

Vedere e non vedere. Sulla visualizzazione dei dati nelle discipline umanistiche

Franco Moretti e Oleg Sobchuk*

Istogrammi, *scatterplots*, serie temporali, diagrammi, reti ... Dieci-quindici anni fa, negli articoli di cinema e musica, di letteratura e arti visive, non si trovava niente di tutto questo. Adesso però sì, ed è anzi proprio quello che distingue immediatamente le *digital humanities* dalle “altre” discipline umanistiche. Qui, riflettiamo su alcune premesse (tacite, e spesso probabilmente inconscie) di questa nuova prassi, concentrandoci su sessanta-settanta studi che hanno avuto una particolare influenza nello sviluppo della nuova disciplina, ed esaminando il modo in cui presentano i loro dati.¹ Quello che ci interessa è portare alla luce la visualizzazione come *pratica*, perché siamo convinti che la pratica – ciò che impariamo a fare “facendolo”, sviluppando pian piano un’abitudine professionale che troviamo perfettamente “naturale” – ha delle fortissime implicazioni teoriche; che a volte contraddicono persino le dichiarazioni teoriche esplicite e guidano di fatto il processo di ricerca.²

I. Storia

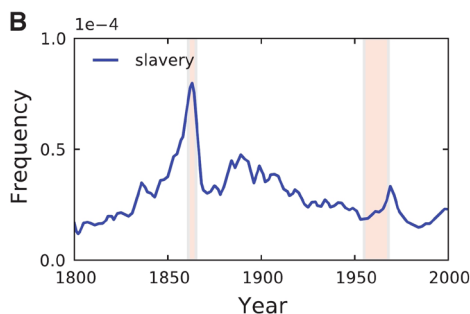
La forma del tempo

Cominciamo con l’articolo che annunciò la creazione degli *n-grams* di Google, caputlando l’approccio digitale-quantitativo ben al di là dei confini di una piccola nicchia accademica: “Quantitative Analysis of Culture Using Millions of Digitized Books”, pubblicato su *Science* nel gennaio del 2011.³ La figura 1 è la prima immagine che si incontra nell’articolo, e dà il tono a tutto quel che segue: l’asse orizzontale misura il passare del tempo, e quello verticale la frequenza della parola *slavery*. Una “serie temporale”, come di solito chiamano questo tipo di diagrammi: gli anni passano, e la presenza di *slavery* cambia: raddoppia durante la Guerra civile, scende lentamente fino al valore iniziale, risale, ma non del tutto, al tempo delle lotte per i diritti civili, e così via. “Quantitative Analysis of Culture” contiene 33 diagrammi, e 27 – l’ottanta per cento – sono di questo tipo.

Negli ultimi dieci anni, le serie temporali sono diventate una sorta di “firma” delle *digital humanities*: non mancano mai, per esempio, in quegli articoli che – pubblicati da riviste “centrali” della critica letteraria, come *Modern Language Quarterly*, *Poetics Today*, o *New Literary History* – hanno funzionato da “ponte” tra il vecchio e il nuovo. La semplicità, come suggeriscono Rosenberg e Grafton, ha certamente giocato a loro favore. Due soli elementi: storia, e semantica. Una parola (figura 1), due (figura 2), quattro (figura 3), o anche centinaia, come nei “campi semantici” e nelle *topics* delle figure 4 e 5: le quantità cambiano, e cambia anche l’oggetto dell’analisi (libri, articoli

di giornale, rapporti della Banca Mondiale, romanzi, saggi critici): quello che *non* cambia, è la centralità del contenuto. *Topic modelling; content analysis; text mining*: come dimostra la terminologia corrente, per le *digital humanities* il significato è una sorta di materia prima, grezza, indifferente all'organizzazione del testo. I corpora sono "sacchi di parole" (*bags of words*) da cui il significato deve essere "estratto" (*text mining*): fatto questo, è tutto perfettamente chiaro: "i mutamenti del discorso *rivelano* più vasti mutamenti storici e socio-culturali"; "i modelli [...] *rivelano* un marcato declino delle emozioni positive"; "questo metodo *rivela* tendenze importanti ma rimaste finora invisibili". Il linguaggio *rivela*; non mente mai; non nasconde nulla. È un'idea edenica della cultura: il regno dell'esplicito.

