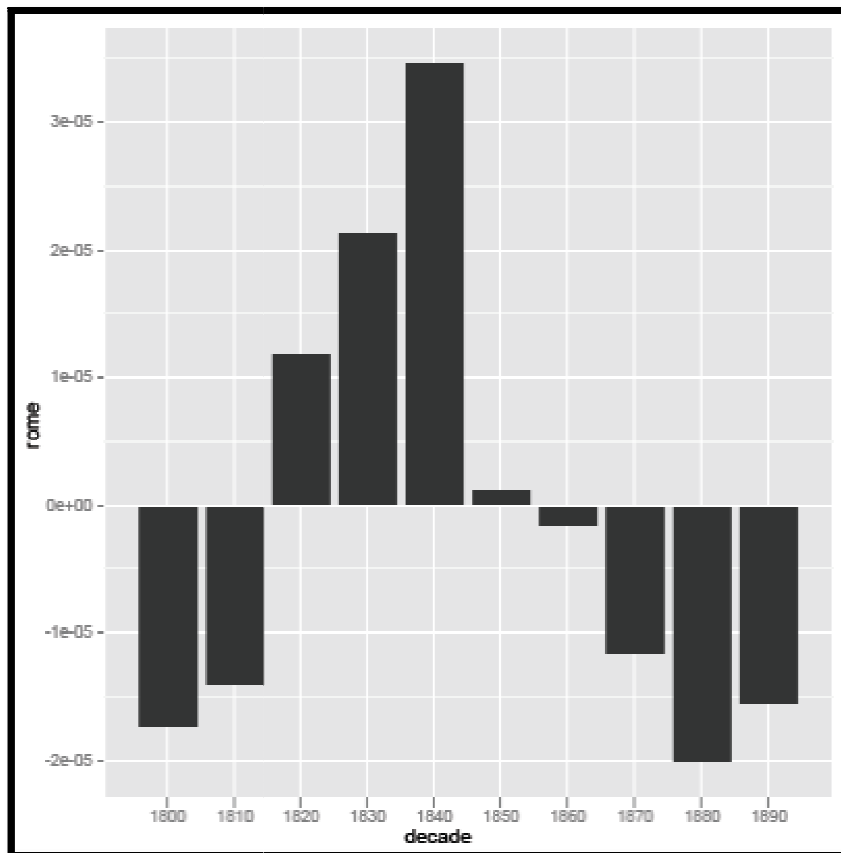


Figure 3: Corpus Presence of "Rome"

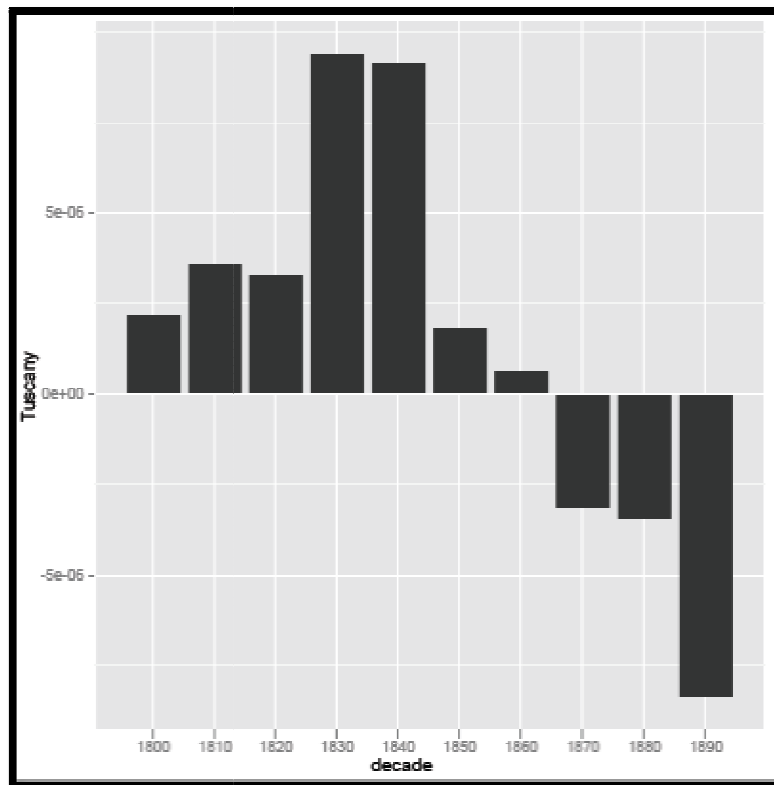


Figure 4: Corpus Presence of "Tuscany"

