



LA LETTERATURA IN LABORATORIO

FRANCO MORETTI

ALGEE-HEWITT, ALLISON, GEMMA, HEUSER, IMPETT,
JOCKERS, KANATOVA, KATSMAN, MILYAKINA, PESTRE,
PIUPOVEC, SHELYA, SOBCHUK, STEINER, TEVEL, TINITS,
YAMBOLIEV, WALSER, WITMORE

A CURA DI GIUSEPPE EPISCOPO

Federico II University Press



fedOA Press

Indice

Premessa

Stanford-Napoli, Opificio 4.0

di Francesco de Cristofaro

9

Introduzione

Macchine che pensano. Le Digital Humanities secondo Franco Moretti

di Giuseppe Episcopo

17

Prima parte

PER UNA TEORIA

1. *Misurare le letterature*

di Franco Moretti

33

2. *«Operazionalizzare»: o la funzione della misurazione nella moderna teoria letteraria*

di Franco Moretti

43

3. *Interpretare pattern*

di Franco Moretti

63

Seconda parte

SAGGI DI METODO

4. *Lo stile al livello della frase*

di Franco Moretti e Sarah Allison, Marissa Gemma, Ryan Heuser, Amir Tevel, Irena Yamboliev

83

5. *Canon/archivio. Dinamiche su larga scala nel campo letterario*

di Franco Moretti e Mark Algee-Hewitt, Sarah Allison, Marissa Gemma, Ryan Heuser, Hannah Walser

125

6. *Sui paragrafi. Scale, temi e forme della narrazione*

di Franco Moretti e Mark Algee-Hewitt, Ryan Heuser

173

Terza parte

ESPERIMENTI SUL ROMANZO

7. *Un esperimento di formalismo quantitativo* 207
di Franco Moretti e Sarah Allison, Ryan Heuser, Mathew Jockers,
Michael Witmore
8. *Il rumore nel romanzo* 241
di Holst Katsma
9. *Le emozioni di Londra* 279
di Franco Moretti e Ryan Heuser, Erik Steiner

Quarta parte

OLTRE LA LETTERATURA

10. *Come parla una banca. La lingua nei rapporti della Banca Mondiale, 1946-2012* 307
di Franco Moretti e Dominique Pestre
11. *Il tempo a pezzi. Anacronie nel cinema contemporaneo* 339
di Maria Kanatova, Alexandra Milyakina, Tatyana Pilipovec,
Artjom Shelya, Oleg Sobchuk, Peeter Tinitis
12. *Totentanz. «Operazionalizzare» le Pathosformeln di Aby Warburg* 369
di Franco Moretti e Leonardo Impett

Postfazione

- Le lettere e l'algoritmo. L'ars mensurandi del Lab* 403
di Paola Di Gennaro

Traduzioni di Antonio Bibbò (2), Mirta Cimmino (8), Sara de Simone (10), Paola Di Gennaro (3), Giuseppe Episcopo (1, 4, 5, 6), Luca Marangolo (5, 7), Serena Messina (11), Pasquale Palmieri (4), Alfredo Palomba (6), Chiara Salierno (9), Valentina Sturli (12).

¶5. Canone/archivio. Dinamiche su larga scala nel campo letterario*

Franco Moretti e Mark Algee-Hewitt,
Sarah Allison, Marissa Gemma,
Ryan Heuser, Hannah Walser

1. Indicatori metrici sociologici

1. Le verdure e la dote nuziale

Tra le innovazioni introdotte dalla digitalizzazione nello studio della letteratura, l'ampiezza dell'archivio è probabilmente quella più visibile: eravamo abituati a lavorare su una paio di centinaia di romanzi del XIX secolo e ora possiamo analizzarne migliaia e decine di migliaia, domani centinaia di migliaia. È un momento euforico per la storia letteraria quantitativa: è come avere un telescopio che permette di vedere galassie totalmente inesplorate. Ma è anche la resa dei conti, e allora: i cieli del digitale ci ha rivelato qualcosa in grado di modificare la nostra conoscenza della letteratura?

Non è una domanda retorica. Nel celebre saggio del 1958, in cui salutava «l'avvento della storiografia quantitativa» che avrebbe portato alla «recente rottura con la storiografia tradizionale del XIX secolo», Fernand Braudel aveva indicato come oggetti di studio più tipici «il progresso demografico, la variazione dei salari e dei tassi di interesse [...] la produttività [...] la pro-

* Questo progetto è stato sostenuto da una sovvenzione della Fondazione Maison Sciences de l'Homme di Parigi e dalla Mellon Foundation. La ricerca è stata condotta in collaborazione con un gruppo di lavoro del Labex OBVIL della Sorbonne.

duzione e la domanda di denaro»¹. Si trattava, per ognuno di questi casi, di entità chiaramente quantificabili ma anche di *oggetti completamente nuovi*, se paragonati allo studio della legislazione, delle campagne militari, dei governi, della diplomazia e così via. Fu questa duplice trasformazione che cambiò la pratica storica, non la mera quantificazione. Nel nostro caso, invece, non c'è un cambiamento nell'oggetto di studio: potremo arrivare a studiare 200.000 romanzi invece di 200 ma saranno sempre romanzi. Dov'è esattamente la novità?

Nei 199.000 libri mai studiati da nessuno – suona la risposta più canonica – come potrebbe non essere questa una novità? È una dimensione interamente nuova della storia letteraria. «Cosicché conosciamo gli scambi che recano prestigio meglio degli scambi quotidiani, le prestazioni rituali meglio dei servizi comuni», ha scritto André Leroi-Gourhan in *Il gesto e la parola*, pochi anni dopo Braudel, «molto di più la circolazione dei denari dotati che quella [delle verdure]»². La dote nuziale e le verdure: antitesi perfetta. Entrambi importanti, ma per ragioni opposte: la dote nuziale lo è per la sua unicità, le verdure perché fanno parte delle abitudini alimentari. A prima vista qui sembra mostrarsi una perfetta analogia con la differenza tra i 200 e i 200.000 romanzi. Non appena però l'analisi si fa più approfondita le cose si complicano. Si prendano due romanzi storici pubblicati nello stesso anno, il 1814: *Waverley* di Walter Scott e *Sir Ferdinand of England* di James Brewer. Intuitivamente si sarebbe tentati di identificare il grande romanzo di Scott nella dote nuziale e il *Sir Ferdinand* nell'umile ruolo del legume. A ben vedere, però, il romanzo di Scott rappresentò tanto un grande momento di rottura formale quanto il libro che tutti in Europa leggevano: verdura e dote intrecciate. Ma se è così, che importanza possono avere tutti i *Sir Ferdinand* dell'archivio digitale? Non sapevamo nulla di essi ma ora qualche cosa la sappiamo. Questa conoscenza ha una qualche importanza?³

¹ Fernand Braudel, *Storia e scienze sociali: La lunga durata*, in Id., *Saggi sulla Storia*, trad. it. A. Tenenti, Bompiani, Milano 2003, pp. 42-43.

² André Leroi-Gourhan *Il gesto e la parola*, vol I., *Tecnica e linguaggio* (1965), trad. it. di F. Zannino, Einaudi, Torino 1977, p. 176.

³ Potrebbe anche non averne alcuna. In un saggio, uscito di recente nel numero speciale di «Modern Language Quarterly» *Scale and Value: New and Digital Approaches to Literary History*, James English ha convincentemente affermato che «a sample gathered on the principle that every individual work of new fiction must hold equal value in the analysis» [«un

Affrontiamo il problema con una delle scoperte ottenute dalla nostra ricerca: il declino del campo semantico dei “valori astratti”: parole come modestia, rispetto, virtù e così via, descritti da Ryan Heuser e Long le Khac in *A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method* (FIGURA 1.1). A prescindere da quel puntiglioso 2958, Heuser e Le Khac hanno visto l’ampiezza dell’archivio come un aspetto cruciale della loro ricerca. Se avessero studiato il vecchio e più ristretto canone, sarebbe cambiato qualcosa? La risposta è nella FIGURA 1.2: no. Il canone precede l’archivio di 15-20 anni, ma la traiettoria storica è la stessa. Questo non vuol dire che il nuovo archivio non contenga nuove informazioni, vuol dire che in ogni caso dobbiamo ancora imparare a porre le giuste domande. Ma prima di farlo bisogna fare chiarezza su un punto: canone e archivio, cosa significano questi due termini?

campione raccolto in base al principio che ogni nuova opera individuale di *fiction* debba contenere pari valori di analisi] – vale a dire, un campione molto simile al nostro “archivio” – non è particolarmente «suitable for a sociology of literary production, where “production” is understood to mean not merely (or even primarily) the production of certain kinds of texts by authors but the production of certain kinds of value by a social system, whose agents include readers, reviewers, editors and booksellers, professors and teachers, and all the many moving pieces of literature’s institutional apparatus» [«adatto a una sociologia della produzione letteraria, dove per “produzione” si intende non semplicemente, (o comunque non primariamente) la produzione di certi tipi di testi di autori ma la produzione di certi tipi di valori del sistema sociale, valori i cui attori includono i lettori, i recensori, gli editori e i librai, i professori e gli insegnanti, e tutti i pezzi in gioco nell’apparato letterario»]. James English, *Now, Not Now: Counting Time in Contemporary Fiction Studies*, «Modern Language Quarterly», vol. 77, n. 4, 2016, pp. 395-418: 403. Il fatto che, quando il presente pamphlet si è concentrato sull’archivio, abbia finito per concentrarsi soprattutto sulla produzione di “certi tipi di testo” sembra evidentemente corroborare la tesi di English. D’altro canto, nella misura in cui un “sistema sociale” crea “valore” non semplicemente assegnandolo a determinati testi o autori, ma anche *negandolo* ad altri («In materia di gusti, più che in qualsiasi altra, ogni determinazione è negazione; ed indubbiamente i gusti sono innanzitutto dei disgusti». Pierre Bourdieu, *La distinzione. Critica sociale del gusto*, trad. it. di M. Santoro, il Mulino, Bompiani 1983, p. 56), allora i lettori e il resto dell’“apparato letterario istituzionale” sono presenti nella nostra narrativa, ma *sempre* e solo in un ruolo distruttivo.

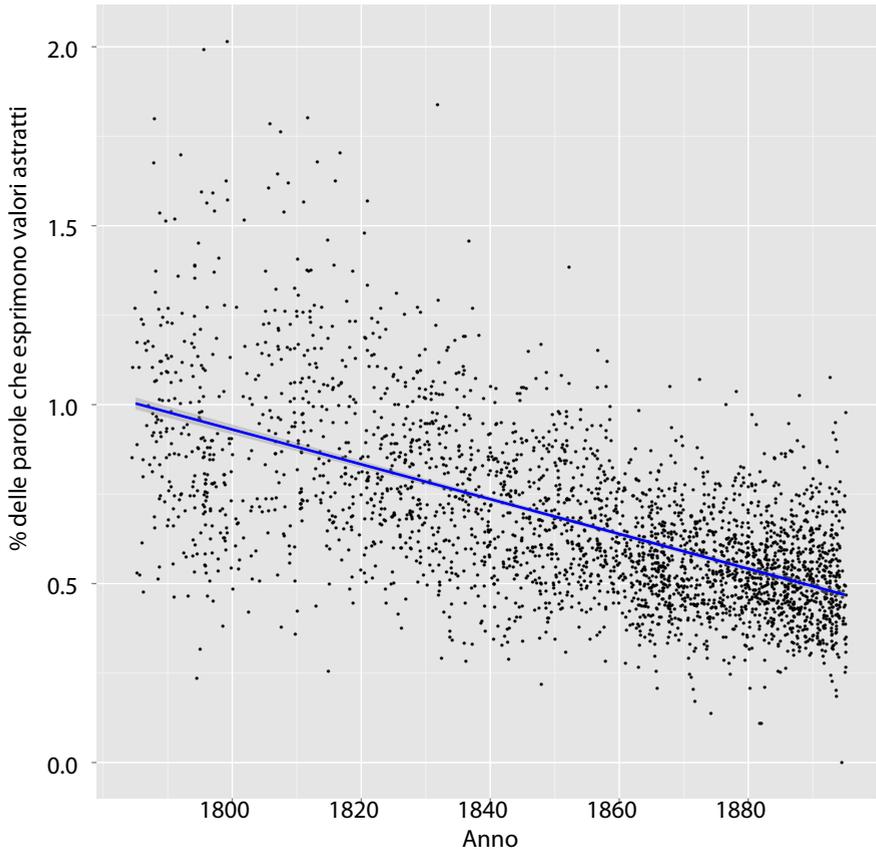


FIGURA 1.1 – I valori astratti nel romanzo inglese, 1785-1900.

Ryan Heuser e Long Le Khac, *A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method* Lab Pamphlet 4, 2012, p. 18.

2. Squilibri d'archivio

Cominciamo da tre concetti principali: il materiale pubblicato, l'archivio e il corpus. Il primo si spiega da sé: è la totalità dei libri che sono stati pubblicati (le opere teatrali messe in scena, le poesie declamate, e così via). Questa la letteratura che è divenuta “pubblica” rappresenta l'orizzonte fondamentale di tutto il lavoro quantitativo (anche se i suoi contorni sono fumosi e potrebbero essere estesi ai libri scritti ma rimasti nel cassetto, a quelli rigettati dagli editori, ecc.). Dal canto suo, l'archivio non è altro che quella parte di letteratura pubblicata che è stata preservata – nelle biblioteche o in altri luoghi – e che adesso viene progressivamente digitalizzata. Il corpus,

infine, è quella porzione di archivio che viene selezionato, per una ragione o per un'altra, allo scopo di raggiungere uno specifico obiettivo di ricerca. Il corpus è quindi più piccolo dell'archivio che, a sua volta, è più piccolo del materiale pubblicato: come delle matrioske, si inseriscono perfettamente l'uno nell'altro. Ma con la tecnologia digitale, la relazione tra questi tre campi è mutata: il corpus di un progetto può acquisire (più o meno) le dimensioni dell'archivio, mentre l'archivio sta diventando in sé – per lo meno in tempi moderni – vasto (più o meno) quanto tutta la letteratura pubblicata. Quando si usa il termine “archivio” ciò che abbiamo in mente è precisamente questa potenziale convergenza dei tre campi in uno: in quella “storia totale della letteratura”, per prendere in prestito un'espressione de *Les Annales*, che un tempo era un miraggio e che presto potrà essere una realtà.

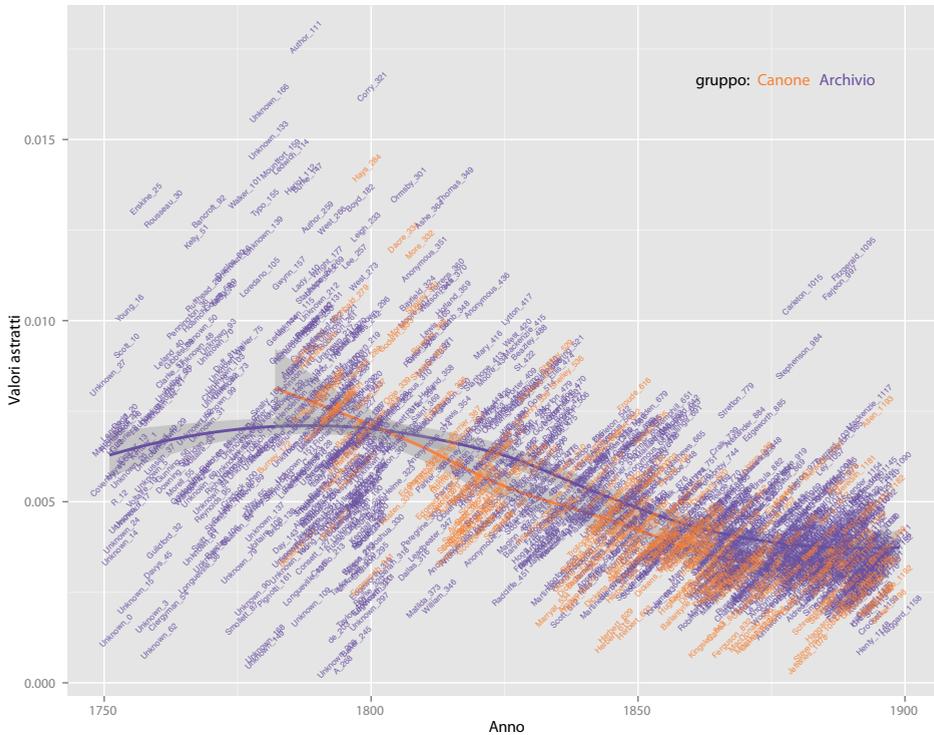


FIGURA 1.2 – Valori astratti, canone, e archivio nei romanzi britannici dal 1750 al 1900.

In questa figura il canone è rappresentato dai 250 romanzi presenti originariamente nella *Chadwyck-Healey Nineteenth-Century Fiction Collection*. Ragioneremo sulla scelta della *Chadwyck-Healey* nella III parte.

Tutto ciò in teoria. In pratica le cose non sono così semplici. Prendiamo il presente progetto. Il suo corpus iniziale consisteva di circa 4000 romanzi anglosassoni dal 1750 al 1880: per il XVIII secolo provenivano dall'ECCO, per il XIX secolo dal corpus della *Chadwick-Healey Nineteenth-Century Fiction Collection* e dall'archivio Internet dell'Università dell'Illinois⁴. Se per i vecchi canoni della storia letteraria 4000 romanzi rappresentano un corpus molto vasto, la reale copertura però è risultata alquanto diseguale. Ad esempio, per il periodo fra il 1770 e il 1830 ci siamo ritrovati con circa 1/3 dei titoli elencati nella bibliografia raccolta da James Raven, Peter Garside e Rainer Schöwerling; per l'ultimo periodo del XIX secolo, la percentuale era anche inferiore, intorno al 10%. Lo stesso è accaduto per dei generi specifici: abbiamo tenuto il 96% della bibliografia sui romanzi *silver-fork* di Allison Adburgham, ma solo il 77% di quella sul romanzo industriale di Catherine Gallagher, il 53% di quella sul romanzo storico pre-Walter Scott di Anne Stevens e il 35% della bibliografia sul romanzo gotico di Federica Perazzini⁵.

Era, chiaramente, un terreno statistico scivoloso. Paragonati alla manciata di testi considerati canonici, i nostri 190 romanzi gotici costituivano già un numero molto grande; difficile resistere alla tentazione di identificarli con l'archivio *tout court*: ma erano effettivamente rappresentativi dell'intera "popolazione" del romanzo gotico inglese? Quasi certamente no. Volendo semplificare, un campione diventa rappresentativo quando è scelto in modo casuale all'interno di una data popolazione: i nostri 190 romanzi decisamente non erano stati scelti in quel modo. Tutti loro, in definitiva, provenivano da poche grandi biblioteche e le biblioteche non acquistano i libri per raccogliere campioni statistici: vogliono invece libri che considerano sia importante conservare. Buoni libri, buoni secondo principi che sono verosimilmente non molto diversi da quelli che presiedono alla formazione dei canoni. Sebbene il nostro corpus fosse venti volte più grande del canone tradizionale, era però certamente possibile che il principio di selezione col quale era composto

⁴ Si veda <https://archive.org/details/19thcennov>. Ecco (*Eighteenth Century Collections Online*) è una collezione digitale in due parti composta da materiali del XVII secolo basati sull'English Short Title Catalogue (ESTC) e provenienti da una serie di biblioteche negli Stati Uniti e nel Regno Unito; la parte II di ECCO è un aggiornamento che consiste in testi o edizioni che non erano disponibili quando uscì l'ECCO originale.