Fig. 1. "Quantitative Analysis of Culture Using Millions of Digitized Books" (2011)



Soltanto alla metà del XVIII secolo si diffuse una terminologia visiva condivisa per le mappe del tempo. Ma i nuovi formati lineari del Settecento furono accettati così rapidamente che, nel giro di pochi decenni, diventò difficile ricordare un tempo in cui non fossero in uso. Apparve chiaro che il problema fondamentale delle cronografie non era quello di tracciare schemi visivi più complessi (che era stato l'approccio di molti aspiranti innovatori del secolo precedente), ma piuttosto quello di semplificare, di creare uno schema visivo che trasmettesse chiaramente l'uniformità, la direzionalità e l'irreversibilità del tempo storico. Daniel Rosenberg e Anthony Grafton, *Cartografie del tempo*

Fig. 2. "Content Analysis of 150 Years of British Periodicals" (2016)

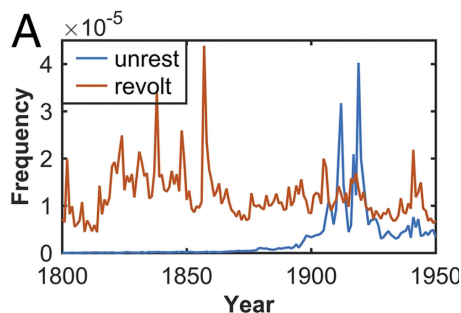


Fig. 3. "Come parla una banca. La lingua nei rapporti della Banca Mondiale 1946-2012" (2014)⁴

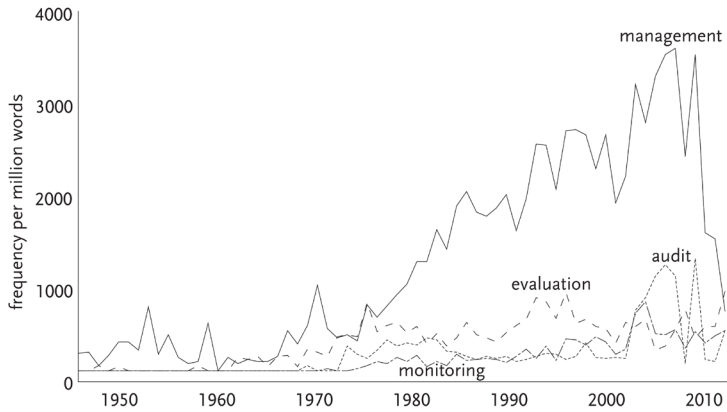


Fig. 4. "A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method" (2012)

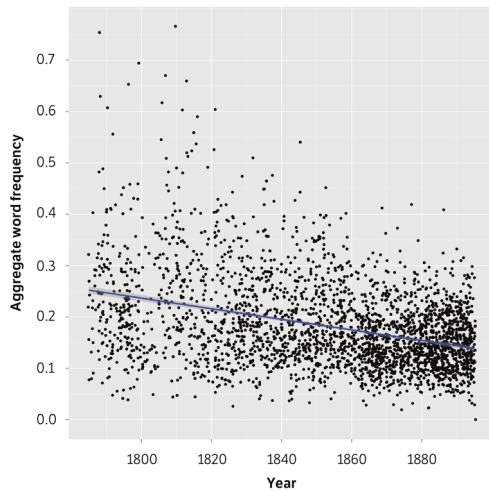
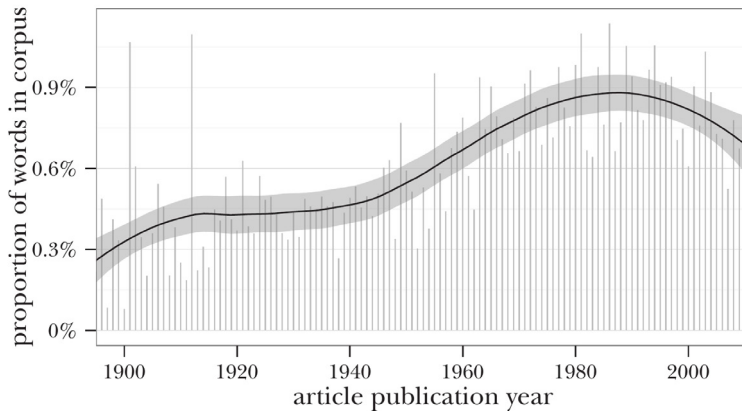


Fig. 5. "The Quiet Transformation of Literary Studies: What Thirteen Thousand Scholars Could Tell Us" (2014)



Ci torneremo. Adesso, spostando l'attenzione dall'asse verticale a quello orizzontale, è impressionante quanto spesso le serie temporali durino esattamente cento anni. I romanzi e i saggi critici delle figure 4 e 5, il lessico della "proprietà" nei dibattiti parlamentari inglesi (figura 6), i bestseller scritti da donne (figura 7), la durata delle inquadrature nel film (figura 8), l'espressione delle emozioni in narrativa (figura 9), le contrazioni nel romanzo americano (figura 10), le ripetizioni nel canone e nell'archivio (figura 11), le recensioni nelle raccolte di poesia (figura 12) Quale che sia l'argomento, il *secolo* è emerso come l'unità di misura fondamentale della storia quantitativa della cultura.