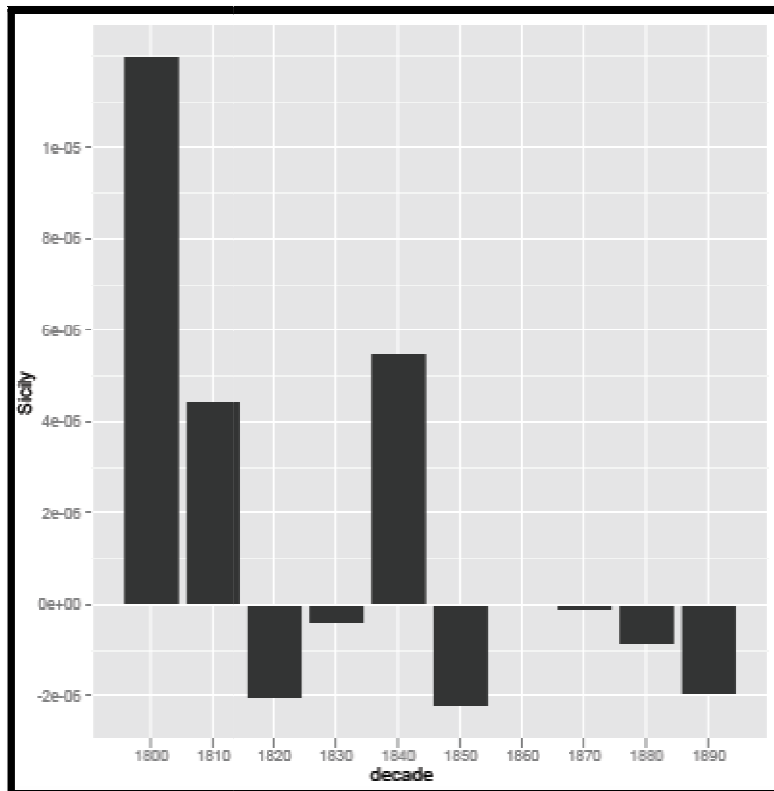


Figure 5: Corpus Presence of "Sicily"

As these graphs suggest, in terms of mere token frequency, the Italian presence in the British novel can be tracked in 3 different directions: Italy as a country, Italian cities and Italian regions. Though these three categories do not follow a homogeneous diachronical behavior, all of them seem to have in common the fact of losing their statistical significance after the 1840s.

On the other hand, in terms of highest scores, if there was a period in which the use of the Italian setting took a center stage this was certainly between the end of the XVIII and the first decade of the XIX century. From the 1780s to the 1800s, in fact, the Bel Paese imposes itself in the British imagery both in terms of a generalized idea of nation, with a regular occurrence of “Italy” throughout the whole interval, as well as a more localized set of locations like “Venice”, “Naples”, “Palermo” and “Sicily”. In addition to that, as the century progresses, a new peak is recorded between the 1830s and 1840s when the city of “Rome” and the Tuscan region make their tardy appearance in relatively consistent percentages.

Quite predictably, the early years of triumph of the Italian setting, and in particular Venice as well as other southern Italy locations, coincide with the decades of maximum effluence of the Gothic Novel. What about “Rome” and “Florence” though? Is their late, circumscribed peak in the 1830s and 1840s part of the same gothic discourse or can it be seen as an index of a shift in the Italian semantic identity?

In order to answer the above questions, I moved on with phase two of my experiment aiming to explore, always from a computational perspective, the semantic behaviour of the following sets of locations: Italy as a country; Southern Italy area; Rome; Florence and Tuscany areas.