⁵ Allison Adburgham, *Silver Fork Society*, 1983; Catherine Gallagher, *The industrial Reformation of English Fiction*, Chicago, 1985; Anne H. Stevens, *British Historical Fiction Before Scott*, London, 2010; Federica Perazzini, *Il Gotico @ distanza*, Roma, 2013.

lo facesse somigliare molto di più al canone che non all'archivio in quanto tale. Questo era il problema⁶.

Volevamo che i nostri risultati fossero affidabili e perciò abbiamo estratto un campione casuale dal campo letterario da studiare: 507 romanzi in tutto per il periodo 1750-1836, 82 romanzi per il genere gotico e 85 romanzi storici per l'epoca precedente a Scott⁷. In totale 674 romanzi. Nell'era digitale, non ci avrebbero impegnati per molto tempo.

Abbiamo creato il campione alla fine delle attività accademiche, nel giugno del 2014. Poi ci siamo rimessi a lavoro sul nostro database, nel quale abbiamo trovato 35 degli 82 romanzi gotici, 35 degli 85 romanzi storici e 145 dei 507 romanzi contenuti nelle raccolte bibliografie di Raven e Garside. A inizio luglio, abbiamo passato la lista di titoli che non eravamo riusciti a trovare (circa 460) a Glenn Worthey e Rebecca Wingfield delle biblioteche di Stanford che hanno prontamente estratto dalla matassa pochi grandi nuclei fondamentali. Circa 300 testi erano conservati (in proporzioni più o meno uguali) dall'Hathi Trust e da Gale (attraverso NCCO e ECCO II)⁸. Altri 30 erano presenti in raccolte, in seconde edizioni, nascosti sotto titoli in parte diversi, in collezioni di microfiches o di microfilm, e così via; circa 100 di questi esi-

⁶ A complicare ulteriormente la faccenda, generi differenti subiscono variazioni diverse fra canone e archivio: se i romanzi epistolari e i *silver-fork* hanno un archivio relativamente grande e un canone molto ristretto, per il romanzo industriale e i romanzi di formazione è il contrario: entrambi attrassero grandi scrittori vittoriani; i due sovra-generi del gotico e del romanzo storico si trovano invece da qualche parte fra questi due estremi. Su questo – e su molto altro – abbiamo bisogno di molte più prove materiali.

⁷ Quest'ultimo gruppo non era un campione a caso: la bibliografia di Ann Scott includeva 85 romanzi storici nell'era precedente Scott, abbiamo deciso di cercarli tutti.

⁸ Hathi Trust è partner delle più grosse biblioteche di ricerca e funziona da archivio delle collezioni digitali; di queste fanno parte i volumi digitalizzati per il progetto Google e per l'*Internet Archive*, così come per altri progetti locali più piccoli. NCCO (*Nineteenth Century Collections Online*) è una collezione digitale di materiali del XIX secolo, di solito provenienti da grandi collezioni, che spazia fra varie discipline differenti (letteratura, scienza e tecnologia, fotografia ecc.). Per questa regione esistono dodici parti di NCCO: una di queste è la *Corvey novel collection*; a differenza di ECCO, NCCO non è basato sulla bibliografia standard del settore ed è perciò difficile prevedere cosa verrà aggiunto. Gale è un grande conglomerato di informazioni e servizi per l'istruzione – gestito a fini di profitto – che vende contenuti e servizi alle biblioteche; pubblica sia lavori a stampa (libri di testo e narrativa) che collezioni elettroniche (fra cui appunto ECCO, NCCO, e altri). La società controllante è Cengage Learning, che si definisce “una compagnia leader nel settore dell'istruzione, della tecnologia per gli studi superiori e la scuola e per il mercato bibliotecario e professionale mondiale.”

stevano solo a stampa, e per 10 di questi romanzi non c'erano copie esistenti conosciute. Ad agosto, le richieste furono inviate a Hathi e a Gale – con entrambe Stanford aveva un accordo finanziario di lunga data – per i trecento volumi in loro possesso. Dei 100 romanzi che esistevano solo a stampa, la metà circa era posseduta dalla British Library di Londra, che solo pochi mesi prima aveva offerto gentilmente al nostro *Literary Lab* una collezione di 65.000 volumi digitalizzati, sfortunatamente però nessuno dei libri che cercavamo era lì. Le collezioni speciali della UCLA e di Harvard, in cui erano conservati circa 50 di quei libri, ci hanno mandato una serie di preventivi che variavano (a seconda, piuttosto ragionevolmente, delle condizioni dell'originale, delle complessità del produrre il materiale in copia fotografica) tra i 1000 e i 20.000 dollari a romanzo. Per concludere, 6 romanzi che facevano parte di collezioni più grandi erano di proprietà di Proquest e ci sarebbero costate – al netto dello sconto generoso del 50 % – fra i 25.000 e i 147.000 dollari per titolo⁹.

Si tenga presente che questa ricerca coinvolgeva: molti bibliotecari di eccellenza fra Londra, Cambridge, Los Angeles e ovviamente Stanford; una mezza dozzina di ricercatori del *Literary Lab*; e in più il personale di Hathi Trust, di Gale, e così via. I libri che cercavamo erano vecchi solo di due secoli ed erano stati stampati in un regime di almeno 750-1000 copie per titolo, in una parte del mondo che, all'epoca, possedeva già delle biblioteche efficienti. Il *Literary Lab* possiede alcuni fondi di ricerca (anche se, a scanso di equivoci, non *quel* genere di denaro): in altre parole, non si poteva sperare in risorse migliori. Ciò nonostante, ci vollero circa sei mesi per ricevere da Hathi Trust e Gale i libri che ci avrebbero permesso di crescere dall'iniziale 30% fino all'attuale 70-80% del campione casuale¹⁰: una cifra che avrebbe reso pro-

⁹ A queste cifre deve essere aggiunto quel che le *Stanford libraries* avevano pagato per ECCO, ECCO II e NCCO tanto per cominciare: con i soliti sconti generosi, eravamo arrivati a circa un milione di dollari per le tre collezioni. ProQuest è un altro servizio educativo *for-profit* i cui prodotti includono la *Historical Newspapers series*, *Literature Online*, *Dissertation Abstracts*, e altri. La sua compagnia di riferimento è il *Cambridge Information Group*.

¹⁰ “*Avrebbe* dovuto permetterci” perché ricevere un testo da queste collezioni non è la stessa cosa che poterci lavorare su. Molti dei dati provenienti dalla *Chadwyck-Healey* e da ECCO di solito venivano recapitati su nastri informatici che richiedevano dei programmi che erano difficili sia da trovare sia da usare; fonti di informazioni più “agevoli” (come il trasferimento dati via rete, o con un hard drive esterno) hanno anch'essi le loro complicazioni, che vanno dalla capienza dei sistemi mail a strambe incompatibilità firewall sino ad assurde richieste di accordi contrattuali per il loro utilizzo. (Molte degli accordi di licenza d'utilizzo delle *Stan-*

tabilmente i risultati delle nostre ricerche contestabili, poiché quel 20-30% mancante avrebbe rappresentato, quasi per definizione, la parte esclusa da ogni ulteriore prospettiva di canonizzazione.

Prova evidente che l'idea che la digitalizzazione avesse reso ogni cosa disponibile e a buon mercato – per non dire addirittura “gratis” – fosse un mito. A mano a mano che questi problemi divenivano progressivamente evidenti, decidemmo di iniziare a lavorare con una selezione del corpus che era a nostra disposizione: un database di 1117 opere, 263 delle quali provenivano dalle collezioni *Chadwyck-Healey* e 854 da varie fonti archivistiche. I risultati iniziali ci spinsero subito in una direzione, quelli nuovi diedero ulteriore spinta alla ricerca cosicché quando il campione (quasi) casuale era finalmente (quasi) pronto, eravamo ormai troppo coinvolti nel nostro lavoro per ricominciare da zero. Non intendiamo presentare questo approccio come un modello ideale di ricerca, e siamo consapevoli che i nostri risultati sono indeboliti proprio a causa della nostra decisione. Ma le ricerche collettive, specialmente se condotte in una sorta di spazio istituzionale “interstiziale” –

ford Libraries, per esempio, erano piuttosto vaghe sulle miniere testuali, o sul condividere materiale al di fuori delle strutture di conservazione della biblioteca; negli ultimi cinque anni le biblioteche avevano esplicitamente insistito sull'esigenza di acquisire i diritti delle miniere testuali in normative attuali, ma gli accordi precedenti rimanevano piuttosto ambigui).

Alla fine, estrarre dati da un oceano di nastri e hard drive, con metadati insufficienti o scorretti e senza alcun database che fornisca assistenza, è veramente un processo bizantino. Le librerie avrebbero cercato sul database di ECCO, per esempio, usando un'interfaccia di *Gale*, citando la sua URL secondo le procedure di quest'interfaccia. Ma perché le biblioteche potessero inviare un file non elaborato al laboratorio, era necessario passare attraverso un paio di hard drive (o un paio di nastri) che contenevano centinaia di migliaia di directory salvate con nomi composti solo di serie numeriche a caso; i metadati del “manifesto” che Gale invia assieme a questi file grezzi è contenuto in circa dieci file Word formattati per la stampa: due colonne, gli autori in grassetto, un catalogo basilare dei dati, un documento di riconoscimento, un identificativo dell'*English Short Title Catalogue*, ed il percorso delle directory. Questi documenti sono immensi: ECCO II, il modulo di letteratura e linguaggio, gli autori da L a Z – che rappresentano circa un decimo della mole di informazioni di ECCO – è un documento di circa 2750 pagine. In secondo luogo i codici identificativi non sono gli stessi dell'interfaccia Gale; sono interni, sono numeri invisibili. Quindi, per quanto tutto il lavoro effettuato dal nostro Lab per identificare le risorse di ECCO usando il database e annotandovi il numero identificativo ufficiale assegnato in Gale, le biblioteche sono state obbligate a effettuare di nuovo la ricerca di ogni oggetto, per autore o per titolo, per trovare il nome di un file o un file da copiare – il numero identificativo di Gale – che non era mai presente nei dati del file. «La mia morale» – ha concluso un bibliotecario che ci ha assistito in tutto il processo – «è questa: anche quando abbiamo trovato il file di cui abbiamo bisogno, non l'abbiamo veramente trovato».

com'è tutt'ora il nostro – ha le sue tempistiche: aspettare mesi e mesi prima che una ricerca possa incominciare ucciderebbe qualsiasi progetto. Forse in futuro manderemo un esploratore, con un anno in anticipo, a raccogliere il campione. O forse continueremo a lavorare con quel che abbiamo, con la consapevolezza dei limiti e delle falle nei dati di riferimento. È meglio ritrovarsi con le mani sporche che a mani vuote.

3. Dal canone al campo letterario

Se la composizione del nostro archivio era stata determinata dalle pratiche seguite dalle biblioteche nel corso della storia (quali romanzi erano sugli scaffali? quali erano facili da digitalizzare?) quella del nostro canone era una questione di giudizio critico, sebbene non il nostro. Il primo canone a cui ci siamo rivolti per questo progetto, la *Chadwyck-Healey Nineteenth Century Fiction Collection*, era stato definito da un comitato editoriale di due persone, Danny Karlin e Tom Keymer¹¹. Si tratta di un insieme di circa 250 romanzi scelti perché così meritevoli di essere preservati, e di un tale valore per gli studiosi, che le biblioteche sono disposte a pagare per avere un accesso digitale.

Compilata nei tardi anni '90, con aggiunte successive di nuovi romanzi, la *Nineteenth-Century Fiction Collection* «riunisce», come si afferma nei materiali pubblicitari, «i più grandi successi del canone vittoriano e rispecchia i punti di riferimento del periodo», ma comprende al contempo «molti lavori misconosciuti o poco conosciuti, molti dei quali sono fuori stampa o difficili da trovare». Tra le opere del 1794, ad esempio, la collezione include *The Mysteries of Udolpho* di Ann Radcliffe e *Caleb Williams* di William Godwin, ma anche *Lady Susan* di Jane Austen (un romanzo molto breve, scritto più o meno in quel periodo ma pubblicato postumo nel 1871) e il radicale *The Adventures of Hugh Trevor* di Thomas Holcroft. Le prime due scelte sono scontate, le altre due meno. Sembra allora che selezionare 250 testi possa concedere spazio a libri meno conosciuti ma di grande importanza storica e critica: non solo i sei maggiori romanzi di Austen, ma anche *Lady Susan*; non solo Godwin, ma anche Holcroft. Fintanto che riteniamo che un “canone” debba consistere in un piccolo numero di testi selezionati e consacrati ai fini di uno studio ravvicinato, allora la *Chadwyck-Heaney* – una collezione facil-

¹¹ Una personale comunicazione con Steven Hall ha confermato che i curatori sono stati totalmente liberi nella scelta dei testi.

mente consultabile dai ricercatori¹² – non è male come approssimazione.

Ma è pur sempre di un'approssimazione e noi ci siamo resi conto che affidarsi a una sola fonte era un modo sbagliato per approcciarsi a un concetto sfaccettato ed elusivo come il canone. Nel loro *Between Canon and Corpus: Six Perspectives on 20th-Century Novels* (Literary Lab Pamphlet 8, 2015) Mark Algee-Hewitt e Mark McGurl si sono trovati ad affrontare un problema simile quando hanno presentato diverse liste di “migliori romanzi del xx secolo”, selezionate da gruppi molto diversi tra loro, di cui hanno analizzato i differenti gradi di affinità. Noi abbiamo seguito un percorso differente che ci ha portato dal breve catalogo di libri della *Chadwyck-Healey* a due lunghe liste di autori, quelli menzionati dal *Dictionary of National Biography* (DNB) e quelli catalogati come “oggetto principale di ricerca” negli articoli accademici indicizzati nella bibliografia della Modern Language Association (MLA). In un progetto collaterale, abbiamo incluso anche i testi presenti nelle liste d'esame del dottorato di ricerca di Stanford degli ultimi trent'anni. Così facendo non cercavamo né la “giusta” definizione di canone (che non era nessuna di queste), né tantomeno speravamo che il DNB, la MLA, e Stanford fossero d'accordo tra loro (e infatti non lo erano)¹³. Al contrario, la ricognizione sui differenti strumenti di misurazione aveva lo scopo di riprodurre gli aspetti più disparati dell'idea di canone: il fatto che la cultura nazionale (cioè il DNB) lo definisca in un modo e che l'accademia internazionale (ovvero la MLA) in un altro; il fatto che il canone possa essere concepito nei termini di una serie di personalità (e questo vale sia per il DNB che per la MLA), oppure nei termini di una mera collezione di testi (come per le liste del dottorato). Le scelte specifiche rimanevano discutibili – è ovvio! – ma i criteri che avremmo seguito sarebbero stati molteplici, espliciti e misurabili: è qui che risiede la novità.

¹² Ammesso che i suddetti ricercatori siano affiliati a un'istituzione che possa fornir loro le risorse necessarie. Secondo un campione rappresentativo dell'università rilevato dalla *Proquest*, nel mondo sono appena “poco più di 600” le università che si iscrivono al *Literature Online database* (LION).

¹³ Pur lasciando da parte la rappresentatività degli esami di dottorato di Stanford, un approccio tutto incentrato sul ruolo dell'autore del DNB e dell'MLA piazza il testo di Scott *Castle Dangerous* o quello di Thackeray *Catherine* sullo stesso piano di *Weaverley* e di *Vanity Fair*, cosa che certamente non può essere giusta. Tuttavia criteri alternativi fornivano comunque risultati simili, oppure avevano tempi eccessivamente lunghi.

Poi abbiamo compreso che c'erano altre caratteristiche del campo romanzesco che sarebbero dovute rientrare nell'equazione. Nelle loro bibliografie, ad esempio, Raven e Garside avevano identificato i romanzi che erano stati ristampati nelle isole britanniche o tradotti in francese e in tedesco dal 1770 al 1830. Da qui si potrebbero prefigurare dati simili per una ricerca futura: dalla tiratura di un'opera alla sua presenza nelle biblioteche popolari, nelle sale di lettura e molto altro. Anche in questi casi i criteri sarebbero altrettanto molteplici, espliciti e misurabili ma con una grande differenza rispetto al DNB e alla MLA. Ristampe e traduzioni misurano il gradimento dei romanzi presso un pubblico "generalista" e lo fanno attraverso le istituzioni del mercato letterario, il DNB e la MLA invece si focalizzano sui lettori "specializzati" e sulle istituzioni universitarie. Il primo modello misura la "popolarità" dei romanzi, l'altro il "prestigio"¹⁴.

¹⁴ La popolarità è misurata su dati del XIX secolo, il prestigio deriva da fonti del XX secolo, e questo è, ovviamente, un problema. Gli studi del XX secolo sono più affidabili da questo punto di vista: in *Becoming Yourself: the Afterlife of Reception* (Literary Lab Pamphlet 3, 2011) per esempio, Ed Finn ha classificato gli autori contemporanei nel campo letterario americano utilizzando due categorie – consumo e conversazione – che appartenevano allo stesso periodo cronologico: il "consumo" era ricavato dai dati di "hanno anche acquistato" di Amazon.com, la "conversazione" dalle recensioni letterarie coeve. È interessante notare che "consumo" e "conversazione" si allineano abbastanza bene a "popolarità" e "prestigio"; per quanto i sei canoni discussi da Algee-Hewitt e McGuff ruotino anch'essi attorno al successo, dall'altro lato prendono in considerazione una selezione culturale più "qualificata".

Nel tentativo di correggere la discrepanza fra i dati dell'Ottocento e del Novecento, studi successivi potrebbero estendere le misurazioni del prestigio tenendo presente anche i libri di testo e le antologie scolastiche (come sta facendo Martine Jej in Francia) premi (James English, *L'economia del prestigio*), recensioni su periodici del XVIII secolo e del XIX secolo, o le prime collezioni di romanzi come quelle di Barbaud o Ballantyne, o Bentley. Non c'è alcuna certezza, comunque, che queste collezioni e recensioni possano essere viste come indicatori di prestigio e non invece meri ingranaggi interni al mercato editoriale; in un interessante saggio molto recente Michael Gamer ha sostenuto entrambe le possibilità, sia quella delle ambizioni canonizzanti che quella del puro commercio: si veda "*To a collected selection: Barbaud, Scott, and the rise of the Novel (reprinted)*", in Jillian Heydt-Stevenson e Charlot Sussman (a cura di), *Recognizing the Romantic Novel*, Liverpool, 2008. Per parte sua, William St. Clair, ha espresso un netto scetticismo sul ruolo delle riviste («in generale, il peso delle riviste sembra essere stato ingigantito sia all'epoca sia dagli scrittori successivi [...] Non riesco a vedere alcuna correlazione fra le riviste, la reputazione e le vendite»). Sul prestigio dei romanzi nella prima parte del XIX secolo ha invece scritto: «per quanto riguarda la narrativa del periodo romantico non esiste alcun canone coevo riconosciuto. In effetti, la stessa definizione di canone aveva poco senso, dato che la maggior parte dei romanzi erano pubblicati anonimi. Un solo autore dominava la scena, 'l'autore di Waverley', non pubblicamente riconosciuto come il famoso poeta Sir Walter Scott fino alla metà degli anni '20 dell'Ottocento». William St.