Fig. 6. "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora" (2018)

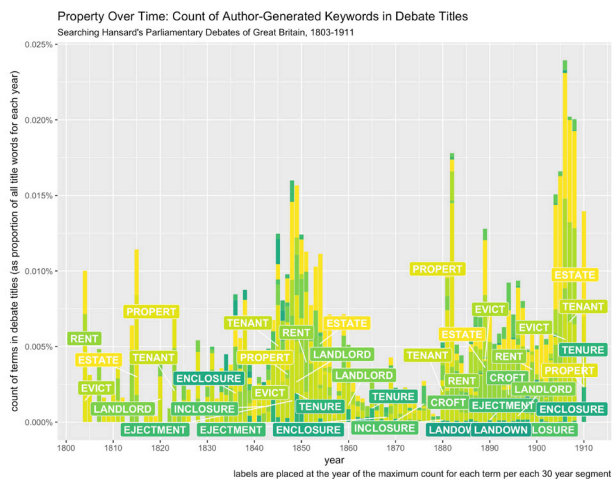


Fig. 7. “The Transformation of Gender in English-Language Fiction” (2018)

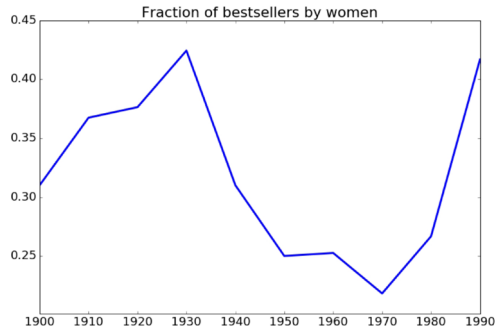


Fig. 8. “Shot Durations, Shot Classes, and the Increased Pace of Popular Movies” (2015)

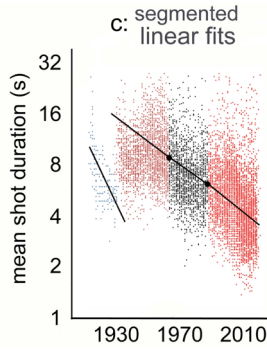


Fig. 9. “Birth of the Cool: A Two-Centuries Decline in Emotional Expression in Anglophone Fiction” (2016)

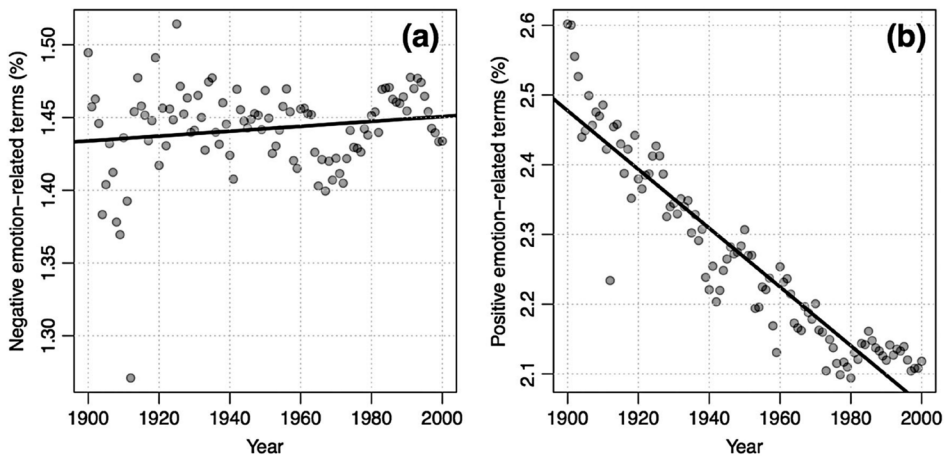


Fig. 10. "The Making of Middle American Style" (2016)