### 3.2. Mining Semantic Fields

As regards the methodology, this further phase of the project required the creation of a customised script named “words-cohort correlation”. “The words-cohort correlation”, better known as “Correlator”, is a tool developed by Ryan Heuser capable of generating unsupervised semantic fields by grouping together words considered to share a similar frequency trend throughout the historical range of the corpus. Originally applied to the analyze the semantic behavior of the “abstract values” during the nineteenth century (Heuser, Le-Khac, 2012), this method allowed me to isolate the semantic differentiating factor within each of the four groups of Italian locations. Here is the semantic field generated by the keyword “Italy”. The positive numbers in the brackets represent the score of standard deviation above the average each of the correlated words records:

retribution (+0.95212)	dressing-room (+0.83568)	committed (+0.85166)
mysterious (+0.95135)	rushed (+0.83531)	paces(+0.84876)
vice (+0.93082)	reverie (+0.83436)	branches (+0.84869)
haunt (+0.90634)	lightning (+0.83427)	hinges (+0.84812)
traces (+0.90544)	valet (+0.83389)	dense (+0.84745)
bride (+0.90355)	graceful (+0.83217)	hearts (+0.84694)
expended (+0.90317)	orphans (+0.83214)	enveloped (+0.84497)
prison (+0.89489)	steps (+0.82999)	require (+0.84465)
ceremonious (+0.88746)	underwent (+0.82891)	coaches (+0.84271)
pinning (+0.88617)	villain (+0.82841)	elapsed (+0.84163)
chateau (+0.88377)	beauty (+0.82756)	tainted (+0.83950)
struggles (+0.88148)	regained (+0.82678)	hollow (+0.83574)
tenants (+0.88099)	regions (+0.82603)	expiration (+0.81264)
enables (+0.87999)	lofty (+0.82583)	gazed (+0.80937)
roof (+0.87934)	fulfill (+0.82531)	debtor (+0.80879)
creditor (+0.87531)	Proceeding (+0.82362)	monsters (+0.80870)
fearfully (+0.87190)	voluptuous (+0.82300)	covering (+0.80809)
riveted (+0.87035)	glistened (+0.82295)	stupor (+0.80784)
depicted (+0.86593)	murdered (+0.82097)	ringlets (+0.80764)
estates (+0.86338)	soothe (+0.82054)	plunged (+0.80754)
deceit (+0.86242)	cavalier (+0.81873)	ruined (+0.80514)
inquiries (+0.86193)	death-like (+0.81734)	fearful (+0.80444)
death-bed (+0.85857)	indications (+0.81719)	arrive (+0.85203)
refuge (+0.85439)	sire (+0.81572)	myriads (+0.81416)
treacherous (+0.85305)	plebeian (+0.81418)	

Table 1

Similarly, to what happens to Italy as a country, the use of southern locations such as “Naples”, “Palermo” and “Sicily”, appears to be connected to the same semantic domain:

artifices (+0.89130)	obligations (+0.80676)	moments (+0.77348)	philanthropy (+0.82424)
dungeons (+0.88229)	gloomy (+0.80638)	rewarded (+0.76880)	horrors (+0.82384)
assiduously (+0.88071)	deceit (+0.80455)	confine (+0.76777)	medicines (+0.82373)
inquisitor (+0.87937)	adieu (+0.80315)	beholding (+0.76625)	endeavor (+0.82347)
dreadful (+0.87777)	crimes (+0.80274)	tranquility (+0.76568)	hero (+0.82274)
ignominy (+0.87315)	terrify (+0.80259)	vice (+0.76563)	misfortunes (+0.82245)
recurred (+0.87191)	signor (+0.80255)	captivated (+0.76497)	don (+0.82160)
signora (+0.87131)	fraternal (+0.80194)	terror (+0.76456)	deluded (+0.82148)
convent (+0.86628)	terrified (+0.80123)	lamentations (+0.76397)	vicious (+0.81727)
trembled (+0.86267)	ruffians (+0.79973)	conjured (+0.76350)	emotions (+0.81552)
assassin (+0.86151)	infamous (+0.79818)	palpitating (+0.76338)	embracing (+0.81379)
fear (+0.85964)	tremble (+0.79753)	charmed (+0.76337)	conjectures (+0.81296)
Madrid (+0.85547)	caverns (+0.79354)	inquisition (+0.76296)	alleviate (+0.81244)
court (+0.85539)	paces (+0.79226)	proofs (+0.76286)	artful (+0.81220)
tormented (+0.85472)	pallid (+0.79154)	diffident (+0.76285)	nun (+0.81211)
gaming (+0.85406)	obedient (+0.79138)	intention (+0.76230)	mournfully (+0.81141)
rigid (+0.84978)	injure (+0.78793)	depravity (+0.76181)	groan (+0.80944)
hero's (+0.84810)	unacquainted (+0.78756)	ill-fated (+0.75935)	endeavor (+0.80797)
fails (+0.84545)	tortures (+0.78751)	cavern (+0.75909)	vices (+0.80735)
unmindful (+0.84443)	robber (+0.78725)	expiring (+0.75893)	alighting (+0.80706)
monk (+0.84399)	calamities (+0.78566)	deprive (+0.75821)	receives (+0.80697)
mortally (+0.84201)	efforts (+0.78276)	awaits (+0.75693)	signora (+0.87131)
horror (+0.84192)	ardently (+0.78264)	chamber (+0.75588)	convent (+0.86628)
grieves (+0.84136)	infinite (+0.78060)	deprived (+0.75587)	trembled (+0.86267)
atrocious (+0.84032)	vaults (+0.77977)	expired (+0.75540)	assassin (+0.86151)
obdurate (+0.83827)	humane (+0.77974)	fortune (+0.75511)	fear (+0.85964)
dungeon (+0.83539)	hapless (+0.77966)	innocence (+0.75478)	Madrid (+0.85547)
confident (+0.83515)	innumerable (+0.77859)	fainting (+0.75383)	court (+0.85539)
perfections (+0.83344)	captivating (+0.77757)	confinement (+0.75382)	tormented (+0.85472)
amiable (+0.83287)	cultivate (+0.77730)	groans (+0.75275)	gaming (+0.85406)
fainted (+0.83246)	torn (+0.77660)	seduced (+0.75241)	rigid (+0.84978)
caprices (+0.83169)	engaging (+0.77652)	libertine (+0.75162)	hero's (+0.84810)
extremely (+0.82924)	unheard (+0.77527)	horrid (+0.75147)	fails (+0.84545)
rural (+0.82767)	fix (+0.77464)	wretched (+0.75051)	unmindful (+0.84443)
informing (+0.82763)	opulent (+0.77370)	enraptured (+0.75013)	monk (+0.84399)

Table 2

On the other hand, here is the result for the word "Rome":

richness (+0.95963)	strife (+0.89318)	palaces (+0.82123)	weapon (+0.87620)
embroidered (+0.95363)	gaudy (+0.89289)	magistrate (+0.81834)	midst (+0.87615)
demonstrations (+0.95185)	calculation (+0.89217)	payment (+0.91491)	rein (+0.87559)
crowded (+0.94994)	furnished (+0.89189)	enterprise (+0.91490)	galloped (+0.87490)
bustling (+0.94918)	brow (+0.89100)	arrest (+0.91486)	expediency (+0.87201)
falsely (+0.94733)	retainers (+0.89045)	sprinkled (+0.91437)	vast (+0.87197)
number (+0.94358)	royal (+0.88723)	mounted (+0.91395)	attorney (+0.87100)
walls (+0.94173)	perfumed (+0.88537)	harassing (+0.91349)	highness (+0.87052)
flourishing (+0.93954)	pestilence (+0.88429)	splendid (+0.91309)	statues (+0.86716)
mob (+0.93818)	halls (+0.88418)	conflicting (+0.91282)	rivalry (+0.86662)
draught (+0.93648)	defy (+0.88307)	ground (+0.91199)	cannon (+0.86598)
demon (+0.93632)	magnificent (+0.88304)	pope (+0.91124)	dwellings (+0.86420)
spectacle (+0.92652)	truth (+0.88228)	swear (+0.90796)	glassy (+0.86337)
realm (+0.92524)	vessels (+0.88130)	devouring (+0.90695)	sophisticated (+0.86183)
figured (+0.92521)	ushered (+0.88114)	constituted (+0.90689)	follower (+0.86083)
dimensions (+0.92413)	high-minded (+0.88085)	familiarly (+0.90659)	gallantly (+0.85948)
wine (+0.92061)	affair (+0.88069)	building (+0.90457)	holiness (+0.84982)
lordly (+0.91850)	bloodshed (+0.88067)	pompous (+0.90274)	canvass (+0.84724)
squire (+0.91828)	demeanor (+0.88051)	dazzling (+0.90272)	apprentice (+0.82125)
beset (+0.91822)	arrested (+0.88010)	leathern (+0.89771)	
tread (+0.91692)	habitual (+0.87991)	sentinel (+0.89653)	
single (+0.91602)	regiments (+0.87967)	gates (+0.89652)	
calculating (+0.87633)	court (+0.87869)	cloak (+0.89632)	