Popolarità e prestigio. Con questa coppia di concetti, la nostra ricerca si è trovata sullo stesso terreno della classificazione radicalmente innovativa del campo letterario francese portata avanti da Bourdieu (FIGURA 3.1). Collocando i dati di popolarità sull'asse orizzontale ("bassi/alti risultati economici") e il prestigio sull'asse verticale ("alta/bassa consacrazione"), possiamo dare "una versione britannica" della classificazione di Bourdieu. Per ora, questa copre un solo genere e una manciata di decenni: eppure, a questo punto, una cartografia del campo letterario non è più un sogno ad occhi aperti (FIGURA 3.2).

Nella FIGURA 3.2, tutti i dati sono ridimensionati dagli incredibili punteggi ottenuti da Walter Scott: solo due romanzieri riescono a superarlo sull'asse del prestigio (Goethe e Austen), e nessuno gli si avvicina minimamente in termini di popolarità: l'autore più prossimo su quest'asse – si tratta di Thomas Day, autore del *best-seller* rousseauiano *The History of Sandford and Merton* (1789) – si trova a uno scarto di ben sette gradi da Scott¹⁵. Rimuovendo dal grafico i risultati fuori scala di Walter Scott, diviene visibile una tripartizione interna al sistema del romanzo britannico (FIGURA 3.3).

Iniziamo dal gruppo vicino all'asse orizzontale: scrittori con un alto risultato in termini di qualità – con uno scarto standard di 5, 8, 10 e 3 misure sopra la media – ma molto bassi in termini di prestigio: ad una coppia di misure standard in meno rispetto alla media, ma spesso solo ad una, o meno di una. Qui troviamo: il *Man of Feeling* di McKenzie e il bestseller educativo di Day; il gruppo di romanzi gotici, con i loro frequenti toni melodrammatici (Radcliffe, Reeve, Roche, Helme, Maturin); i romanzi giacobini e antigiacobini (Charlotte Smith, Opie); le epopee nazionali (Edgeworth, Morgan); la nuova forma egemonica del romanzo storico (Galt, Genlis, Horace Smith,

Clair, *The Reading Nation in the romantic period*, Cambridge 2004, p. 189.

Dall'altro lato, l'esistenza di una relazione fra le riviste e la reputazione è stata recentemente – e convincentemente – proposta da Ted Underwood e Jordan Sellers in *How Quickly do Literary Standards Change?* http://figshare.com/articles/How_Quickly_Do_Literary_Standards_Change_/1418394. Underwood e Sellers studiano la poesia anziché i romanzi e iniziano la loro indagine dal 1820, vale a dire dal punto in cui si arrestano sia il libro di St. Clair e che il nostro corpus: c'è una disomogeneità nell'oggetto di ricerca e nel periodo cronologico tale da impedire una vera e propria comparazione. Ma ci stiamo lentamente avvicinando al momento in cui le prove raccolte da studi indipendenti potranno essere felicemente messe a confronto e anche integrarsi.

¹⁵ Non misurando la produzione delle rotative, la classifica in verità *sottovaluta* la popolarità di Walter Scott: mentre la maggior parte dei romanzi coevi ha una prima stima di circa mille esemplari, le tre successivi ristampe di *Waverley* risultano rispettivamente di 6000, 8000 e 10.000.

Porter, Cooper). Potremmo chiamare questo “lo spazio del genere”, nel senso di tutti i generi: “il” romanzo si sviluppa come una famiglia di forme distinte le cui convenzioni, facilmente riconoscibili, facilitano la scalata al successo. Il capitolo d’apertura di *Weaverley*, in cui l’autore commenta le connotazioni di genere dei titoli dei romanzi, è in effetti il sintomo esemplare di questo stato di cose.

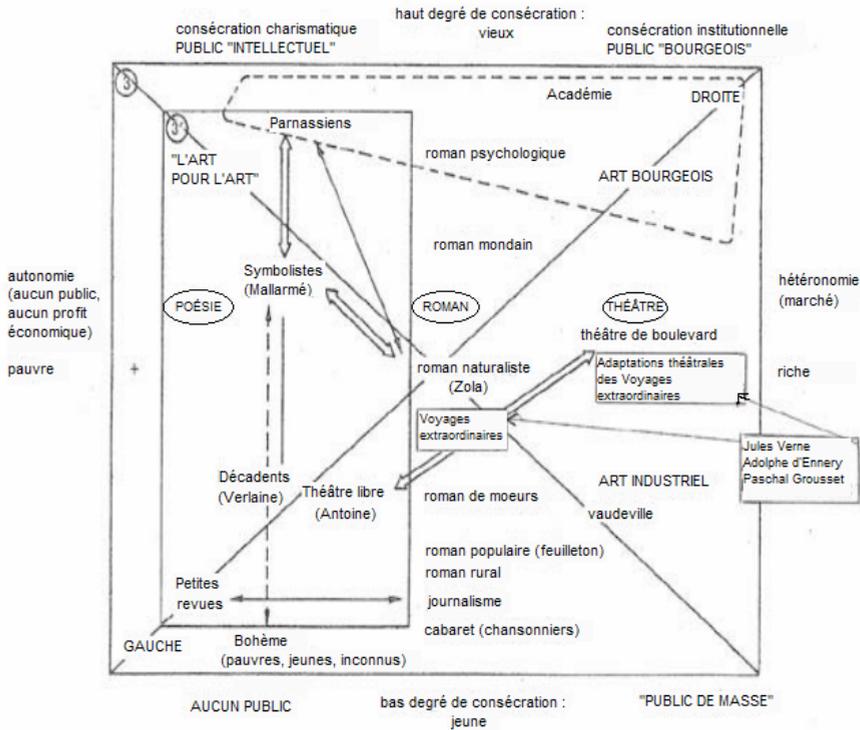


FIGURA 3.1 – Il campo letterario francese alla fine del XIX secolo.

Il diagramma di Bourdieu del campo letterario, sebbene enormemente suggestivo, non offre alcuna prova empirica a sostegno per la posizione occupata dai diversi generi e movimenti. Nella mancanza di un criterio evidente e misurabile risiede probabilmente il motivo per cui – malgrado la sua eleganza e la profonda influenza – il grafico di Bourdieu non sia mai diventato un vero e proprio strumento di ricerca, ripreso e riadattato da altri studiosi. L'incredibile regolarità della distribuzione, così diversa da quelle della FIGURA 3.2 e 3.3 nonché da quella proposta dello stesso Bourdieu in *La distinzione*, è probabilmente essa stessa una conseguenza dell'origine speculativa del diagramma. Pierre Bourdieu, *Les Règles de l'art. Genèse et structure du champ littéraire*, Paris, Seuil, 1992, p. 176.

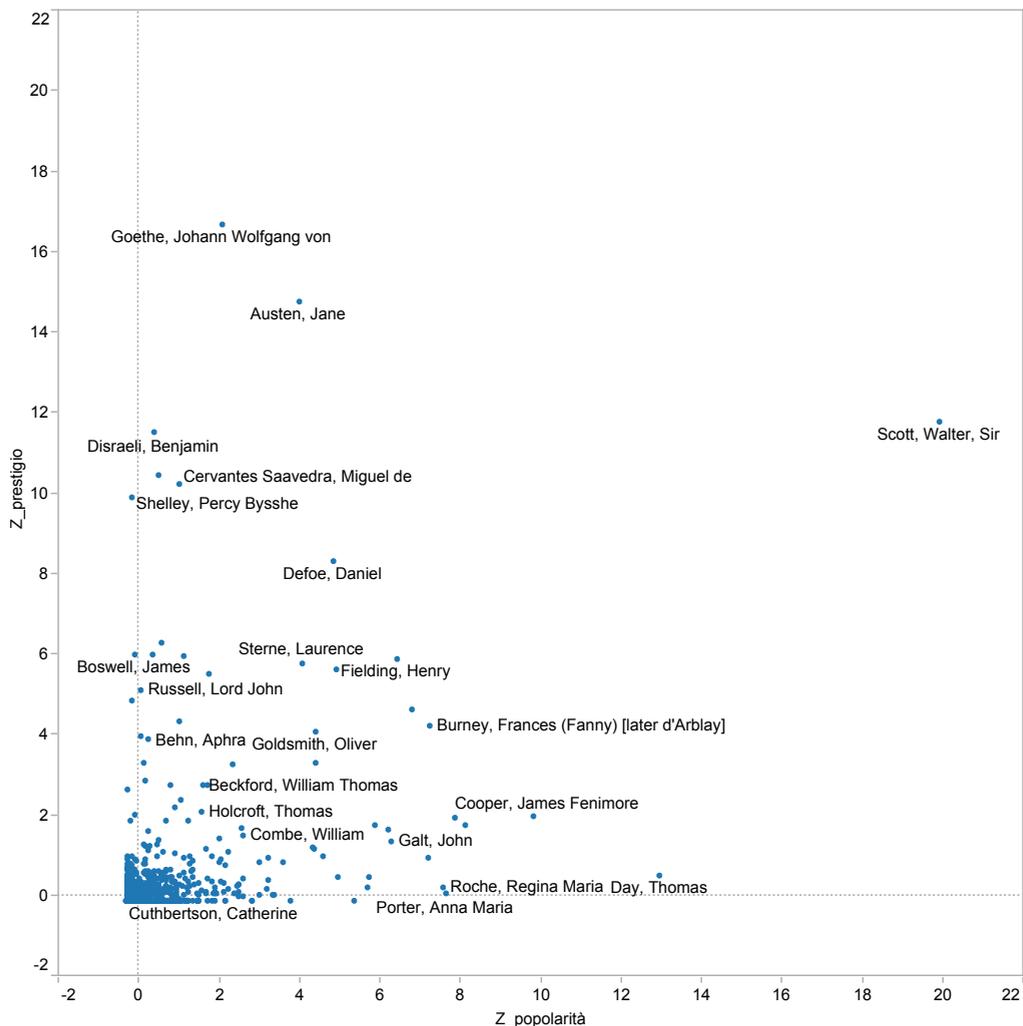


FIGURA 3.2 – Il campo romanzesco britannico, 1770–1830.

Per l'asse della popolarità i risultati sono basati sul numero delle ristampe (nelle isole britanniche) e sulle traduzioni (in francese e tedesco). Per l'asse del prestigio, i riferimenti sono stati il numero di citazioni per "autore principale" nella bibliografia MLA e la lunghezza delle voci della DNB.

La posizione degli scrittori è determinata dal numero della deviazione standard sopra la media del campo; John Galt, ad esempio, ha un valore di 7,5 deviazione standard sopra la media sull'asse della popolarità e di 1 sull'asse del prestigio. All'estremo opposto, Percy Shelley ha 10 di deviazione standard sopra la media in termini di prestigio, ma è appena sotto la media del campo in termini di popolarità.

Spostandoci al di sopra di questa area, la parte centrale del diagramma ci porta in un territorio tutto differente. Semmai si fosse giustificati ad affermare semplicemente “questo è il canone”, sarebbe in questa circostanza: Defoe, Richardson, Fielding, Sterne, Goldsmith, Burney, Goodwin... Tutti loro sono raggruppati in questo spazio perfettamente bilanciato (tra le 4 e le 7 misure standard al di sopra della popolarità media, e da 3 a 8 sopra quella del prestigio) in cui la formula della narrativa di successo si mescola pressoché indistinguibilmente con la letteratura con un alto riconoscimento culturale. Guardare a questa area centrale permette di “vedere” il processo di canonizzazione come una combinazione di due processi: la popolarità che si restringe lentamente, col passare degli anni, sull’asse orizzontale – in questo senso, la maggior parte dei giganti del XVIII secolo sono ben al di sotto dei vari Roche, Porter, Charlotte Smith e Opie – laddove il prestigio cresce sull’asse verticale¹⁶. Per quanto ci sia naturalmente più di un modo di diventare uno scrittore del canone¹⁷, la lezione principale che proviene da quest’immagine è che il canone non è “il del mondo economico rovesciato”, secondo la formula di Bourdieu sull’autonomia del campo letterario. Il canone, o almeno questo canone, è composto da autori dalle cui opere gli editori si aspettano di trarre profitti per almeno due o tre generazioni dopo il loro successo iniziale. E il prestigio, per parte sua, non è necessariamente in antitesi con la popolarità: nel nostro caso, sembra piuttosto fluire da essa, sembra che “distilli” i guadagni economici in qualcosa di più intangibile, ma anche di più durevole¹⁸.

¹⁶ Parlando di diminuzione della popolarità, Austen e i suoi contemporanei costituirebbero un perfetto *case study*: come mostra la FIGURA 3.2, circa 25 autori (un terzo di essi del XVIII secolo) erano più popolari di Austen nei sessant’anni coperti dal diagramma. Dal momento che dal XIX secolo le bibliografie sui romanzi diventarono più affidabili, sapremo quanti di loro furono più popolari di lei una o due generazioni dopo. (I primi risultati dagli anni ’30 agli anni ’40 dell’Ottocento dicono Scott, e nessun altro).

¹⁷ Il successo e la popolarità immediati di Scott sono totalmente diverse da quelli di Austen, che ha un passo significativamente più lento, o dall’ambiguo status di autori a lungo confinati in nicchie specifiche a causa del loro pubblico di partenza (Carroll) o del genere (Radcliffe, Doyle). E poi, certo, c’è la *nemesi* di ogni teoria generale del canone: *Moby Dick*.

¹⁸ Per quanto i nostri risultati fossero completamente differenti dall’idea di Bourdieu del campo letterario, non necessariamente falsificano la sua tesi dato che stiamo lavorando esclusivamente sui romanzi (a esclusione della poesia, dei drammi, delle riviste, e così via), su una nazione e su un periodo differenti. A dirla tutta, abbiamo bisogno di molte mappe empiriche dei campi letterari (al plurale) da culture ed epoche differenti per far sì che il campo letterario (al singolare) divenga un concetto solido.

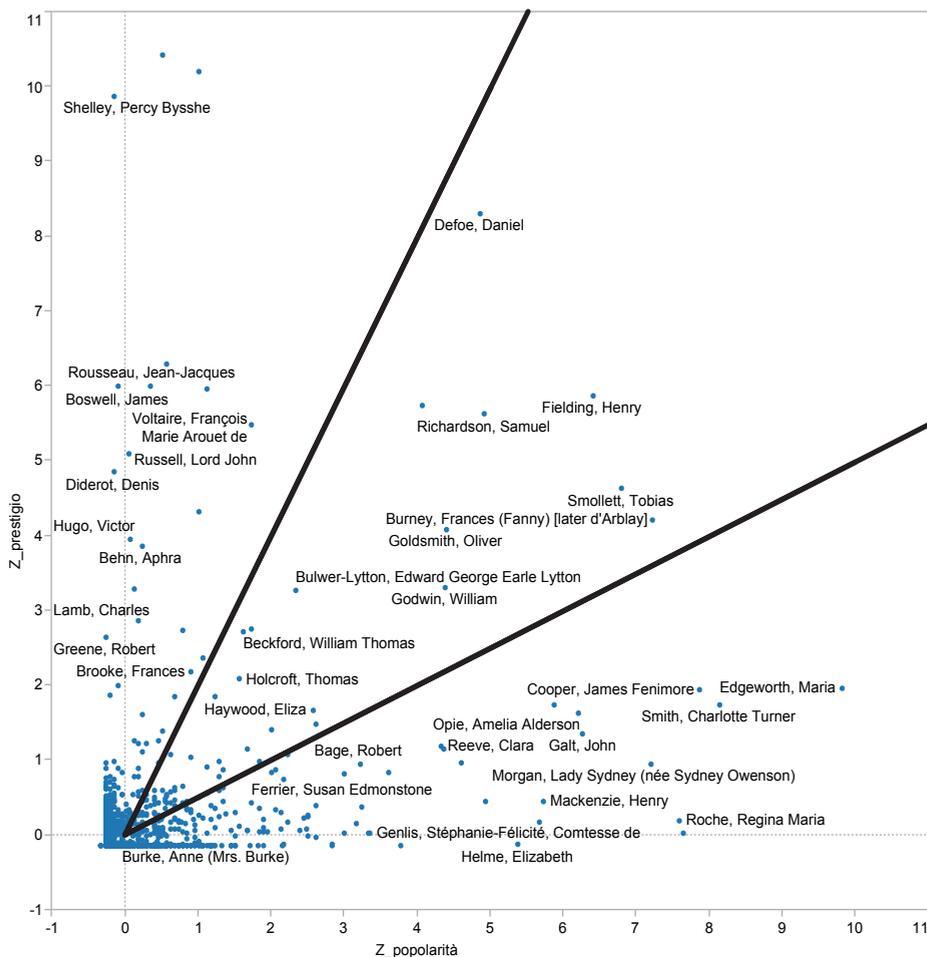


FIGURA 3.3 – Le tre regioni del campo romanzesco britannico, 1770–1830.

Le tre regioni di questo diagramma raccontano differenti tipi di relazione tra popolarità e prestigio. Nell'area vicino all'asse verticale il prestigio ha un punteggio almeno due volte più alto di quello della popolarità. L'area accanto all'asse orizzontale rovescia gli equilibri in modo speculare dal momento che il punteggio espresso dalla popolarità è almeno il doppio di quello del prestigio. Nell'area centrale i due gruppi di misurazione tendono a bilanciarsi l'uno con l'altro. Uno studio della popolarità e del prestigio su un periodo più vasto è in corso al Literary Lab sotto la direzione di J.D. Porter, i dati sono raccolti sia attraverso degli algoritmi che grazie a un gruppo di laureandi guidati da Micah Siegel.

Le cose vanno diversamente nella zona del “prestigio alto” della FIGURA 3.3, che è dominata da scrittori non anglofoni (Cervantes, Voltaire, Diderot, Rousseau, Goethe, Schiller, Hugo...) o da quegli scrittori britannici che, nello scrivere almeno un romanzo o al massimo due, possono essere a stento visti come romanzieri “professionisti”. Fra di essi si annoverano: la figura dell’enciclopedista Samuel Johnson e dell’altrettanto versatile Horace Walpole; poeti come Percy Shelley (e, più in basso, Thomas “Anacreonte” Moore e James Hogg); il romanziere e politico Disraeli e il politico-politico Lord Russell (che nel 1822 ha pubblicato un’improbabile *Nun of Arrouca*); saggisti come James Boswell e Charles Lamb; al grado più basso di prestigio, il musicista e drammaturgo Charles Dibdin, la drammaturga e attrice Charlotte Cibber Chakle, l’economista e scrittore di viaggi Arthur Young. Fra i pochi romanzieri-romanzieri, la politica gioca un ruolo insolitamente molto forte: a parte Russell e Disraeli, incontriamo la *bluestocking* Sara Scott (autrice di *Millennium Hall* e *Desmond*), Mary Shelley e Hanna More, il cui *Coelebs in Search of a Wife*, secondo la leggenda, fu l’unico romanzo che la regina Vittoria ritenne degno di approvazione.