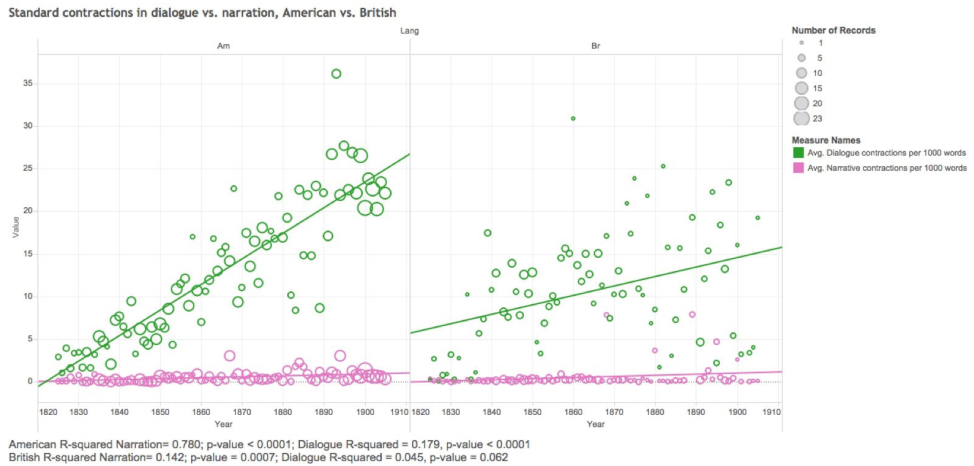
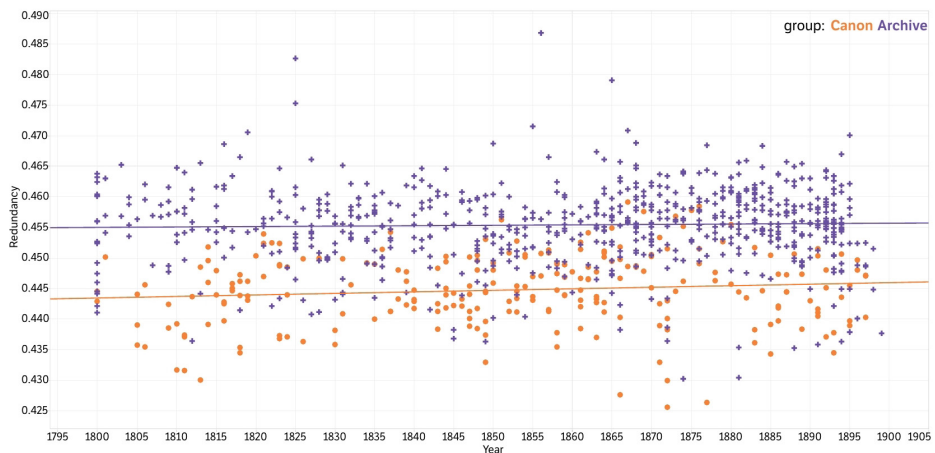


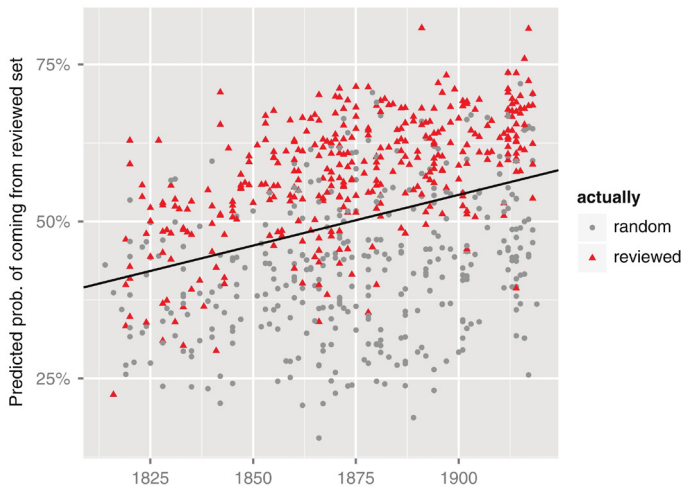
Fig. 11. "Canone/ Archivio. Dinamiche su larga scala nel campo letterario" (2016)



Ci sono dei casi in cui la cosa è quasi inevitabile: il cinema esiste da poco più di cent'anni, ad esempio, e i libri pubblicati tra il 1800 e il 1900 hanno il duplice vantaggio di essere facilmente scannerizzabili (a differenza di quelli anteriori), e di non essere soggetti al diritto d'autore (a differenza di quelli posteriori); il che spiega perché vi siano così tanti studi sull'Ottocento. Ma il predominio del secolo nella nuova immaginazione storica è probabilmente dovuto a qualcos'altro, e cioè all'idea, avanzata in "The Quiet Transformation of Literary Studies", che le *digital humanities* siano in grado di "scoprire e interpretare *patterns* su una scala storica nuova e diversa".⁵ Una scala nuova e diversa: è la dichiarazione d'indipendenza

dalle “vecchie” discipline umanistiche. Ma *in che modo* diversa? Ecco, il secolo è un modo di rispondere a questa domanda.