Table 3

Whereas this are the correlations for “Florence” and “Tuscany”:

staggering (+0.96213)	slap (+0.90467)	table (+0.87884)
purchases (+0.95717)	cushion (+0.90395)	boudoir (+0.87869)
flinging (+0.95351)	overhanging (+0.90366)	oriental (+0.87108)
grinning (+0.94502)	interposed (+0.90353)	waistcoat (+0.87058)
glance (+0.93277)	glancing (+0.90122)	mother-in-law(+0.86992)
thrusting (+0.93115)	astounding (+0.90119)	hand(+0.86924)
confidential (+0.92511)	unusual (+0.89982)	mastery(+0.86779)
hissing (+0.92213)	tumbler (+0.89636)	blanched (+0.90628)
winked (+0.92143)	bolted (+0.89588)	drawing (+0.90564)
exciting (+0.91920)	handkerchief (+0.89500)	tone (+0.90559)
aloud (+0.91736)	rambling (+0.89443)	clothing (+0.88157)
spinster (+0.91728)	satin (+0.89236)	glass (+0.88056)
glimpse (+0.91629)	loan (+0.89236)	tranquilly (+0.88019)
intently (+0.91537)	sprang (+0.89204)	bust (+0.90725)
ghastly (+0.91229)	appreciate (+0.89170)	shirt (+0.90638)
sternly (+0.91215)	awaiting (+0.89094)	expression (+0.88300)
moment's (+0.91183)	handkerchiefs (+0.88940)	wink (+0.88287)
pause (+0.90799)		

Table 4

The data resulting from the unsupervised word-correlation procedure show an interesting differentiation in the significant traits of these four sets of locations. In fact, whereas group 1 (“Italy”) and 2 (“Naples”, “Palermo”, “Sicily”) seem to be ascribable to the same type of semantic identity, even with slightly variations, the same cannot be said for group 3 (“Rome”) and 4 (“Florence, Tuscany”).

As a whole country, the location named “Italy” is proven to be highly correlated to a specific set of genre-determined words: Mysterious, Vice, Haunt, Chateau, Orphan, Treacherous, Villain, Death-like, Ruined, Prison, Confinement, Voluptuous, Libertine. Likewise, and even more genre-determined, the correlations generated by “Naples”, “Palermo” and “Sicily” evoke dark imageries of Artifice, Dungeons, Nun, Monk, Inquisition, Horror, Fainting, Ruffian, Fear, Misfortune, Torture, Pallid, Robber, and Captivating. In both cases, the correlation experiment associates the Italian semantic identity to motifs and features belonging to the Gothic genre.