Con i diagrammi sul rapporto fra prestigio e popolarità, un primo arco del nostro progetto aveva trovato una sua naturale conclusione. Sebbene, contro le nostre intenzioni originali, ci fossimo fermati ben lontani dalla nozione di archivio¹⁹, la nostra operazionalizzazione del concetto di canone era stata tanto sorprendente quanto sostanziosa: aveva riportato la nozione alla concretezza, risolvendola in elementi più semplici come la popolarità e il prestigio, che in parole ancora più semplici sono il mercato e la scuola. All’interno di queste nuove coordinate, il canone rimane visibile come sempre, ma *perde la sua autonomia concettuale* per diventare il prodotto contingente di un incontro tra due forze contrapposte. Sono *queste forze*, allora, che necessitano di essere investigate più a fondo, se vogliamo saperne qualcosa di più del canone²⁰: nelle ricerche future si potranno facilmente aggiungere le tirature e

¹⁹ Nelle FIGURE 3.2-3.3, che avevano il loro punto di interruzione due o tre deviazioni standard sopra la media del campo, tutti gli autori di alto prestigio e di medio prestigio, e circa la metà di quelli ad alta popolarità, possono essere considerati canonici. Scendendo più “in basso”, la tripartizione del campo rimane visibile un po’ di più, e poi scompare. Quel che accade dopo è una domanda affascinante, ma per un altro saggio.

²⁰ O, più precisamente, *se si vuole decomporre il concetto di canone nei due elementi sottostanti di popolarità e prestigio*. Risulta qui utile paragonare la scelta epistemologica iniziale di questo progetto con quella di Algee-Hewitt e McGuff in *Between Canon and Corpus*. La principale

la presenza nelle biblioteche popolari ai parametri della popolarità, così come gli estratti dai libri di testo o le citazioni negli archivi di opere saggistiche ai parametri del prestigio²¹. Ad ogni nuova aggiunta guadagneremo qualcosa in più nella comprensione della natura composita del canone e della sua natura *storica*: il canone del 1770-1830 (e immaginiamo anche dei successivi 70-80 anni) fu il prodotto di un'età d'oro della borghesia europea, in un momento in cui gli imperativi del successo e dell'educazione sembravano compatibili tra loro, come si addiceva ad una classe dominante che, per la prima volta nella storia, si sentiva a proprio agio sia a scuola che sul mercato. Rendere visibile la duplicità di questo canone del XIX secolo, è il merito principale di queste sezioni iniziali²².

differenza non è fra testi (*Between Canon and Corpus*) e autori (*Canone/archivio*) – che potrebbe essere facilmente appianata – ma fra un'analisi basata sui network e una basata sui diagrammi cartesiani. I network sono più utili a *indagare le relazioni fra nodi individuali* (il cluster “ipercanonico” illustrato dalla FIGURA 3 dello studio, la singolare centralità di *Grapes of Wrath*, il distacco fra bestseller e altri gruppi), *ma non può collegare i nodi con qualcosa al di fuori della rete stessa*. I diagrammi cartesiani, incorporano il “fuori” nei loro assi (qui la popolarità e il prestigio), ma inevitabilmente *allentano le relazioni tra i singoli dati individuali* (in un diagramma non c'è nulla di simile ai link delle reti o alle misurazioni del *clustering*). Com'è ovvio non è una questione di quale metodo sia “migliore”: si tratta invece dei progetti di ricerca, i quali, volendo indagare diverse proprietà del sistema, scelgono gli strumenti di analisi di conseguenza.

²¹ Inutile aggiungere che simili misurazioni potrebbero essere discontinue e difficili da ottenere (le tirature, ad esempio), mentre altre (i libri di testo) potrebbero partire da una data significativamente tarda. Ma se la nozione di campo letterario dev'essere d'aiuto nel comprendere le diverse epoche e nazioni, fare ricorso alle misurazioni storiografiche più disparate sarà inevitabile: invece di sperare in una – chimerica – omogeneità delle fonti, dobbiamo imparare a rendere concettualmente compatibili dei dati eterogenei.

²² *Between Canon and Corpus* mostra quanto le cose siano cambiate da allora: il canone (o i canoni) sono tutti caratterizzati da «una differenziazione sistematica, se non da una contraddizione fra valore artistico e valore commerciale». È proprio questa differenziazione/contraddizione che è assente dalla regione “canonica” della FIGURA 3.3.

II. Caratteristiche morfologiche

4. Misurare la ridondanza

Per quanto differenti da quelli di Bourdieu sotto molti aspetti, i grafici presentati nella sezione precedente ne condividono la principale premessa metodologica: hanno un fondamento sociologico piuttosto che letterario²³. Per ottenere la **FIGURA 3.3** non c'è bisogno di aprire neanche un romanzo. In quanto storici della letteratura, comunque, noi *volevamo* aprire i romanzi, e scoprire se il loro destino sociale – popolare, di prestigio, entrambi, o nessuno dei due – avesse una qualche connessione con le loro caratteristiche morfologiche. Così, mentre lavoravamo ai diagrammi del campo letterario, ci siamo anche concentrati sulla composizione interna del *Chadwyck-Healey* e del campione più vasto dell'archivio. Qui il primo passo è consistito nel misurare la quantità di ridondanza e di informazione presente nel corpus. È convinzione diffusa, un luogo comune, che i lettori preferiscano i testi informativi a quelli ridondanti – da qui la sopravvivenza dei primi e l'estinzione dei secondi – ma volevamo metterla alla prova. Prendendo spunto dalla teoria dell'informazione, Mark Algee-Hewitt aveva misurato quella che si chiama “ridondanza di secondo grado” (la prevedibilità delle parole singole) apportando una modifica alla misura della quantità di informazioni di Shannon che determina il contenuto informativo di ciascun testo basandosi sulla prevedibilità del passaggio da una parola alla successiva all'interno di una gamma di transizioni possibili. Dato che la preposizione “of” è molto più spesso seguita dall'articolo “the” che da una negazione, ad esempio “no”, va da sé che la coppia “of the” è assai più comune di quella “of no”²⁴. Le **FIGURE 4.1** e **4.2** sintetizzano l'indagine di Algee-Hewitt.

²³ «Suggerisco che il problema di ciò che si chiama formazione del canone», scrive John Guilroy nella stessa ottica, «possa essere compreso più a fondo in termini di costituzione e distribuzione del capitale culturale, o più specificatamente, un problema di accesso ai mezzi di produzione e di consumo della letteratura». John Guilroy, *The Cultural Capital, the problem of literary canon formation*. Chicago U.P., Chicago (IL) 1995, p. ix.

²⁴ In questo pamphlet useremo i termini “ridondanza” e “ripetizione” in maniera pressoché intercambiabile, mettendoli in contrapposizione alla coppia “informazione” e “varietà”; per quanto si tratti di una semplificazione, non pensiamo che infici la qualità del lavoro né il tipo di risultati che abbiamo trovato. A tal riguardo, la relazione tra informazione e ridondanza è spesso definita “entropia”: abbiamo optato per definizioni diverse per rendere i diversi aspetti di questa ricerca comparabili il più possibile.

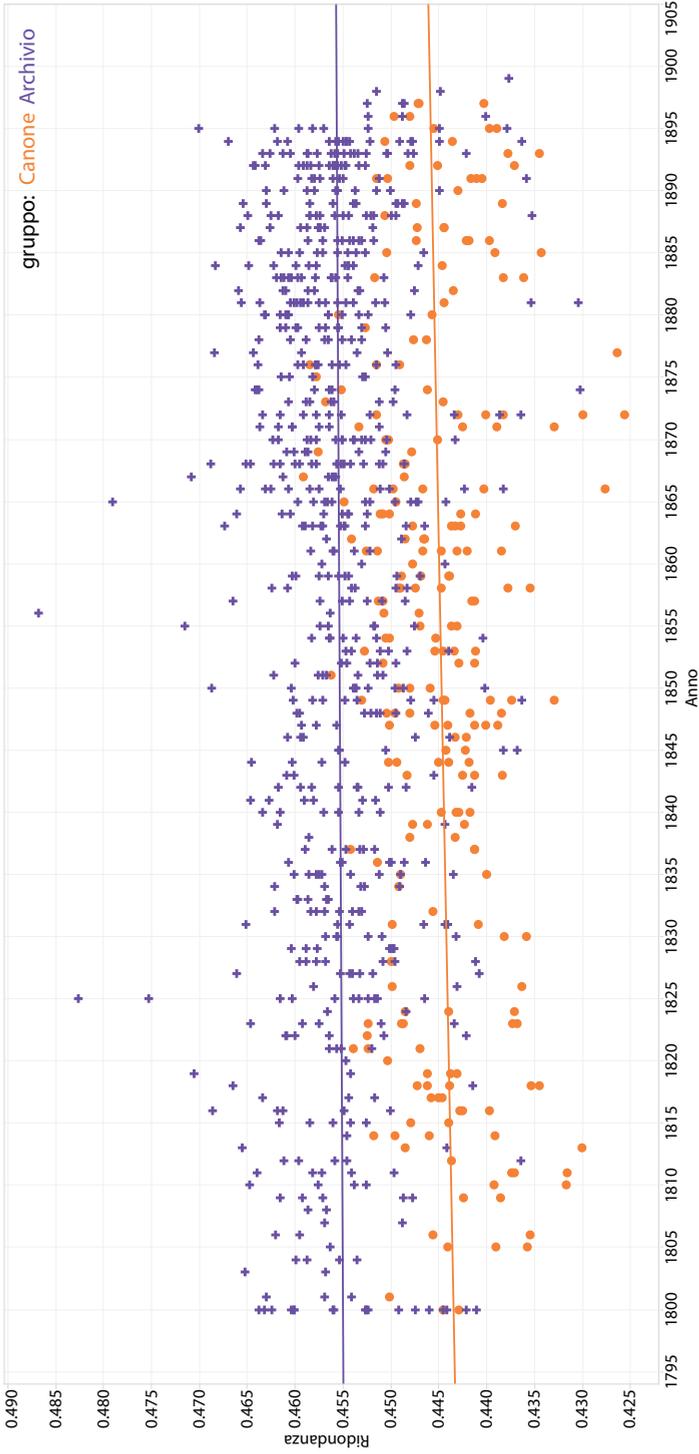


FIGURA 4.1 – Misurazione della ridondanza, 1800-1900.

Le croci viola indicano i romanzi d'archivio, i cerchi arancioni quelli canonici.



FIGURA 4.2 – La ridondanza nel XIX secolo: un diagramma sintetico.

Questa figura aggrega i dati della FIGURA 4.1 in due sub-corpora di canone e archivio. Ogni “rettangolo” include un quartile del gruppo, separato da una linea che indica i valori mediani del gruppo; i “baffi” che spuntano dalle forme rettangolari rappresentano i due quartili più estremi, i casi limite invece sono indicati ciascuno da un puntino.

La FIGURA 4.2 mostrava dei risultati sorprendenti: i tre quarti della *Chadwyck-Healey* erano meno ridondanti dei tre quarti dell'archivio. Si trattava di una separazione molto più netta di quella che immaginavamo di trovare. Non eravamo però ancora pienamente soddisfatti. La chiarezza del contrasto aveva semplicemente confermato un luogo comune: gli autori dimenticati utilizzavano il linguaggio in forma ridondante. Se non erano stati letti era perché non valeva la pena leggerli. E viceversa: apprezziamo Austen ancora oggi perché è il paradigma della precisione informativa, come mostra chiaramente lo specchietto della FIGURA 4.3.

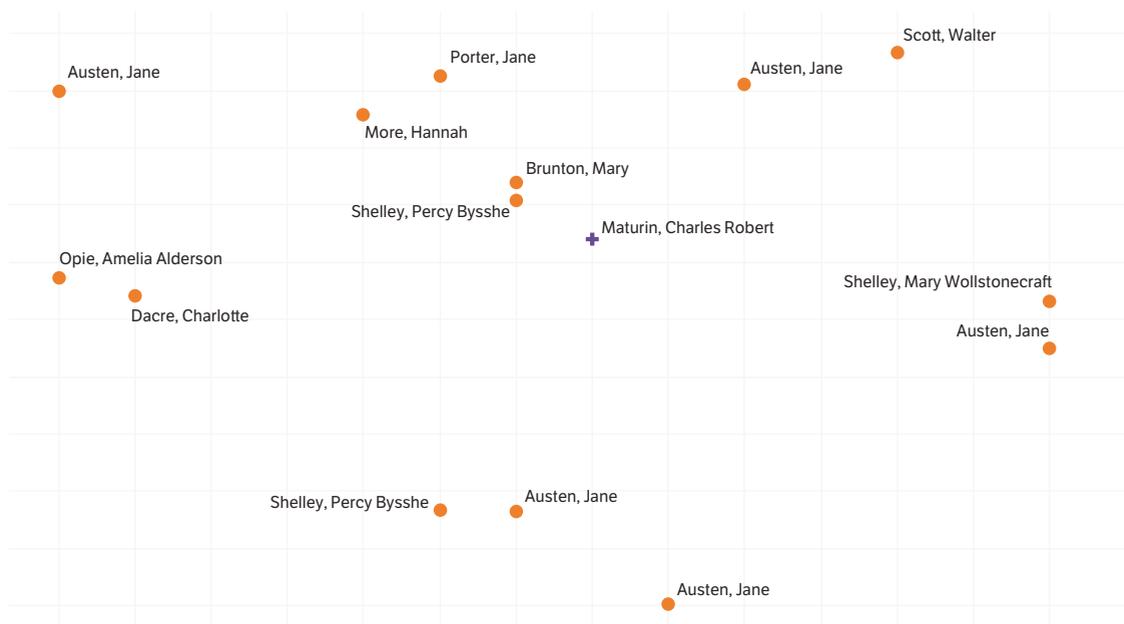


FIGURA 4.3 – Ridondanza estremamente bassa nella prima parte del XIX secolo.

Un romanzo nel quale non si ripetesse neanche una parola avrebbe una ridondanza pari a 0 e un grado d'informazione del 100%. Quest'informazione però non avrebbe alcun valore perché il libro diverrebbe presto incomprensibile. Il senso dipende sempre da una mescolanza di novità e di ripetizione: è per questo che i valori di questa figura oscillano sempre in uno spazio molto ridotto. Tuttavia le differenze all'interno di questo spazio sono comunque significative, come mostra l'ingrandimento dell'area in basso a sinistra della FIGURA 4.1.

Non c'è nulla di esaltante nel corroborare un luogo comune²⁵. Ora, però, c'è un secondo problema che viene a galla. Benché Algee-Hewitt avesse operazionalizzato il concetto di ridondanza e ottenuto risultati sorprendenti, non era ancora chiaro come dovessimo disaggregare i dati generali e osservare i risultati per verificare quale coppia di parole ricorresse e quale no. Siamo stati in grado di misurare la ridondanza, ma non riuscivamo veramente ad *analizzarla*: ci stavamo allontanando pericolosamente da quella interazione tra misurazione quantitativa e interpretazione qualitativa che è stata una costante del nostro lavoro sin dall'inizio. Qui, i valori statistici sembravano impermeabili a qualsiasi valore critico. Il “testo” creato estraendo cento dei bigrammi più frequenti da ogni romanzo del corpus componeva un grafico enorme con più di centomila caselle: “leggerle” era fuori questione (FIGURA 4.4).

of the 144	in the 67	to the 63	of his 34	of a 33	and the 31	on the 29	by the 28	said the 26
of the 114	in the 57	to the 521	of his 30	of a 30	the master 294	and the 26	with the 23	by the 22
of the 26	in the 9	to the 95	on the 8	of his 6	by the 6	of her 6	from the 5	at the 4
of the 94	in the 40	to the 36	of his 24	to be 19	and the 17	by the 17	on the 17	of a 16
of the 148	in the 78	to the 63	of his 36	of a 36	and the 35	on the 30	to be 30	he had 27
of the 74	to be 61	in the 57	it was 38	she had 36	of her 34	she was 32	to the 32	to her 30
of the 67	in the 40	sir ulick 34	to the 33	to be 29	he had 29	he was 27	of his 23	that he 22
of the 70	in the 49	of her 44	to the 40	to be 38	lady juliana 27	to her 27	and the 25	of a 24
of the 35	said i 23	in the 21	to the 19	i am 18	of my 17	and the 17	and i 17	that i 17
of the 38	in the 29	to be 27	of her 19	i am 18	to the 16	it was 16	she had 15	of a 13
of the 45	to be 42	in the 38	i am 29	of her 26	to the 25	it was 24	mr darcy 23	she was 21
of the 22	in the 14	mr glowry 8	and the 7	of a 6	to the 6	it is 5	on the 5	to be 5
of the 16	in the 9	to be 8	to the 5	i am 4	of a 4	a very 3	for the 3	at the 3
of the 34	to the 24	the marquis 21	in the 21	of his 21	and the 16	he had 15	to his 12	for the 11
of the 24	to the 19	in the 18	mrs villars 12	of her 11	she was 10	she had 10	on the 9	i am 8
of the 64	in the 41	to the 33	i have 28	it is 26	i am 20	on the 19	that i 18	and the 16
of the 60	in the 47	to the 33	to be 24	of her 22	had been 20	on the 19	she had 19	i have 16
of the 60	to the 32	in the 29	of his 23	to be 20	he had 18	and the 18	on the 17	he was 17
of the 42	to the 30	in the 28	said i 26	of a 21	and the 18	at the 18	i was 15	in a 15
of the 38	in the 19	on the 15	and the 13	to the 12	of my 10	i had 10	it was 9	i was 8
of the 162	in the 132	of her 131	to the 128	to her 97	of his 93	for the 68	from the 63	she had 55
of the 100	in the 62	to the 57	he had 44	he was 38	to be 36	it was 35	she had 35	of her 33
of the 75	in the 45	to the 41	he had 39	she had 35	to be 29	he was 29	she was 28	on the 27
of the 479	in the 332	to the 251	to be 219	of her 214	she had 207	she was 193	of a 186	it was 160
of the 116	to the 72	in the 67	to her 51	of her 49	to be 49	i am 40	she had 38	of a 36
of the 168	to the 70	in the 65	said the 52	and the 40	of a 34	of his 34	it is 33	to be 32
of the 130	in the 58	to the 58	of his 36	and the 31	on the 30	as he 27	by the 25	with the 23
of the 163	in the 68	to the 64	the earl 40	of his 38	and the 37	sir simon 36	by the 36	i will 35
of the 146	to the 55	in the 52	said the 37	of his 35	and the 30	on the 27	of a 27	from the 25

FIGURA 4.4 – Lettura dei bigrammi: 0.00003% dei dati.