Fig. 12. “The Longue Durée of Literary Prestige” (2016)



Intuitivamente, un secolo è lungo; non è antropomorfo (perché di rado viviamo così a lungo), ed è dunque estraneo alla vecchia centralità dei singoli autori; è il tempo *del mondo*, non della vita. Le lingue romanze se ne servono per periodizzazioni un po' alla buona (*el Siglo de Oro*, il Quattrocento, *une dix-neuvièmiste*), le università americane per molte delle loro assunzioni. L'idea era lì, nella *doxa* che tutti respiriamo; un bel numero tondo, abbastanza ampio da suggerire una scala storica nuova – ma non così ampio da diventare incontrollabile. Certo, non era davvero un *concetto*; non aveva alcun ruolo, per esempio, nella tripartizione del tempo storico – *longue durée*, ciclo, evento – elaborata al tempo della prima svolta quantitativa dalla scuola delle *Annales*; e non è stato certo adottato in seguito a una riflessione teorica. Era la pratica, che prendeva silenziosamente la guida della teoria e le imponeva la sua forma.

Trend

Una scala storica diversa: in concreto, una in cui diventino visibili dei *trends* – delle “tendenze”. Fino a pochi anni fa, nessuno parlava di *trend*, nelle discipline umanistiche; finché si studiavano pochi testi racchiusi in pochi anni, la cosa non aveva senso. Sulla scala del secolo, un senso lo ha. Una nuova disciplina ha bisogno di parole chiave, e *trends* – con la sua miscela di quantificazione e orientamento – era perfetta per le *digital humanities*; e infatti eccola lì, che compare già nell'*abstract* dell'articolo di *Science* 2011 della figura 1 (“L'analisi di questo corpus ci permette di indagare trend culturali”) – e torna poi, come una sorta di ritornello, negli studi

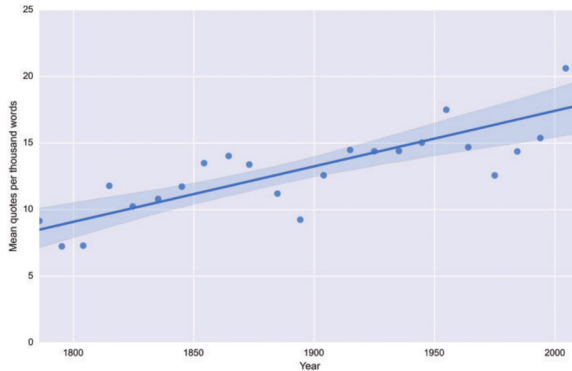
più diversi: l'evoluzione del jazz, le inquadrature del film, la critica letteraria, i giornali, le recensioni di poesia, gli archivi legali, le espressioni delle emozioni. C'è un articolo ("Quantitative Literary History of 2,958 Novels", 2012), in cui compare sessantotto volte. Sessantotto! E non è solo una parola: nelle figure 4, 8, 9, 10, 11, 12, e 13 i *trends* sono presenti sotto forma di regressioni e altre elaborazioni statistiche. Non abbiamo solo "parlato" di *trends*; li abbiamo resi *visibili*, e posti al centro delle nostre visualizzazioni.

Oxford English Dictionary: "Trend": "La direzione di massima che viene presa da una corrente, una costa, una catena di montagne, una valle, uno strato e così via." Questo, verso fine Settecento (prima, il termine era in circolazione – a partire dall'undicesimo secolo, stando allo OED – come verbo: un fiume "trends", una costa "trends"). Poi, a fine Ottocento, emerge il senso figurato: una discussione "trends": "prende una certa direzione"; "acquista una tendenza di massima"; "la tendenza, deriva, o corso generale (di un'azione, del pensiero ecc.)". Corso, tendenza, deriva, direzione ...; sempre al singolare, perché pensare in termini di *trends* significa *ridurre la pluralità a un corso singolo*: ridurre una nuvola di dati a una linea. A differenza dei tentativi precedenti, che avevano cercato di visualizzare "le molte traiettorie che si intersecano nella storia", scrivono Rosenberg e Grafton in *Cartografie del tempo*, "la formula della linea del tempo [...] metteva in rilievo la trama principale e le traiettorie ad ampio raggio".⁶ Da una molteplicità di traiettorie a una storia singola: è un passaggio decisivo. I dati sono sempre sparpagliati, confusi, pieni di "rumore"; le linee di tendenza sono perfettamente univoche. È per questo che gli abbiamo dato tanto rilievo: rendono i dati facili da leggere. Gli danno *un senso*.