On the other hand, the semantic field related to the central areas of the Italian peninsula does not seem to be constructed around this same univocal connotation. In fact, the like-trend words that correlate with “Rome” can be divided in at least two categories: Turmoil (Demonstration, Mob, Walls, Weapon, Regiment, Arrested, Sentinel, Cannon, Rivalry, Conflicting) and Nobility (Gallantly, Court, Palace, Royal, Cloak, Squire, Sophisticated, Furnished, Vessel, Mounted, Pompous, Galloped, Demeanor, Highness, Calculation). As regards “Florence” and “Tuscany”, instead, a narrower version of this same aristocratic imagery seems to be the structuring principle of their semantic field. Here, in fact, the presence of correlations such as Purchase, Clothing, Satin, Glass, Table, Boudoir, Oriental, Waistcoat, Awaiting, Handkerchiefs, Flinging, Glancing, Expression, Wink, Loan, Pause suggests a further increase in the cultural significance of a new type of upper-class gentele sensibility, both in terms of material culture and social habits.

#### 4. Conclusion

What these two experiments seem to have proven so far is the idea of a direct proportionality between the historical occurrence of these 4 sets of Italian locations and their semantic behavior, thus connecting the emergence of new geographical regions in the space of British cultural imagery with the rise of specific genres.

The novelistic form that more than any other found in Italy its major source of inspiration is unquestionably the Gothic Novel. It is indeed during the dark season of this genre that Southern Italy imposes itself as a synecdoche of the whole country popularizing the stereotype of Italy as an exotic laba, to use Baudelaire’s words, of sublime landscapes, violence, despotic ruling classes and degenerate passions. Not too dissimilar to the late-century Orientalist-reverie, Italy was turned by the Gothic writers into a de-domesticized space of strangeness and subversion where the tensions and the bewilderment of the “age of revolutions” could find their environmental conditions of possibility.

As the XIX century progresses though, southern regions give way to the rise of central Italy locations such as “Rome” “Florence” and “Tuscany”. Although it is clear that the representation of these new areas eludes the specific constraints of the Gothic imagery, both in terms of chronological order and in those of semantic features, the variety of correlations that these locations produce does not seem to allow a conclusive classification in terms of genre. One possible speculation, though, could be that this movement of the Italian semantic identity toward a more refined and aristocratic set of values is to be seen as a consequence of the development of the Silver fork novel in the Victorian era. However, to turn this into a more empirically supported thesis will take another experiment and, above all, another article.

By now, I just want to conclude this work stressing, once again, the potential of the computer based macro-analysis not only in terms of enabling the investigation of larger scale literary phenomena but also, and especially, for its capacity of suggesting new stimulating challenges, otherwise unperceived and unperceivable from another perspective.

## 5. References

- i. Allison, S., Heuser, R., Jockers, M. et al. (2011) *Quantitative Formalism: An Experiment*. Literary Lab, Pamphlet 1 (Jan 2011). Stanford University Press.
- ii. Barthes, R., (1970). *Empire of the Signs*, NY: Hill and Wang.
- iii. Brad C., (2011). *Italy and the English romantics, the Italianate fashion in early nineteenth-century England*, Cambridge University Press.
- iv. Boitani, P., (1983). *Chaucer and the Italian Trecento*, Cambridge Paperback Library.
- v. Cavaliero, R. (2005). *Italia Romantica*, Tauris Parke Paperbacks.
- vi. Churchill, K. (1980). *Italy and English Literature 1764-1930*, London: Macmillan
- vii. Demata, M. (2006). "Introduction: Italy and the Gothic", *Gothic Studies*, vol. 8/1, pp.1-8.
- viii. Hebron, S. (2006). *The Romantics and the British Landscape*, British Library Publishing Division.
- ix. Heuser, R., Le-Khac, L. (2012). *A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method*. Literary Lab, Pamphlet 4 (May 2012). Stanford University Press.
- x. Jockers, M., (2013). *Macroanalysis: Digital Methods and Literary History*. Urbana-Champaign: University of Illinois Press.
- xi. Jockers, M., (2014). *Text Analysis With R for Students of Literature*. New York: Springer, 2014.
- xii. Laven, D., (2003). *Turner and Venice*, London: Tate Gallery Pubn.
- xiii. Marrapodi M., (ed.)1993,*Shakespeare's Italy: Functions of Italian locations in Renaissance drama*, Manchester University Press
- xiv. Moretti, F. (1997). *Atlas of the European Novel*. London: Verso.
- xv. Moretti, F., (2005). *Maps, Graphs, Trees: Abstract Models for a Literary History*. London: Verso.