Una sezione del foglio utilizzato per i calcoli alla base delle FIGURE 4.1 e 4.2. Sebbene i bigrammi in quanto tali siano perfettamente identificati, è pressoché impossibile “interpretare” quel che significano da un punto di vista statistico. Walser e Algee-Hewitt hanno osservato che i bigrammi assomigliano alle “progressioni demografiche” e alle “variazioni dei tassi di interesse” di Braudel: tutti fenomeni che non potevano essere percepiti passaggio-per-passaggio, la scala che di solito privilegiamo nelle nostre analisi.

²⁵ Ed è già la seconda volta: nella FIGURA 1.2 il fatto che il canone precedesse regolarmente l'archivio di 15-20 anni sembra “provare” un altro luogo comune secondo il quale i grandi scrittori aprono la strada che altri seguiranno.

Un approccio più tecnico – seguire cioè la curva di declino dei sintagmi più frequenti – è risultato ugualmente improduttivo: i bigrammi più frequenti (“there is”, “I am”, “to the”) mostravano una frequenza simile in tutti i testi, presentando una trascurabile quantità di variazioni nelle tracce finali della curva. In più c’erano in ogni romanzo così *tanti* bigrammi che i loro effetti si manifestavano attraverso un numero enorme di cambiamenti estremamente piccoli: in un testo relativamente breve di 65.500 parole, ad esempio, c’erano 66.499 bigrammi, 40.000 dei quali non si ripetevano mai. E per quanto il numero delle parole condivise fra due testi fosse ragguardevole – circa 3000, 4000 – i bigrammi condivisi erano di solito meno di 1000: troppo pochi per una solida analisi comparativa.

Ci sembrava di aver creato una versione fatta in casa del principio di indeterminazione; quanto più precisamente misuravamo la ridondanza, tanto più diveniva difficile determinare “dove” si trovasse effettivamente. La ridondanza agiva su una scala onnipervasiva, ed evidentemente decisiva nello stabilire il destino dei libri: eppure l’intero processo avveniva talmente al di sotto della lettura cosciente da essere praticamente invisibile. In futuro, in un futuro forse anche prossimo, di questo aspetto potrebbe occuparsi la psicologia sperimentale, ma per ora ci siamo affidati a uno strumento linguistico classico di misurazione della varietà lessicale noto come “type-token ratio” (o rapporto tipo/occorrenza)²⁶. Minore la ridondanza in un testo – così ragionavamo – maggiore la sua varietà: convesso e concavo. Avremmo dovuto ottenere un’immagine che doveva essere l’esatto opposto della FIGURA 4.2. Così abbiamo fatto i nostri calcoli e il risultato è la FIGURA 4.5.

Le FIGURE 4.2 e 4.5, l’una accanto all’altra, producevano il seguente paradosso: il canone risultava meno ripetitivo (quindi molto più vario) dell’archivio al livello delle coppie di parole e dell’intero testo, ma meno vario (e quindi più ripetitivo) a livello di singole parole e dei segmenti di 1000. Il fatto che il testo a diverse scale si comportasse diversamente non costituiva di per sé una sorpresa: due dei nostri pamphlet, *Lo stile al livello della frase* e *Sui paragrafi*, si sono occupati proprio di questo [si vedano i Capp. 4 e 6]. In quei casi, tuttavia, ogni scala era stata associata a caratteristiche che riguardavano

²⁶ Questo è il modo in cui la *Longman Grammar of Written and Spoken English* definisce il rapporto type-token: «La relazione fra il numero dei vocaboli (*types*) differenti e il numero totale delle occorrenze (*Tokens*) è detta *type-token ratio* (o TTR)». Si veda Biber, Johansson, Leech, Conrad, Finegan, *Longman Grammar of Spoken and Written English*, Harlow 1999, pp. 52-3.

tutt'altro: le frasi con lo stile e i paragrafi con i temi. Invece qui i tratti che chiamiamo in causa sono strettamente collegati. Come si possono travasare i risultati nel momento in cui si passa da due a mille parole? Ma questo “come” va inteso alla lettera, non con valore enfatico, e quindi: quale meccanismo testuale potrebbe, concretamente, trasformare il primo risultato nel secondo?

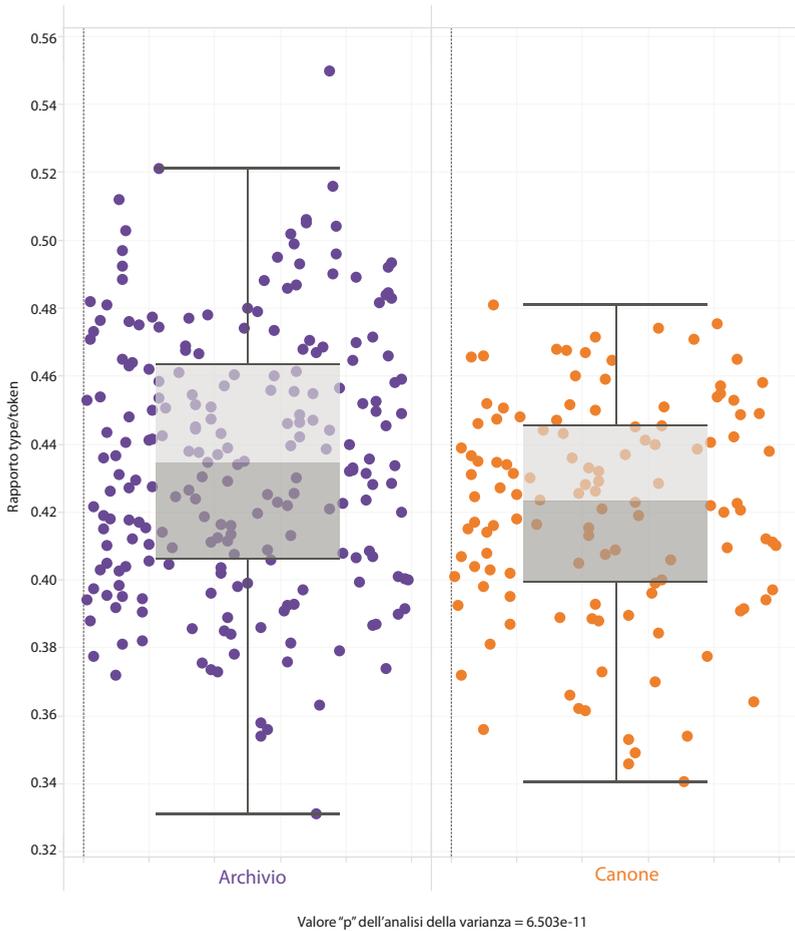


FIGURA 4.5 – Misurazione della varietà: diagramma sintetico del rapporto type-token. Sebbene la distinzione fra i due *subcorpora* sia in questo caso molto meno netta che nella FIGURA 4.2 il risultato è di fatto notevole. La FIGURA 4.2 aveva pienamente confermato le nostre attese sul canone e l'archivio, mentre questo grafico le contraddice in pieno: il lessico del canone non è più vario di quello dell'archivio, ma al contrario lo è significativamente meno. (La procedura seguita per determinare il rapporto type-token è descritta nella nota 28, all'inizio della prossima sezione).

Algee-Hewitt ha risolto questo problema “traducendo” tutte le parole sotto forma di parti del discorso, riformulando così la ridondanza sotto forma di *categorie* di bigrammi piuttosto che di coppie separate: “clever little” e “first cruel”, ad esempio, diventano entrambi locuzioni aggettivali; “a condition” e “the kitchen” diventano «articolo-nome», e così via. Ricalcolando tutto in termini di «ridondanza grammaticale» diventava possibile identificare i tipi di bigrammi appartenenti al canone o all’archivio (FIGURE 4.6 – 4.7)²⁷.

Questa volta, i due subcorpora rivelavano due centri di gravità assai diversi fra di loro: l’archivio era dominato dai nomi, mentre il canone aveva una presenza assai larga di funzioni verbali (coniunzioni, articoli, proposizioni). L’archivio indulgeva in apposizioni (“il conte Goldstein”, “zio Gerard”), nella puntigliosità verso i luoghi e le persone (“in Irlanda”, “da Shirley”), in un’ab-

Preposizione-nome proprio (IN_NNP): to Shirley; in Ireland
 Aggettivo-aggettivo (JJ_JJ): young happy; first cruel
 Nome-aggettivo (NN_JJ): child incapable; nomenclature peculiar
 Nome-nome (NN_NN): iron will; evening sky
 Nome-nome proprio (NN_NNP): count Goldstein; uncle Gerard
 Nome-nome plurale (NN_NNS): iron bars; autumn tints
 Nome proprio-preposizione (NNP_IN): Alps of; Shelburne upon
 Nome proprio-nome (NNP_NN): Agnes’ wedding; Manchester cotton
 Nome proprio-nome plurale (NNP_NNS): Cumberland coasts, Hector’s lodgings
 Nome-pronome (NN_PRP): tail itself, driver himself.

FIGURA 4.6 – Bigrammi grammaticali più caratteristici: archivio.

Congiunzione-gerundio (CC_VBG): and walking; and taking
 Determinativo-aggettivo (DT_JJ): the silly; an eventful
 Determinativo-nome (DT_NN): a condition; the kitchen
 Determinativo-nome plurale (DT_NNS): the environs; the travelers
 Preposizione-determinativo (IN_DT): at the; in a
 Aggettivo-nome plurale (JJ_NNS): folded arms; harsh features
 Nome-preposizione (NN_IN): account of; sense of
 Nome plurale-preposizione (NNS_IN): grains of; years of
 Pronome possessivo-nome plurale (PRP\$_NNS): their excursions; our girls

FIGURA 4.7 – Bigrammi grammaticali più caratteristici: canone.

²⁷ Per questa parte del lavoro Algee-Hewitt ha usato il Part-of-Speech Tagger di Stanford; le abbreviazioni chiuse tra parentesi (IN_NNP ecc.) sono in ogni caso utilizzate dal Trebank project (<http://www.cis.upenn.edu/~trebank/>) dell’Università della Pennsylvania.

bondanza generale di nomi propri (“gli alloggi di Hector”, “Shelburne upon”) e in definitiva ci consegnava un indizio sul suo alto livello di ridondanza; “il conte Goldstein” e “Shelburne upon” potevano non comparire molto spesso in un romanzo, ma quando lo facevano, le due parole tornavano pressoché sempre insieme, facendo aumentare la ridondanza del testo. Lo stesso valeva per le costruzioni in cui è presente un complemento attributivo: “volontà d’acciaio” e “tinte autunnali”. Non era la risposta alla nostra domanda, ma era un inizio. Per riuscire a risolvere l’altro lato del paradosso, siamo tornati all’indice di diversità lessicale, al rapporto type-token.

5. “Ma non potevo andarmene”

Nel caso del rapporto type-token, la prima cosa che andava fatta era trovare un metodo di analisi adeguato a un corpus composto da romanzi che per la maggior parte non erano ristampati da uno o due secoli, il che rendeva il riconoscimento ottico difficile con la conseguenza di invalidare, potenzialmente, tutti i calcoli che ne sarebbero derivati. Ryan Heuser, che nelle fasi embrionali del progetto ci ha spinti a considerare il rapporto type-token, trovò un metodo di misurazione affidabile per testi dalla qualità di stampa molto diversa²⁸. Una volta ottenuti i risultati, abbiamo iniziato a osservare i valori bassi del rapporto type-token per comparare quello specifico tipo di ripetiti-

²⁸ Heuser aveva cominciato col creare un dizionario assai ampio dell’inglese romanzesco – 232.845 parole distinte – e col dividere tutti i testi in segmenti da 1000 “parole da dizionario” (in verità i segmenti erano lunghi fra le 1000 e le 1500 parole, a seconda di quante parole “non da dizionario” – cose come errori di scansione, *hapax*, ecc. – erano presenti). Dato che il numero delle parole occorrenze (token) era di 1000, dalla divisione del numero dei vocaboli (type) di ogni segmento per 1000 si ottiene il risultato di ciascun segmento-base e dalla loro media il rapporto type-token dell’intero testo.

La funzione è stata scritta con due parametri: «*slice_len*» (la lunghezza del segmento, fissata a 1000) e «*force_english*» (per le parole che non appartengono al dizionario di lingua inglese, impostato su «falso»). La ragione alle spalle del parametro «*force_English*» – che esclude tutte le parole non inglesi – risiede nel fatto che senza di esso l’archivio avrebbe creato un rapporto type-token più alto per meri errori di scansione del testo. Viceversa se il sistema avesse accettato tutte le parole non inglesi avrebbe prodotto, sempre per i problemi di scansione, un rapporto type-token molto basso. Se il segmento avesse dovuto espandersi oltre le 1500 parole “reali” per trovarne 1000 inglesi, allora avrebbe potuto privilegiare termini più brevi e facili da identificare nella scansione, questi termini sono anche i più comuni nella lingua: ciò avrebbe abbassato i valori del rapporto type-token. Alla prova dei fatti, questi risultati indesiderati si sono annullati a vicenda.

vità con la ridondanza calcolata da Algee-Hewitt. Sapevamo già dalla FIGURA 4.6 che una bassa varietà lessicale sarebbe andata spesso a braccetto con i testi canonici, e infatti nella collezione *Chadwyck Healey*, che ammontava al 20% circa del corpus totale, la frequenza crebbe fino al 50% per i 500 segmenti con i valori di rapporto type-token più bassi (quando invece la media era di 3,2% per i 500 segmenti con valori alti). Tra i cinquanta libri con i valori più bassi, circa la metà di provenivano dalla *Chadwyck-Healey*: molti libri per bambini (*Alice*; *Through the Looking Glass*; *The Water Babies*; *Black Beauty*; *Little Lord Faunleroy*; *Island's Night's Entertainments...*) dieci dai romanzi di Trollope (*The Last Chronicle of Borset*; *Phineas Finn the Irish Member*; *Can You Forgive Her?*; *Tha Eustace Diamonds...*), più due romanzi irlandesi (*Castle Rackrent* di Edgeworth e *Father Tom and the Pope* di Samuel Ferguson; *The Absentee* seguiva a breve distanza). In sé, questo mix non era particolarmente rappresentativo del canone (in qualunque senso lo si voglia intendere). Sembrava più significativo che i risultati dei testi della *Chadwyck-Healey* rimanessero bassi per tutto il XIX secolo (FIGURA 5.1) e che questa tendenza coinvolgesse alcuni dei più raf-

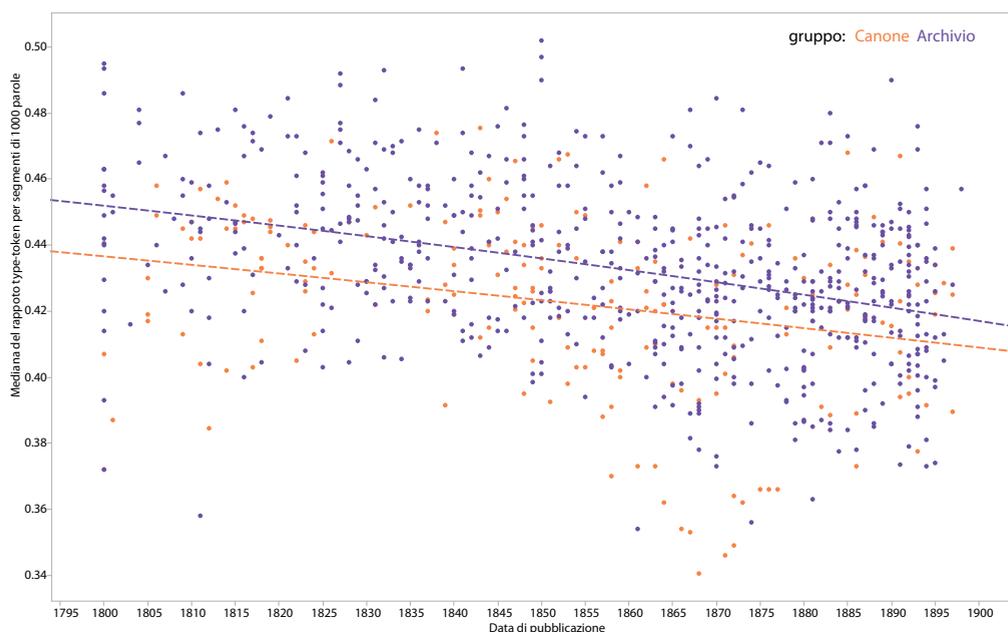


FIGURA 5.1 – Rapporto type-token 1800-1900.

L'influenza della letteratura per l'infanzia nell'abbassare il rapporto type-token è visibile nel periodo tra il 1860 e il 1880. In generale, tuttavia, il rapporto type-token rimane stabile per tutto l'Ottocento sia per il canone che per l'archivio.

finati prosatori secolo: tutta Austen era sotto della media del corpus (con *Persuasion*, *Sense and Sensibility*, e *Mansfield Park* inferiori al 20%); tutto Dickens era sotto della media (con *Little Dorrit*, *A Tale of Two Cities*, *David Copperfield*, *Our Mutual Friend*, *Bleak House*, e *Great Expectations* inferiori al 20%); tutta George Eliot era sotto la media e *Adam Bede* conteneva il passaggio il più basso rapporto type-token dell'intero secolo.

Ora, *Adam Bede* è un romanzo anomalo per un risultato del genere perché sono presenti le famose riflessioni di Eliot sulla pittura fiamminga: un manifesto dell'estetica della precisione e della varietà, *scritto* con straordinaria precisione e varietà (FIGURA 5.2).

Le prime cento parole di questo passaggio mostrano un rapporto type-token pari a 79: più alto di qualsiasi altro, in qualsiasi registro, mai commentato dalla *Longman Grammar*. Ma poi, più avanti nel romanzo, lo stile di Eliot si muove sull'estremo opposto. (FIGURA 5.3).

It is for this rare, precious quality of truthfulness that I delight in many Dutch paintings, which lofty-minded people despise. I find a source of delicious sympathy in these faithful pictures of a monotonous homely existence, which has been the fate of so many more among my fellow-mortals than a life of pomp or of absolute indigence, of tragic suffering or of world-stirring actions. I turn, without shrinking, from cloud-borne angels, from prophets, sibyls, and heroic warriors, to an old woman bending over her flower-pot, or eating her solitary dinner, while the noonday light, softened perhaps by a screen of leaves, falls on her mob-cap, and just touches the rim of her spinning-wheel, and her stone jug, and all those cheap common things which are the precious necessaries of life to her—or I turn to that village wedding, kept between four brown walls, where an awkward bridegroom opens the dance with a high-shouldered, broad-faced bride, while elderly and middle-aged friends look on, with very irregular noses and lips, and probably with quart-pots in their hands, but with an expression of unmistakable contentment and goodwill. "Foh!" says my idealistic friend, "what vulgar details!"

FIGURA 5.2 – «This rare, precious quality of truthfulness».