Tentazione irresistibile, dare un senso alla storia. Figura 13: la frequenza del dialogo nei romanzi di lingua inglese. Il diagramma "mostra che [i romanzieri] aggiungono circa una frase di dialogo ogni mille parole ogni 25 anni, ossia circa una frase per pagina per secolo".⁷ Veramente, il diagramma mostra che il dialogo raddoppia nei primi 50 anni, oscilla molto, ma senza una direzione precisa, per altri 50, resta più o meno stabile per altri 70-80, e poi aumenta di colpo del 50% negli ultimi vent'anni. (O forse siamo davanti a due cicli di 80 anni l'uno, dove un aumento notevole finisce in un brusco declino, con il terzo ciclo che inizia intorno al 1980). In altri termini: i romanzieri fanno ogni sorta di cose, *tranne aggiungere una frase ogni 25 anni*; è la media statistica che fa questo. O si torni alla figura 8, sulla lunghezza delle inquadrature filmiche tra il 1913 e il 2013. "Il declino non è brusco", scrivono gli autori del saggio, "bensì uniforme e graduale lungo l'arco di almeno 80 anni".⁸ Non è brusco, quando nel film muto la durata delle inquadrature viene più che dimezzata in quindici anni? Uniforme per 80 anni, quando c'è un chiarissimo *aumento* tra il 1930 e il 1960? Ma li guardano, i loro dati? Certo che li guardano; il problema è che le *trendlines* hanno cambiato *il modo* in cui guardiamo i dati. La visualizzazione si rivolge alla nostra intuizione: se ci presenta un pulviscolo con in mezzo una linea, vedremo solo la linea. È inevitabile. E così, invece di aiutarci ad *analizzare* i dati, le medie ci hanno spesso aiutato a dimenticarli. Che pazzia: ci siamo rivolti alla quantificazione perché volevamo conoscere – e, per

cominciare, *vedere* – tutto ciò che il predominio del canone aveva reso invisibile; e ora che lo abbiamo davanti agli occhi, abbiamo trovato il modo di non vederlo!⁹

Fig. 13. “Dialogism in the Novel: A Computational Model of the Dialogic Nature of Narration and Quotations” (2017)



Serie temporali; contenuto semantico grezzo; secoli; *trends*. Un passo alla volta, la pratica della visualizzazione ha dato forma a una nuova storia della cultura. E ne ha anche definito i bersagli polemici. L’inizio di “The Quiet Transformation of Literary Studies”:

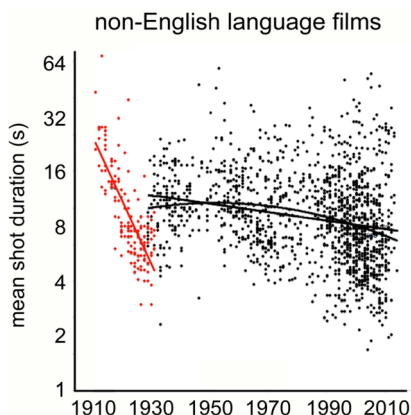
In passato, gli studi letterari sono stati fondamentalmente una narrazione di idee in conflitto. I movimenti critici si scontrano tra loro [...] Sebbene gli studi recenti abbiano complicato il quadro [...] sottolineando il conflitto sociale e istituzionale, il conflitto generazionale resta un parametro fondamentale: invece di lotte tra le idee, ora ci sono lotte tra dilettanti e professionisti e così via. Nel sottolineare il conflitto, questo tipo di storiografia trascura delle dimensioni di grande rilievo per la storia della nostra disciplina: le premesse che cambiano pian piano, senza discussione esplicita; i modelli soggiacenti che sopravvivono ai conflitti visibili alla superficie delle cose; le trasformazioni di lungo periodo causate dal mutamento sociale.¹⁰

Ancora qualche frase, e gli autori menzionano il “trend secolare” del loro primo diagramma (riprodotto qui alla figura 5): il trend, come meccanismo decisivo del divenire storico, al posto della vecchia “narrazione di idee in conflitto”. (Dopo essere stati menzionati 8 volte nelle prime 14 righe, “conflitto”, “lotta” e “scontro” tornano solo 6 volte nelle successive 25 pagine; il “trend secolare”, per parte sua, compare in 21 dei 23 diagrammi). E in effetti, di fronte a una trendline riesce difficile pensare a un conflitto: il trend è sempre uno. E così, in modo probabilmente del tutto involontario, un eccessivo ricorso ai trends ha di fatto bandito il conflitto dalla ricerca quantitativa, installando al suo posto – per usare il famoso slogan del neoliberalesimo – una plumbea atmosfera da “Non c’è alternativa” (“There Is No Alternative”). Diciamo la verità: la storia della cultura merita di meglio.

Scala della ragione

C'è anche un altro problema, con i *trends*. Gli autori di "Shot Durations, Shot Classes" sono convinti che le inquadrature cinematografiche siano diventate più brevi a causa della "crescente attenzione del pubblico"; poi però aggiungono anche di avere osservato che "la durata media delle inquadrature [era] di 4.64 secondi per gli *action films* [...] e di 8.55 secondi per *comedies* e *dramas*".¹¹ Le cifre sono plausibili: film dominati da bruschi movimenti del corpo tendono ad abbreviare le inquadrature molto più di quelli che si basano sul dialogo; così uno torna alla figura 8 e si chiede: possibile che dietro il *trend* costituito dall'abbreviarsi delle inquadrature vi sia l'aumento di "action films" sul mercato americano? Visto che nulla ci viene detto sulla composizione del corpus, la cosa è perfettamente plausibile, e resa ancor più tale dal diagramma di film non americani della figura 14: qui, dove gli *action films* sono certamente assai meno numerosi, la durata delle inquadrature quasi non cambia tra il 1930 e il 2010. Se la durata dipendesse da un aumento dell'attenzione, come vuole la tesi di fondo di "Shot Durations, Shot Classes", i risultati dovrebbero essere gli stessi ovunque (a meno di voler sostenere la tesi – grottesca – che il pubblico americano sia il più attento del mondo); visto che i risultati sono invece molto diversi, la presenza (crescente) degli *action films* può ben essere la variabile decisiva.¹² Invece di un *trend* a una dimensione, e della diagnosi, euforica, di un pubblico cinematografico sempre più attento, quello che abbiamo sotto gli occhi può ben essere una polarizzazione dell'industria cinematografica in direzioni opposte: in ultima analisi, un conflitto tra forze avverse, nel quale sta avendo la meglio una brutale semplificazione della logica narrativa. È un modo assai diverso di immaginare la storia della cultura.

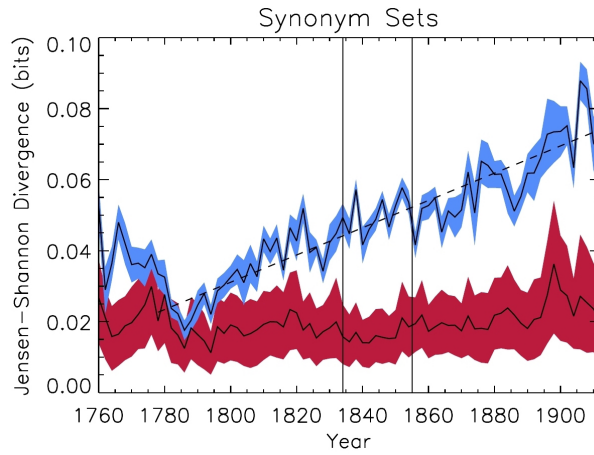
Fig. 14. "Shot Duration, Shot Classes, and the Increased Pace of Popular Movies" (2015)