«È per questa rara, preziosa qualità di veridicità che apprezzo molti dipinti olandesi, disprezzati dalle persone altezzose. Trovo una fonte di deliziosa vicinanza in queste fedeli immagini di un'esistenza familiare monotona, che è stata la sorte di tanti dei miei fratelli mortali, e non in quelle di una vita di ricchezza o di assoluta indigenza, di tragica sofferenza o di azioni che scuotono il mondo. Mi sposto, senza sentirmi rimpicciolire, dagli angeli appollaiati sulle nuvole, dai profeti, le sibille e gli eroici guerrieri, a una vecchiaia che si china sul suo vaso di fiori o sulla sua cena solitaria mentre la luce di mezzogiorno, ammorbidita forse da uno schermo di foglie, le cade sulla cuffia e tocca appena il bordo dell'arcoliaio, la brocca di pietra e tutte quelle cose comuni poco costose che per lei sono le preziose necessità della vita. Oppure mi rivolgo a un matrimonio di villaggio, tenuto tra quattro pareti marroni, dove uno sposo imbarazzato apre le danze con una sposa dalle spalle e dal viso larghi mentre gli anziani e gli amici di mezza età guardano, con nasi e labbra molto irregolari, e probabilmente con un quartino in mano, ma con un'espressione inconfondibile di contentezza e simpatia. "Puah!" dice il mio amico idealista, "che dettagli volgari!"» [FIGURA 5.2: trad. it. nostra]

came all of a sudden, as I was lying in the bed, and it got stronger and# stronger#... I# longed so to go back again... I# could n't* bear being so# lonely, and# coming to# beg for want. And# it# gave me strength and# resolution to# get up and# dress myself. I# felt I# must do it#... I# did n't* know how... I# thought I#d find a# pool, if I# could#, like that other, in# the# corner of# the# field, in# the# dark. And# when the# woman went out, I# felt# as# if# I# was# strong enough to# do# anything... I# thought# I# should get# rid of# all# my misery, and# go# back# home, and# never let'em know# why I# ran away. I# put on my# bonnet and# shawl, and# went# out# into the# dark# street, with the# baby under my# cloak; and# I# walked fast till I# got# into# a# street# a# good way off, and# there was# a# public, and# I# got# some warm stuff to# drink and# some# bread. And# I# walked# on# and# on#, and# I# hardly felt# the# ground I# trod on#; and# it# got# lighter, for# there# came# the# moon-- O, Dinah, it# frightened me# when# it# first looked at me# out# o#' the# clouds-- it# never# looked# so# before; and# I# turned out# of# the# road into# the# fields, for# I# was# afraid o#' meeting anybody with# the# moon# shining on# me#. And# I# came# to# a# haystack, where I# thought# I# could# lie down and# keep myself# warm# all# night. There# was# a# place cut into# it#, where I# could# make me# a# bed#; and# I# lay comfortable, and# the# baby# was# warm# against me#; and# I# must# have gone to# sleep for# a# good# while, for# when# I# woke it# was# morning, but not very light, and# the# baby# was# crying. And# I# saw a# wood a# little way# off#... I# thought# there#d perhaps be a# ditch or a# pond there#... and# it# was# so# early I# thought# I# could# hide the# child there#, and# get# a# long way# off# before# folks was# up#. And# then I# thought# I#d go# home#-- I#d get# rides in# carts and# go# home#, and# tell'em I#d been to# try and# see for# a# place#, and# could# n't* get# one. I# longed# so# for# it#, Dinah#-- I# longed# so# to# be# safe at# home#. I# do# n't* know# how# I# felt# about the# baby#. I# seemed to# hate it#-- it# was# like# a# heavy weight hanging round my# neck; and# yet its crying# went# through me#, and# I# dared n't* look at# its# little# hands and# face. But# I# went# on# to# the# wood#, and# I# walked# about#, but# there# was# no water#... Hetty shuddered. She was# silent for# some# moments, and# when# she# began again#, it# was# in# a# whisper. `` I# came# to# a# place# where# there# was# lots of# chips and# turf, and# I# sat down# on# the# trunk of# a# tree to# think what I# should# do#. And# all# of# a# sudden# I# saw# a# hole under# the# nut-tree*, like# a# little# grave. And# it# darted into# me# like# lightning-- I#d lay# the# baby# there#, and# cover it# with# the# grass and# the# chips#. I# could# n't* kill it# any other# way#. And# I#d done it# in# a# minute; and# O#, it# cried so#, Dinah#-- I# could# n't* cover# it# quite up#-- I# thought# perhaps# somebody `ud* come and# take care of# it#, and# then# it# would n't* die. And# I# made haste out# of# the# wood#, but# I# could# hear it# crying# all# the# while#; and# when# I# got# out# into# the# fields#, it# was# as# if# I# was# held fast#-- I# could# n't* go# away#, for# all# I# wanted so# to# go#. And# I# sat# against# the# haystack# to# watch if# anybody `ud* come#: I# was# very# hungry, and# I#d only a# bit of# bread# left; but# I# could# n't* go# away#. And# after ever such a# while#-- hours and# hours#-- the# man came#-- him in# a# smock-frock*, and# he looked# at# me# so#, I# was# frightened#, and# I# made# haste# and# went# on#. I# thought# he# n't* was# going to# the# wood#, and# would# perhaps# find# the# baby#. And# I# went# right on#, till# I# came# to# a# village, a# long# way# off# from the# wood#; and# I# was# very# sick, and# faint, and# hungry#. I# got# something to# eat there#, and# bought a# loaf. But# I# was# frightened# to# stay. I# heard the# baby# crying#, and# thought# the# other# folks# heard# it# too,-- and# I# went# on#. But# I# was# so# tried, and# it# was# getting towards dark#. And# at# last, by the# roadside there# was# a# barn-- ever# such# a# way# off# any# house-- like# the# barn# in# Abbot's Close; and# I# thought# I# could# go# in# there# and# hide# myself# among the# hay and# straw, and# nobody `ud* be# likely to# come#. I# went# in#, and# it# was# half full o# trusses of# straw#, and# there# was# some# hay#, too#. And# I# made# myself# a# bed#, ever# so# far behind, where# nobody# could# find# me#; and# I# was# so# tired and# weak, I# went# to# sleep#.... But# oh, the# baby#'s crying# kept waking me#; and# I# thought# that# man# as# looked# at# me# so# was# come# and# laying hold of# me#. But# I# must# have# slept a# long#

FIGURA 5.3 – Il passo più ripetitivo dell'Ottocento: la confessione di Hetty in *Adam Bade*.

Il cancelletto indica che una parola è stata ripetuta all'interno di un segmento dato, gli asterischi invece denotano parole che non fanno parte del "dizionario" utilizzato per elaborare i calcoli. Alcuni aspetti di questi e di altri passaggi sono prodotti del *parser* di Stanford, che considera le contrazioni delle frasi negative – ad esempio "n't" alla fine di espressioni come "couldn't" – come una parola a parte.

Questo passaggio del romanzo di Eliot contiene un momento centrale del romanzo: quello della confessione di Hetty a Dinah, quando ricorda di aver abbandonato la sua creatura nei boschi e di aver aspettato la sua morte: *waiting for its death* (è significativo il fatto che Hetty utilizzi il pronome neutro inglese “its” per il bambino). “Aspettare” non è davvero la parola giusta (FIGURA 5.4).

And# I# made haste out# of# the# wood#, but# I# could# hear it# crying# all# the# while#; and# when# I# got# out# into# the# fields#, it# was# as# if# I# was# held fast#-- I# could# n't* go# away#, for# all# I# wanted so# to# go#. And# I# sat# against# the# haystack# to# watch if# anybody# `ud* come#: I# was# very# hungry, and# I#'d only a# bit of# bread# left; but# I# could# n't* go# away#.

FIGURA 5.4 – «But I could hear it crying all the while».

Da un punto di vista grammaticale, la cosa che più colpisce di queste frasi è la marea di forme verbali flesse con Hetty come soggetto: *I made haste... I could hear... I got out... I was held fast... I couldn't go away... I wanted... I sat... I was... I had... I couldn't...* Nell'analisi narrativa, le forme verbali sono di solito viste come indice di “azione”, ed è comprensibile che sia così. Ma qui, mettendo in stridente dissonanza grammatica e semantica, esse significano invece paralisi: Hetty vuole disperatamente “andare via” ma non può. È proprio come non poteva abbandonare il luogo fisico in cui era ambientato l'episodio non poteva rinunciare alle *parole* che lo descrivevano. Non può *dimenticare*: ed è da qui che viene la ripetizione. Meglio: non può né dimenticare *né dire fino in fondo ciò che era avvenuto*. Un classico esempio di opposizione fra “ripetizione” ed “elaborazione”: Hetty continua a ripetere la stessa cosa di continuo perché non riesce a dire l'unica cosa che conta davvero. La parola “morte” non è *mai* ripetuta e appare solo in una costruzione indiretta e ambigua verso la fine del brano²⁹.

²⁹ «But it was morning, for it kept getting lighter, and I turned back the way I'd come. I couldn't help it, Dinah; it was the baby's crying made me go--and yet I was frightened to death. I thought that man in the smock-frock 'ud see me and know I put the baby there.» («Ma era mattina, perché il cielo si schiariva sempre di più; io ripresi la strada già percorsa e feci ritorno. Non potevo fare diversamente, Dinah: il pianto del bambino mi richiamava ed ero spaventata da morire.» *Adam Bede*, trad. it F. Fizi, Castelvetti, Roma 2011, Kindle posizione 7396). Da notare come il concetto di “morte” è riferito a Hetty invece che al bambino.

Perché la ripetizione? Perché c'è stato un trauma e la ripetizione è il modo più efficace per esprimerlo attraverso il linguaggio: un imprigionamento nelle proprie parole la cui forza enigmatica dà conto del perché Eliot, nonostante il suo amore per il dettaglio analitico, potesse scrivere il passo più ripetitivo del XIX secolo. In più, la confessione di Hetty mette in risalto la componente essenzialmente *orale* del rapporto type-token. Accanto alla pagina di Eliot, i due segmenti con la più bassa densità lessicale sono anch'essi confessioni: lo scambio di culle in *Ennui*³⁰ di Edgeworth, e l'amore in *Last Chronicle of Barse*³¹ di Trollope. Nella medesima area a bassa densità lessicale troviamo: i passaggi delle storie per bambini (con una tipica narrazione naturale), i romanzi irlandesi (specializzati nell'imitazione del parlato) e una moltitudine di casi di sticomitie, in cui si replicano i dialoghi piccolo-borghesi, provenienti da Trollope³². Ci sono scene di processi (*The Ordeal of Richard Feverel*, *The Heart of Mid-Lothian*, il *Tale of Briefness Barriester* di William Scargill) scontri ideologici (*Marius the Epicurean*), una visione estatica del "comunismo della felicità" (*Lady Laura* di Mary Christie)³³ e,

³⁰ «I thought, how happy he would be if he had such a fine babby as you; dear; and you was a fine babby to be sure; and then I thought, how happy it would be for you, if you was in the place of the little lord: and then it came into my head, just like a shot, where would be the harm to change you?». («Ho pensato a quanto sarebbe bello avere un bimbo carino come te, caro; e tu eri un bimbo carino di sicuro; e ho pensato quanto sarebbe stato carino, per te, essere al posto di un piccolo lord: e poi mi è venuto in mente come un lampo: che male avrei fatto a scambiarti?»).

³¹ «You are so good and so true, and so excellent, – such a dear, dear, dear friend that I will tell you everything, so that you may read my heart. I will tell you as I tell mamma,– you and her and no one else; – for you are the choice friend of my heart. I can not be your wife because of the love I bear for another man». («Siete così carino e sincero – un caro, caro, caro amico. E vi dirò ogni cosa, così che possiate leggere nel mio cuore. Ve la dirò come lo direi a mamma – a voi, a lei e a nessun altro. Siete l'amico che ho scelto nel mio cuore. Non posso essere vostra moglie per l'amore che provo per un altro uomo»).

³² «Do you think that I am in earnest?» “Yes, I think you are in earnest.” “And do you believe that I love you with all my heart and all my strength and all my soul?” “Oh, John!” “But do you?” “I think you love me.” “Think!”». («Pensate che io sia nel giusto?” “Sì penso che voi siate nel giusto!” “E credete che vi ami con tutto il mio cuore, tutta la mia forza e la mia anima?” “Oh John” “Io credete?” “Penso mi amiate” “Penso!”»).

³³ «All are not equally happy; all can not be equally happy. But there is a sort of communism possible in happiness. The unhappy have a claim upon the happy; the happy have a debt towards the unhappy.” “But how can one share one’s happiness with others? It seems to me impossible. It is what I have most wished to do, but I see no way in which it can be done.”

infine, una grande invettiva contro il denaro (*A very old question* di Thomas Pemberton)³⁴. Ci sono personaggi che parlano troppo per il desiderio di compiacere (*Emma*), o perché, come Van Helsing in *Dracula*, hanno bisogno di sperimentare le prove in continuazione. Difficilmente poteva essere un caso, hanno concluso Allison e Gemma, che i nostri segmenti di 1000 parole con il valore più basso (e appartenenti al canone) fossero nella stessa gamma di densità che aveva la conversazione nella *Longman Grammar*: per quest'ultima la media risultava del 30%, per i nostri 500 segmenti meno densi il valore oscillava tra il 27 e il 33%.

Ci eravamo rivolti al rapporto type-token nella speranza che questo indice ci consentisse di tornare all'analisi testuale e non siamo rimasti delusi: i punteggi più bassi catturavano gli aspetti cruciali della struttura narrativa, segnalando i traumi, l'intensità e l'oralità. Cosa avremmo trovato nei punteggi alti?

“In one sense certainly you can not share your happiness, and you can not give it away. It is essentially your own, a development of your being, a part of yourself that you may not alienate.”». («Non tutti sono ugualmente felici; non si può essere tutti ugualmente felici. Ma c'è una sorta di comunismo nella felicità. Gli infelici hanno delle recriminazioni sui felici, i felici hanno un debito con gli infelici.» “Ma come si può condividere la propria felicità con un infelice? Mi sembra impossibile, è ciò che vorrei fare più di ogni altra cosa, ma non vedo via alcuna per cui possa esser fatto.” “in un certo senso non puoi condividere la felicità, e certamente non puoi darla via. Fa essenzialmente parte di te, una parte di te stesso che certamente non puoi alienare.”»).

³⁴ “Money!” she cried derisively.” Money! What is money to the trouble which has torn my heart ever since I have been married! What is money to those who thirst for love! I never wanted money; without money I was strong and happy; since I have had it I have been weak and miserable. Money broke down my poor father, and it was for money that Percy married, deceived, and has forsaken fine. Thank God that the wretched money has gone”». («Il denaro!” urlò derisoriamente “il denaro! Che cos'è il denaro in confronto al dramma che ha preso il mio cuore da quando mi sono sposata! che cos'è il denaro per coloro che sono assetati d'amore! Non ho mai voluto denaro; senza il denaro ero forte e felice; da quando ce l'ho sono triste e miserabile. Il denaro ha abbattuto il mio povero padre, ed è stato per denaro che Percy si è sposato, ha ingannato, e tradito con leggerezza.”»).

6. “Delle sottili feritoie con cannoni dalla bocca larga”.

La **FIGURA 6.1** mostra i dieci romanzi con gli indici più alti densità lessicale (type-token) di tutto il corpus. La **FIGURA 6.2** mostra il passaggio che ha ottenuto il risultato più alto, tratto dall'*Arthur Montague, or an Only son at Sea*, di Edward Hawker.

Se la posizione privilegiata del canone è sempre legata al privilegio linguistico – Dario Fo, premio Nobel per la letteratura nel 1997, scrisse una commedia intitolata *L'operaio conosce 300 parole, il padrone 1000, per questo è il padrone* –, allora gli autori del canone dovrebbero avere un linguaggio molto più variegato degli autori dimenticati. Tuttavia, se si esamina la diversità lessicale con il rapporto type-token, è vero il contrario. «Tutta la terminologia dell'estetica resta rinchiusa entro un rifiuto primigenio del *facile*» scrive Bourdieu «le opere “volgari” [...] suscitano disagio e disgusto per i metodi di seduzione»³⁵. *Facile* il linguaggio di Hawker? seducente? Semmai, il contrario. Una dicotomia come quella volgare/raffinato il volgare e il raffinato non sarà mai sufficiente a spiegare il legame tra l'archivio e valori alti del rapporto type-token. Dovevamo guardare altrove.

Come è già successo altre volte nel corso di questa ricerca, abbiamo trovato una risposta nella linguistica del corpus. Questa volta era nel concetto di “registro”: gli “obiettivi comunicativi e i contesti situazionali” dei messaggi descritti da Douglas Biber e Susan Conrad in *Register, Genre, and Style*³⁶.

Edward Duros, *Otterbourne; A Story of the English Marches*, 1832
 Edward Hawker, *Arthur Montague, or, An Only Son at sea*, 1850
 Emma Robinson, *The Armourer's Daughter: or, The Border Riders*, 1850
 William Lennox, *Compton Audley; or, Hands Not Hearts*, 1841
 Mary Anne Cursham, *Norman Abbey: A Tale of Sherwood Forest*, 1832
 William Maginn, *Whitehall; or, The Days of George IV*, 1827
 Thomas Surr, *The Mask of Fashion; A Plain Tale, with Anecdotes Foreign and Domestic*, 1807
 James Grant, *The Scottish Cavalier: An Historical Romance*, 1850
 Cecil Clarke, *Love's Loyalty*, 1890
 Jane West, *Ringrove, or Old Fashioned Notions*, 1827

FIGURA 6.1 – Rapporto type-token alto, ovvero il trionfo dell'archivio.

³⁵ Pierre Bourdieu *La distinzione. Critica sociale del gusto*, cit., p. 482-483.

³⁶ Douglas Biber e Susan Conrad, *Register, Genre, and Style*, Cambridge U.P., Cambridge 2009, p. 2.

La letteratura in laboratorio

then cut through some acres of refreshing greensward, studded with the oak, walnut, and hawthorn, ascended a knoll, skirted an expansive sheet of water; afterwards entering an avenue of noble elms, always tenanted* by a countless host of cawing* rooks whose clamorous conclaves* interrupted the stillness that reigned around, and whose visits to adjacent corn-fields* of inviting aspect raised the ire and outcry of the yelling VOL. I. C urchins employed to guard them from depredation. Emerging from this arched vista, a near view was obtained of the mansion, approached through a thick luxuriant shrubbery of full-grown evergreens. It was a straggling stone structure of considerable size and doubtful architecture, having on either side an ornamental wing, surmounted by glazed cupolas*, and indented below with niches containing statues and vases alternate. The face of the building displayed a row of fine Corinthian pillars-- their capitals screened by wire-work* shields, to defend them from the injurious intrusions of the feathery tril* > e, who ever chirped* and hovered about the forbidden spots, coveting the shelter denied them. In the vicinity of the house was a spacious flower-garden*, encompassed by a protecting plantation of bay, holly, augustines*, arbutus, laburnum, yellow and red Barbary, lilac, and Guelder-rose*, ever melodious with the shy, wary blackbird's whistle, the sweet notes of the secreted thrush, and the varied carols of their fellow-choristers*, a conspiring to give motion as well as life to their leafy concealment. To the right, was a rich, park-like* prospect, sprinkled with deer, grazing beneath clumps of commingled oaks and chestnuts or pulling acorns from the low, overhanging? branches of some solitary venerable stout-trinket* tree, whose outspread limbs bent downwards to the earth from whence their life was drawn, as if in thankfulness for the nourishment received. In an opposite direction stretched forth undulating woodland scenery, bordering on an open furzy down, which was frequently occupied by the moveable* abodes* of those houseless rovers-- the hardy, spoiliating*, mendacious tribe, whose forefathers Selim*, on conquering Egypt, was unable to extirpate, but contrived to expel, thereby entailing on Europe their lawless and unpopular posterity, so obnoxious to the proprietors of the localities they select for their temporary residences. I see a column of slow rising smoke Overtop the lofty wood that skirts the wild. c 2 A vagabond and useless tribe there eat Their miserable meal. A kettle slung between two poles upon a stick transverse Receives the morsel* of cock purloined From his accustomed perch. Hardening race! They pick their fuel out of every hedge, Which, kindled with dry leaves, just saves unquenched The spark of life. The sportive wind blows wide Their flutt'ring rags, and shows a tawny skin-- The vellum of the pedigree they claim. Great skill have they in palmistry, and more To conjure clean away the gold they touch, Conveying worthless dross into its place: Loud when they beg, dumb only when they steal." A grove of tall poplars* formed a conspicuous object from the western look-out*, and not far from hence rose, up the slope of a hill, a dense extensive coppice, impervious to the eye, where the lordly chief of the forest reared its head proudly over its arboreous companions, silently asserting its supremacy; and the graceful beech, silvery ash, dark-green* spiral fir, Scotch larch, and stunted hazel, were blended together, and the stream-wooling* willow dipped its pensile shoots into a clear, gurgling stream, that wound its tortuous course along, its sequestered, shady nooks pointing out to the angler the probable haunts of the hungry trout on the alert for its insect diet, and snug spots under the gnarled roots of undermined antique trees growing on the banks of the encroaching brook, hinting to juvenile poachers, setters of night-lines*, the likely lurking-places* of the snake-like*, slimy eel. Situated in a dell, at no great distance off, was the home-farm*, with its roomy barns, high granary, cow-sheds*, and fowl-house*, on entering which, perhaps the cackling her gave notice of her sedentary occupation, or the outstretched neck of the hissing goose apprised you of her displeasure at your approach. Without, probably the clustering poultry, emitting their various cries, surrounded you without alarm, expecting to receive a shower of grain in reward for their courage and confidence; a turkey or two, may be fearing to be late for the fare, running greedily upon the yard to join the rest, and the gaily-dressed* peacock condescending to associate with his inferiors* on the occasion. Roaming about, you doubtless encountered the sheep-dog*, if not away, attending on his fleecy charge in adjoining pastures, the nature of his bark denoting delight oranger*, according to his knowledge or ignorance of your countenance; and in passing the sties*, you naturally glanced at the swollen carcasses* of the noxious inmates, lying in their miry beds surfeited* with food, scarcely willing to open their small eyes or lift their snouts* from the stone troughs on which they rested; a short, low grint* perchance being the only acknowledgment of their consciousness of the presence of a visitor. The farm-house* was the very picture of rustic comfort-- a model of cleanliness and neatness Within; and the brickwork* of the exterior almost totally hidden by the undying ivy that clung tenaciously to every part, as if resolved not to separate from a pleasing acquaintance. A primly clipped box-hedge* bounded it on one side, and, running along in front, was a wattle paling supporting a mass of white jessamine. The dairy lay at the back, and its whitewashed walls were always well garnished with parallel tiers of Stilton, sage, cream, and other cheeses. Thus was ensured a certain supply of creature comforts, contributing in no small degree to create that full contentment that pervaded the household, where food was abundant, beer and cider plentiful, and work light. A few hundred yards from the farmstead was "The Retreat," where I, Arthur Montague, (for it is fitting I should begin to speak in propria* persona) was wont to pass many an hour in listless idleness, looking on the blooming landscape, listening to the humming bees, or teasing a pet jackdaw, who poked his head between the bars of his wicker cage when confine there for misconduct. The said Retreat* was an elegant little two-roomed* Gothic cottage, plastered with sparkling sanded cement

FIGURA 6.2 – Il passo meno ripetitivo dell'Ottocento.

La descrizione del paesaggio di Arthur Hawker ha un rapporto type-token pari a 60, molto al di sopra dei valori (di 46 e 50 per cento, per le notizie) riportati dalla *Longman Grammar* per segmenti di eguale lunghezza.

Nello studio del registro, l'opposizione fondamentale è quella tra l'oralità e la scrittura, ed è un fatto accertato che nella lingua inglese quest'ultima esprima un rapporto type-token molto più alto. Se l'archivio ha una varietà lessicale maggiore del canone, allora la ragione è che *l'archivio si sposta verso il registro "scritto" molto più del canone* (mentre quest'ultimo, come abbiamo visto nella sezione precedente, è molto più a suo agio con le convenzioni orali). Non è che i romanzi d'archivio con alti valori nel rapporto type-token contengano minori passaggi orali (dialoghi, discorsi, esclamazioni, ecc...). La ricerca di Marissa Gemma sul discorso colloquiale suggerisce invece che potrebbero averne anche di più, ma i passaggi "parlati" di questi romanzi hanno un carattere decisamente "scritto". Il romanzo di Jane West, *Ringrove*, contiene ad esempio una gran quantità di passaggi contrassegnati tipograficamente come "dialoghi" che però sono spesso delle tirate composte secondo una formalità che assomiglia più a una dissertazione scritta che a uno scambio orale³⁷.

Un atteggiamento linguistico conservatore è certamente la ragione principale della qualità "scritta" di molte opere d'archivio. Un passo tratto da *The Impostor* di William North – il cui rapporto rapporto type-token si trova in quell'1% del corpus – lo mostra bene:

There has of late years crept into our *belles lettres*, in addition to the *soi-disant* fashionable trash above mentioned, a violent predilection for low life, slang, and vulgarism of every kind. Dickens and Ainsworth led the way, and whole hosts became their followers ... Let us endeavor to reestablish pure classical taste.

[In anni recenti si è intrufolato nelle *belles lettres* nostrane, oltre all'usuale spazzatura di *cosiddetta* moda summenzionata, un gusto virulento per le esistenze di infimo rango, per il linguaggio da strada e per le volgarità di ogni sorta. Dickens e Ainsworth hanno aperto la strada e ora vengono seguiti da una folla di autori... Proviamoci di ristabilire la purezza del gusto classico].

³⁷ Eccone uno, che riguarda l'uso improprio di Byron dei suoi mezzi poetici: «There is a deep condensation of thought, an appropriateness of diction, an elegance of sentiment, and an original glow of poetical imagery; ever happy in illustrating objects, or deepening impressions; -which so fascinate our fancy and bewilder our judgment, that we lose sight of the nature of the deeds he narrates, and the real character of the actors» [«C'è un'ampia condensazione di pensiero, un'appropriatezza di dizione, un'eleganza di sentimento e un bagliore originale di immaginazione poetica; sempre felice di descrivere gli oggetti, o impressioni forti, che colpiscono a tal punto la nostra fantasia e intorbidiscono il nostro giudizio, al punto che perdiamo cognizione della natura dei fatti da lui narrati e della natura dei personaggi descritti»].

«Proviamoci di ristabilire...» Nel loro studio sul prestigio e lo stile, Underwood e Sellers hanno notato che molti oscuri libri «in fondo alla lista [del loro modello] [...] hanno l'ambizione del linguaggio ispirato e oratorio»³⁸. Lo stesso vale qui: «il linguaggio da strada» e «le volgarità», tratti tipici del registro orale, offendono «il gusto classico». Ma il gruppo della FIGURA 6.1 va al contrattacco “elevando” il tono del discorso alla *gravitas* formale della pagina scritta: molti nomi, molti aggettivi, e quante meno possibili forme verbali flesse (FIGURE 6.3-6.4, 6.5)³⁹.

Finora abbiamo spiegato l'affinità fra la grande varietà dei rapporti type-token e del registro scritto come il risultato, per dirla semplicemente, di scelte stilistiche e ideologiche. Tuttavia la loro correlazione ha anche una ragione molto più neutrale e “funzionale”. Secondo la linguistica dei corpora, il livello di massima diversità lessicale corrisponde coerentemente ai testi delle notizie giornalistiche: è un discorso che richiede un livello estremamente denso di elementi nominali, sottolinea il *Longman Grammar*, «refer to a diverse range of people, places, objects, events, etc.» (pp. 53-54) [«per indicare un'ampia gamma di persone, luoghi, oggetti, eventi, ecc...»]. Le fonti giornalistiche presentano una doppia fonte di varietà lessicale: la prima è la necessaria specificità interna a ogni notizia; la seconda è la netta discontinuità *fra* una notizia e la successiva. All'inizio di ogni articolo o corrispondenza il tasso di ripetizione ritorna allo zero e il type-token aumenta di conseguenza. Questa doppia logica ritorna nei testi narrativi con un alto valore nel rapporto type-token: questi accolgono una notevole quantità di

³⁸ Underwood e Sellers, “How Quickly Do Literary Standards Change?”, cit., p. 14.

³⁹ L'alta frequenza di nomi e di aggettivi ci porta di nuovo ai bigrammi grammaticali di cui si è parlato alla fine della sezione 4: le coppie “aggettivo-aggettivo”, “nome proprio-nome”, “nome-aggettivo”. Incrociando questi risultati con quello che è emerso dalla sezione, possiamo finalmente risolvere il paradosso dei testi ad alta ridondanza di bigrammi e di un alta varietà nel rapporto type-token. La funzione *labelling* di bigrammi come “Conte Goldstein” e “zio Gerard” o di cliché retorici come “iron will” e “clever little” può ripetersi facilmente nel corso di un romanzo: da qui la crescita della ridondanza *misurata a quella scala*. Ma anche un mediocre scrittore difficilmente ripeterà “clever little” in un segmento di mille parole, lasciando quindi la varietà lessicale a un livello abbastanza alto. Avverrà l'opposto con i bigrammi “determinativo-nome” o “preposizione-determinativo” che sono tipici dei testi canonici: l'articolo “the” è la parola più frequente della lingua inglese e si ripeterà inevitabilmente decine di volte in un segmento di 1000 parole, diminuendo la varietà percentuale del lessico. Ma dato che il nome accanto all'aggettivo varia con grande facilità, la ridondanza dei bigrammi rimarrà relativamente bassa.

materiali disparati e accentuano la loro diversità utilizzando una pluralità di generi diversi. Jane West, che ha ben sei romanzi nel 3% del corpus per rapporto type-token, cita la poesia in diciassette dei ventiquattro segmenti che hanno raggiunto la percentuale più alta; in mancanza di poesia, ripiega su metafore elaborate («expect a fearful tempest to arise, which will clear the tree of its unsound branches» [«aspettatevi una tremenda tempesta che spezzerà gli spogli rami dell'albero»]) e anche sul *pastiche*⁴⁰. L'introduzione di William North al suo *The Impostor* – metà critica letteraria, metà apologia – tocca una grande quantità di argomenti e presenta una divagazione su... sulla grande varietà di argomenti che ha deciso di trattare nel suo “romance”⁴¹. Thomas Hope la butta sulla profezia politica⁴², Lewis Wingfield si cimenta in una di-

⁴⁰ «First, Venus, queen of gentle devices! taught her prototype, lady Arabella, the use of feigned sighs, artificial tears, and studied fainting: while Aesculapius descended from Olympus, and, assuming the form of a smart physician, stepped out of an elegant chariot, and on viewing the patient, after three sagacious nods, whispered to the trembling aunt, that the young lady's disorder, being purely mental, was beyond the power of the healing art. Reduced to the dire alternative of resigning the fair sufferer to a husband or to the grave, the relenting lady Madelina did not long hesitate». Jane West, *A Tale of the Times*, 1799. [«In principio, Venere, signora di gentili artifizi! il suo idealtipo, Lady Arabella, istruiva all'uso di finti singhiozzi, lacrime artificiali, e dello svenimento calcolato: quando Esculapio scese dall'Olimpo e, prendendo il sembiante di un medico sagace, venne fuori da un'elegante carrozza, osservando il paziente, dopo aver scosso il capo tre volte con fare sornione, sussurrò alla zia che il caos, tutto mentale, di quella giovane donna, era ben oltre il potere di ogni arte curativa. Ridotta alla magra alternativa di scegliere se concedere questa povera degente ad un marito o ad una tomba, intenerita, Lady Arabella, non ebbe tanto da esitare»].

⁴¹ «By introducing literary criticism, satire of political and social evils, and popular illustrations of interesting facts in science, I have hoped to add to the interests of a romance, in which I trust no deficiency of adventure, plot, and carefully developed character will be found. But the day has gone by for mere fashionable novels. The age is utilitarian, and even novelists (the poets of present times) must conform to the mode». [«Inserendo critica letteraria, satira politica e dei malcostumi sociali, e divulgazione popolare di interessanti fatti scientifici, ho sperato di aggiungere interesse al *romance*, nel quale sono sicuro non si troverà mancanza di avventura, trama e personaggi sviluppati con cura. Ma è passato il tempo dei romanzi semplicemente accattivanti. L'epoca è utilitaristica e anche i romanzieri (i poeti dei nostri giorni) si devono uniformare alla moda»].

⁴² «The time is at hand when all the tottering monuments of ignorance, credulity, and superstition, no longer protected by the foolish awe which they formerly inspired, shall strew the earth with their wrecks! Every where the young shoots of reason and liberty, starting from between the rents and crevices of the worn-out* fabrics of feudalism, are becoming too vigorous any longer to be checked: they soon will burst asunder the baseless edifices* of self-interest* and prejudice, which have so long impeded their growth. Religious inquisition,

vagazione architettonica semiparodica⁴³, Edward Hawker si dà all'istruzione naturalistica...⁴⁴

Ma basta esempi. È giunto il tempo di alcune riflessioni finali.

judicial torture, monastic seclusion, tyranny, oppression, fanaticism, and all the other relics of barbarism, are to be driven from the globe». Thomas Hope, *Anastasius, or, Memoirs of a Greek*, 1819. [«È prossimo il tempo in cui i traballanti monumenti dell'ignoranza, della credulità e della superstizione, non più protetti dal ridicolo timore reverenziale che ispiravano in precedenza, copriranno la terra con le loro macerie! Dovunque, i colpi, giovani, della ragione e della libertà, iniziando dall'intercapedini e dalle cervici delle strutture, svuotate, del feudalesimo, diventano ormai troppo vigorosi per essere assorbiti: presto raderanno al suolo questi edifici senza fondamenta di interessi particolari e pregiudizio, che così a lungo hanno impedito loro di crescere. L'inquisizione religiosa, la tortura giuridica, la reclusione monastica, la tirannia, l'oppressione, il fanatismo, e tutti gli altri relitti della barbarie, saranno spazzati via dal globo»].

⁴³ «a stately entrance hall in the most fashionable quarter of the metropolis, embellished with lofty Ionic columns of sham Sienna marble; in front of each a magnificent bust of sham bronze by Mr. NoUekins* on a pedestal of scagliola. From a heavily stuccoed* ceiling, wrought in the classic manner, depend six enormous lanterns in the Pagoda style, wreathed with gaping serpents. Along three sides there are rows of "em pire*" benches, covered with amber damask, on which are lolling a regiment of drowsy myrmidons in rich liveries*. Passing these glorious athletes, you enter an ante-room choked with chairs, sofas, settees*, whose florid gilding is heightened by scarlet cushions. Very beautiful». Lewis Wingfield, *Abigel Rowe. A Chronicle of the Regency*, 1883. [«Una imponente sala di ingresso nella zona più alla moda della città, con flessuose colonne ioniche di finto marmo di Siena; di fronte a ciascuna di esse un magnifico busto di finto bronzo di Mr. NoUenkins* su un piedistallo di scagliola. Dal soffitto divinamente stuccato, arredato alla vecchia maniera, pendono sei enormi lanterne nello stile da Pagoda, ornate da serpenti penzolanti. Lungo le tre pareti c'erano file di panchine stile "impero*", coperte di ambr damascata, sulle quali sedeva un reggimento di robusti mirmidoni in ricche livree*. Oltrepassati questi gloriosi atleti, si entra in una anticamera, piena di sedie, sofà, e divani*, il cui florido bagliore era esaltato da vesti di porpora. Molto bello»].

⁴⁴ «The shield, slung to his neck, bore no emblazonry, and his open baronet and pennonless* lance argued him neither to have undergone the clapham, or knightly box on the ear (!); nor the osculum pads, which more gently signified the chivalric brotherhood. He was, however, well mounted and perfectly armed. Judging from his simple habergeon, and a silver crescent which he bore, more in the way of cognizance than as his own device, he might be pronounced a superior retainer in the service of some great feudatory». Edward Duros, *Otterbourne; A Story of the English Marches*, 1832. [«Lo scudo, che gli pendeva sul collo, non aveva alcun blasone, e la sua lancia da baronetto, senza pennone, faceva arguire che egli non avrebbe indossato né il calaface né l'elmo da cavaliere sulle orecchie (!): e neanche le piastre osculari che significavano, molto più urbanamente, la fratellanza fra cavalieri. In ogni caso era ben arcionato e armato. A giudicare dal suo semplice usbergo, e dalla falce di metallo che portava, più come forma di riconoscimento che come vero e proprio strumento, avrebbe potuto essere nominato vallassore in servizio a qualche grande feudatario»].

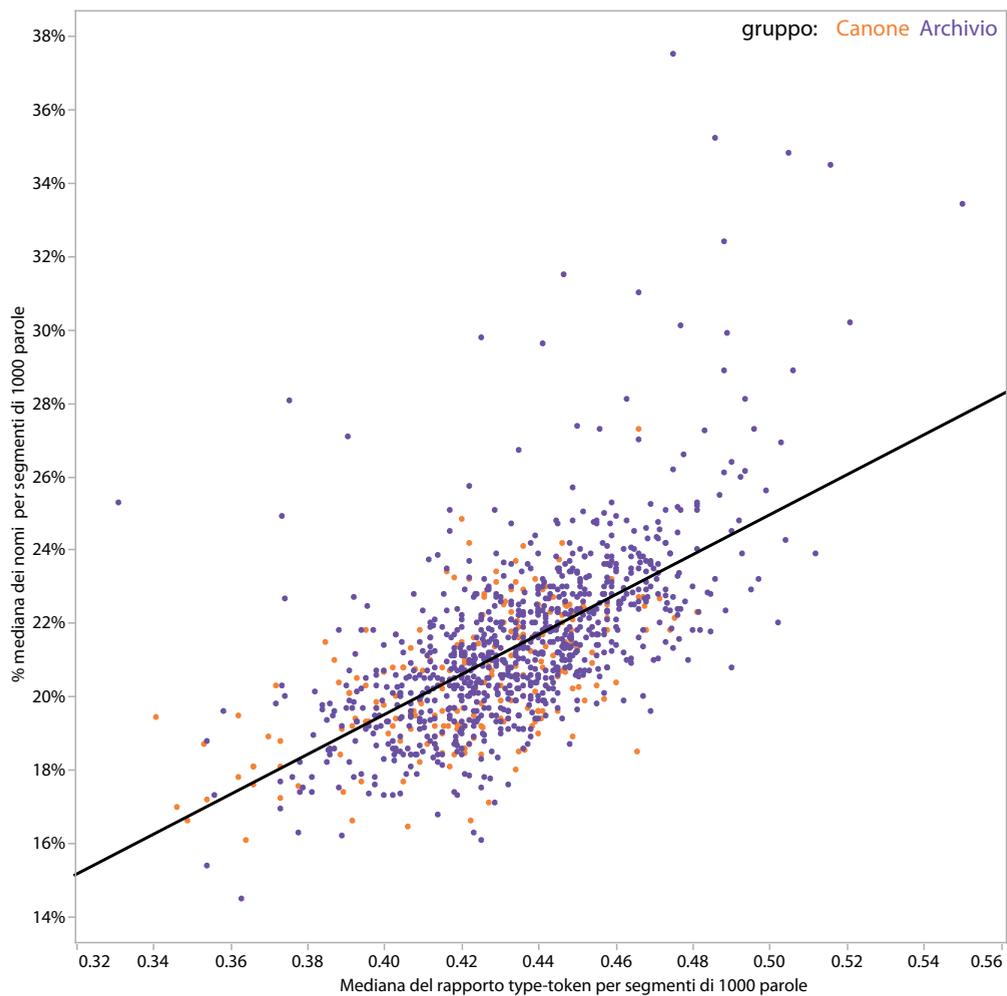


FIGURA 6.3 – Rapporto type-token e nomi.

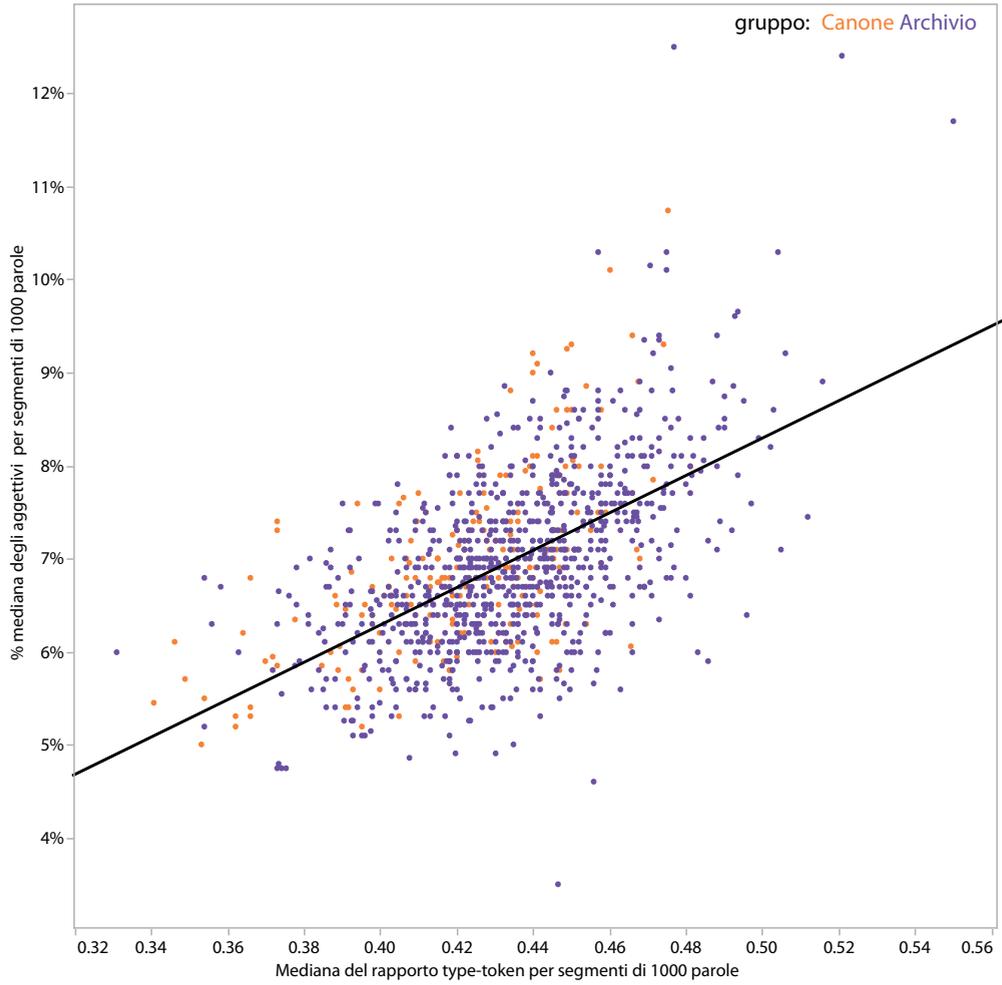


FIGURA 6.4 – Rapporto type-token e aggettivi.

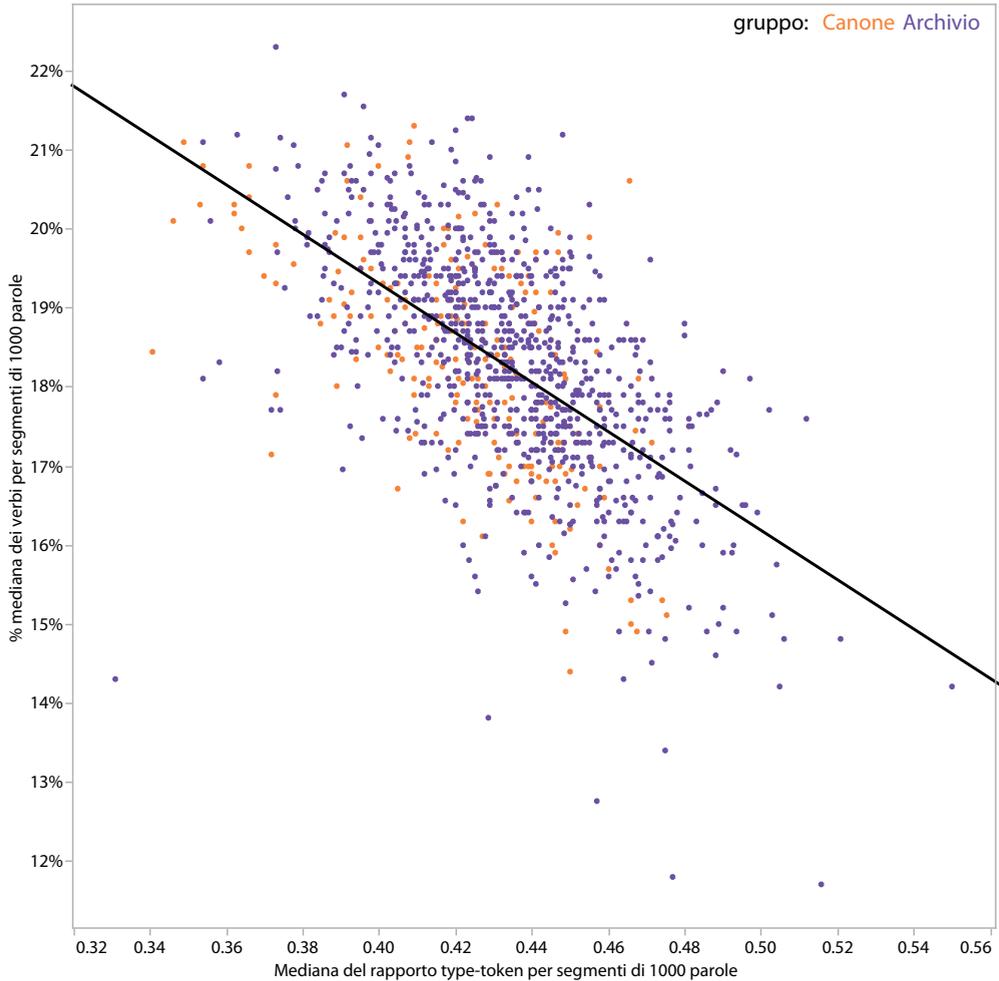


FIGURA 6.5 – Rapporto type-token e verbi.

Nel passaggio di Hawker su Gibilterra, della FIGURA 6.2, gli aggettivi (e i participi) sono più del triplo della media del romanzo ottocentesco. Per converso, il passaggio dell'*Adam Bade* nella FIGURA 5.4 contiene solo quattro nomi e un solo aggettivo, “hangry”, in 75 parole.

III. Dinamiche su larga scala nel campo letterario

Non è facile “concludere” un progetto che si è allontanato così tanto dal proprio obiettivo originale. I nostri oggetti di studio iniziali sono stati il canone e l’archivio e abbiamo utilizzato la ridondanza e l’indice di varietà lessicale (la relazione type-token) quali strumenti d’indagine; poi la relazione fra mezzi e scopi si è silenziosamente rovesciata: il canone e l’archivio si sono spostati sullo sfondo delle nostre discussioni, mentre la ridondanza e la varietà lessicale hanno finito per occupare il centro. Non c’era niente di pianificato in questa inversione, per un bel po’ non ci eravamo neanche resi conto di quel che stava accadendo. Abbiamo infatti passato mesi e mesi a chiederci cosa effettivamente “significassero” quei bigrammi e per quale ragione al mondo fossero così efficaci nel separare i nostri testi. Quando poi, dopo un po’, Allison e Gemma hanno introdotto la questione dei registri orali e scritti, abbiamo impegnato ancora più tempo sull’indice di varietà lessicale, leggendo passi di romanzi sconosciuti pieni di segni quali cancelletti, asterischi, e di parole quali “acclivio”, “laburno”, “mescidanza”. Strano.

Perché l’abbiamo fatto? Perché avevamo la sensazione che lavorare sulla varietà lessicale ci avrebbe permesso di comprendere qualcosa sulle forze “interne” (distinte da quelle esterne, discusse nella sezione 3) che informavano il campo letterario. Il nostro oggetto di studi si è spostato ancora una volta: la linea di presunta demarcazione tra il canone e l’archivio (la barra diagonale visibile nel titolo del nostro articolo) stava perdendo molto del suo interesse, una volta assorbita in un paesaggio più vasto. Con tutte le dovute distanze, c’era una somiglianza con la traiettoria di Bourdieu di quarant’anni prima, quando, a partire da uno studio sulla *Educazione sentimentale* e sulla posizione di Flaubert nella letteratura francese del XIX secolo, aveva sviluppato un quadro generale in cui Flaubert era sì ancora presente ma solo come un elemento fra gli altri. A noi è accaduto lo stesso: il canone e l’archivio sono ancora “dentro” il quadro, identificati da diversi colori, ma l’obiettivo dei nostri diagrammi consiste nell’illuminare il campo letterario nel suo complesso. La polarità stilistica esemplificata da Eliot e Hawker non evoca più per noi il canone e l’archivio ma i registri “orale” e “scritto”. Il fulcro della nostra attenzione è cambiato.

Rimane comunque una differenza fondamentale fra il lavoro di Bourdieu e il nostro. Per noi la sociologia del campo letterario *non può poggiarsi sulla sola sociologia*: c’è bisogno di una componente morfologica

forte. Ecco perché la ridondanza e (soprattutto) il rapporto type-token sono diventati così importanti: la fusione di quantitativo e qualitativo era ideale per stabilire la morfologia del romanzo a cui miravamo come obiettivo finale. Retrospectivamente, dobbiamo ammettere che, sebbene si sia avvicinato, l'obiettivo è rimasto fuori portata. Fuori portata, nel senso che, ove c'era una correlazione molto forte fra forma e destino sociale (è il caso della ridondanza) la natura elusiva dell'unità morfologica dei bigrammi ha reso difficile stabilire una relazione causale. Laddove invece una caratteristica si prestava a un'analisi articolata e accurata (la relazione type-token), la correlazione era più debole e si mostrava incontestabile solo nei casi più estremi. Al tempo stesso, due fenomeni che erano divenuti visibili accanto a questi casi estremi – l'intensità delle voci dei personaggi nei momenti di percentuale più bassa e, all'estremo opposto, la disparata raccolta di argomenti nella prosa del narratore – avevano aperto una nuova linea d'indagine, nella quale il *continuum* quantitativo-qualitativo riemergeva con chiarezza puntando con forza verso due concetti chiave della teoria del romanzo di Michail Bachtin: la polifonia e l'eteroglossia (gli "altri linguaggi" dei discorsi extra-letterari consolidati, come la politica, l'estetica, la geografia, l'architettura, ecc...). Di solito queste due nozioni vengono considerate strettamente connesse (e anche Bachtin sembrava pensarla così): tuttavia, come ha osservato Hannah Walser nelle nostre ultime discussioni, i nostri risultati indicavano che queste nozioni erano in realtà al lavoro ognuna in un estremo opposto del campo letterario: la polifonia risultava tendenzialmente associata a testi canonici, l'eteroglossia ai romanzi dimenticati. Lo stretto legame fra eteroglossia e insuccesso era, in effetti, particolarmente interessante. Per Bachtin, quando il romanzo entra in contatto con altri discorsi, li trasforma in modo creativo, si appropria della loro forza e rafforza la propria centralità nel sistema culturale. È come se, grazie all'eteroglossia, nulla possa andar storto. Invece è esattamente quanto è successo alla nostra piccola armata di autori dimenticati: l'incontro con altri discorsi ha esercitato su di loro un effetto paralizzante e, al posto della vitalità dialogica, ha prodotto pallidi cloni della prosa non romanzesca: una scelta disastrosa per sopravvivere nel sistema letterario inglese.

L'eteroglossia è quindi come una potenziale patologia della struttura romanzesca? «Non esistono fatti normali o patologici in sé», scrive Georges Canguilhem nel suo capolavoro sulle concezioni di normalità del XIX secolo: «L'anomalia o la mutazione non sono in se stesse patologiche. Esse esprimono

altre possibili norme di vita»⁴⁵. Se questa tesi è giusta, quel che ha messo fuori gioco Hawker, North e Duros non è stata tanto la scelta dell'eteroglossia in sé quanto il fatto che la scelta sia avvenuta in un periodo e in un paese – un ecosistema – *in cui il la forma del romanzo andava nella direzione opposta*. Vale a dire che invece di cercare ispirazione in discorsi esterni (come pure stava avvenendo in altri paesi) il romanzo lavorava alla sua meccanica narrativa interna. Anche Dickens, con la sua esperienza del gergo *parlamentare*, ha scritto racconti con un alto tasso di “oralità”. È questa particolare situazione storica che ha fatto degli “altri linguaggi” veicolati dall'eteroglossia uno svantaggio in termini di sopravvivenza.

A questa altezza, una visione storica di più ampio respiro può essere d'aiuto. Alcuni anni fa Niklas Holzberg, esperto di filologia classica, ha scritto un saggio la cui principale metafora cognitiva (la nozione di “fringe”, *frangia*) ha lasciato un'impronta profonda negli studi sul romanzo nell'antichità. Holzberg ha utilizzato questo termine per evocare il fatto che attorno a un gruppo ristretto di romanzi greci e latini “veri e propri”, c'era un gruppo molto più ampio di testi in cui i tratti romanzeschi erano mescolati con elementi tratti da altri discorsi (storiografia, narrativa di viaggio, filosofia, educazione politica, pornografia ...), estendendo nel complesso la portata del romanzo. Nei venti secoli che seguirono, mentre il romanzo “vero e proprio” intensificava la sua produzione, diversificava i generi e migliorava il proprio status all'interno del sistema culturale, parallelamente il ruolo della *frangia* si contraeva al punto tale che gli studiosi di letteratura moderna si sono a malapena interessati a questo concetto⁴⁶. Ma, a conti fatti, la *frangia* non ha mai smesso di esistere e gli scrittori della **FIGURA 6.1** ne sono una versione moderna, con la loro strana proliferazione di temi che è il segno distintivo delle opere situate al confine fra romanzo e altri discorsi. Il vero problema risiede nel fatto che nel corso degli anni, la funzione morfologica di questa zona di confine (terreno fertile per l'incontro tra romanzo e altri discorsi) è diventata più instabile. Un secolo prima, un romanzo che si tingeva delle sfumature dell'autobiografia spirituale, dei meccanismi dell'epistolario o delle discontinuità del “sentimento” poteva ancora guadagnare il titolo di capolavoro e

⁴⁵ Georges Canguilhem, *Il normale e il patologico*, trad. it. di D. Buzzolan, Torino, Einaudi, 1987, p. 114.

⁴⁶ Niklas Holzberg, “The Genre: Novels proper and the Fringe”, 1996, in Gareth Schmelting (a cura di), *The Novel in the Ancient World*, edizione rivisitata, Boston-Leiden, Brill 2003.

dare vita a un sottogenere di successo: *The Pilgrim's Journey*, *Pamela*, *Tristram Shandy* e forse anche *Waverley*, possedevano molte delle caratteristiche della *frangia*. Invece nel corso del XIX secolo (forse a causa della divisione del lavoro intellettuale che ha portato a un crescente divario tra la narrativa e le scienze sociali, rendendo i rispettivi linguaggi sempre più intraducibili l'uno con l'altro) il ruolo dell'eteroglossia nello sviluppo della forma romanzesca è diventato problematico: è questo fenomeno che ha deciso il destino di questi autori dimenticati⁴⁷.

Se questo risponda alla nostra domanda iniziale sul modo in cui l'archivio ha cambiato la nostra conoscenza della letteratura, non spetta a noi dirlo. Quel che possiamo dire è che, con il procedere del lavoro, ci siamo trovati a dedicare sempre più tempo a *Ringrove*, *The Impostor*, e *Arthur Mortague*, e che in certi momenti fortunati, ci siamo accorti che questi romanzi avevano posto delle questioni che un romanzo come, ad esempio, *Adam Bede* non avrebbe mai sollevato. Ma i momenti fortunati sono stati *pochi* perché non è facile mantenere l'attenzione sull'archivio: in parte si deve resistere al richiamo esercitato dagli autori più noti – l'attrazione verso il conosciuto – che riportano a sentieri già battuti; in parte è la natura problematica di quel che gli scrittori dimenticati mettono sotto gli occhi: un enorme spreco di ideali grandiosi, molto diverso dal paesaggio con cui si confrontano gli storici della letteratura. Imparare a esaminare queste rovine senza un atteggiamento di superiorità (e senza pietismi) è quel che il nuovo archivio digitale ci sta chiedendo di fare. Sul lungo termine potrebbe essere un cambiamento anche più grande dell'introduzione della stessa analisi quantitativa.

(Traduzione di Giuseppe Episcopo e Luca Marangolo)

⁴⁷ Sulla stessa logica, i capolavori dell'eteroglossia – come *Moby Dick* o l'*Ulisse* – si sono dovuti allontanare progressivamente dall'asse principale dell'evoluzione romanzesca, acquisendo sempre meno interesse per lettori non accademici.