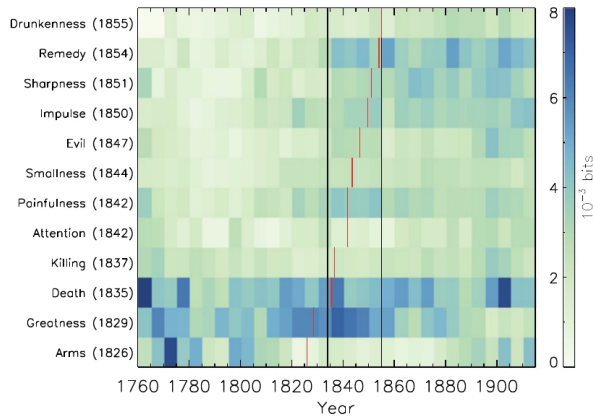
A volte, beninteso, i *trends* esistono – e sono anche uniformi e graduali. Questo però non implica necessariamente che *le forze soggiacenti siano anch'esse uniformi*. Si prenda “The Civilizing Process in London’s Old Bailey”: uno studio di come il linguaggio usato nei documenti legali inglesi per definire i reati violenti e non violenti sia cambiato nel corso di 160 anni (figura 15). Nonostante oscillazioni episodiche, il *trend* è chiaro, e sostanzialmente regolare. Ma questa “tendenza secolare” scrivono gli autori “non è semplicemente l’ampliarsi di una differenza iniziale; al contrario, i campi semantici che distinguono i reati violenti da quelli non violenti cambiano a loro volta nel corso del tempo”.¹³ Lo si vede dalla figura 16, che segue decennio per decennio le trasformazioni del linguaggio, e dove di regolare non c’è proprio nulla. Alcuni gruppi semantici, come i sinonimi di “morte”, conservano più o meno la stessa frequenza per tutto il periodo; “arms” raggiunge il valore più alto nella prima metà dell’Ottocento, e “remedy” (che annuncia l’entrata in scena della medicina) nella seconda; tra i casi non violenti, il “nascondiglio” e “vestiario” dei piccoli furti sono frequenti a inizio secolo e poi scompaiono; tutto il contrario per il “denaro” e i “documenti” caratteristici dei reati da colletti bianchi. Il *trend* era graduale e uniforme; le sue basi eterogenee e sconnesse fra loro. “The Making of Middle American Style”: un aumento costante dello stile colloquiale lungo più di un secolo (vedi sopra, figura 10) dovuto a cinque ondate distinte, distribuite su altrettante generazioni: dagli arcaismi irlandesi, scozzesi e inglesi (‘o’, ‘thee’, ‘thy’), alle formule di cortesia (“sir”, “doctor”), termini del linguaggio afro-americano e popolare (‘fer’, ‘ter’, ‘tuh’, ‘dat’), fino ai nomi propri – la *first-name basis* – del dialogo novecentesco (si veda lo studio complementare “Operationalizing the colloquial style: Repetition in 19th-century American fiction”, 2017). L’aumento era perfettamente regolare; le sue fasi del tutto contingenti. Si cambia la scala d’indagine, e si trova una storia diversa.

Un’analogia aiuterà a capire. In molte città italiane esiste un cosiddetto “Palazzo della ragione”, che all’inizio dell’età moderna ospitava tribunali, notai, magistrati, e altre funzioni del governo locale. A Verona, a metà Quattrocento, al palazzo venne aggiunta, nel Cortile del mercato vecchio, una “Scala della ragione”: una splendida costruzione nel marmo rosato che è tipico della città, e che, pur virando bruscamente a 90 gradi a metà percorso, continua a salire in modo perfettamente regolare: una sorta di idealizzazione della ragione stessa. Questa splendida regolarità poggia però su quattro archi che non potrebbero essere più diversi (figura 17) – proprio come il *trend* dell’Old Bailey emergeva da un mosaico di voci discordi. E uno si chiede: in casi come questi, quale dev’essere l’oggetto della spiegazione storica? Il *trend* regolare di lunga durata, o gli archi irregolari su cui tutto poggia? E in questo caso, dovremmo forse ritenere il *trend* un mero effetto di superficie – una sorta di fantasma statistico? O dovremmo forse puntare a un’architettura concettuale che sia, come la Scala, rettilinea e dissonante a un tempo? Ma che cosa significherebbe, questo, in termini di categorizzazione storica?

Fig. 15-16. "The Civilizing Process in London's Old Bailey" (2014)



(a) Partial KL, Violent to Mixture



(b) Partial KL, Non-Violent to Mixture

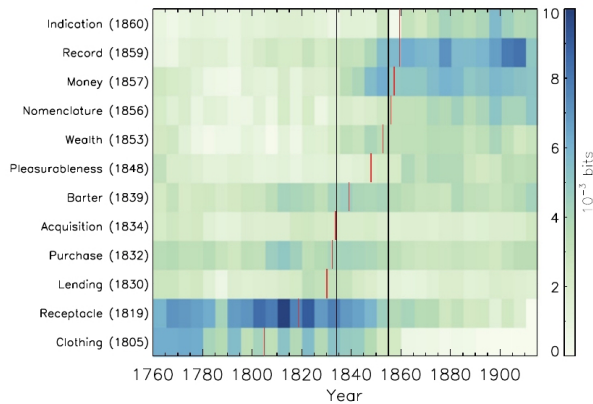


Fig. 17. Verona, Scala della ragione, ca. 1450



II. Morfologia

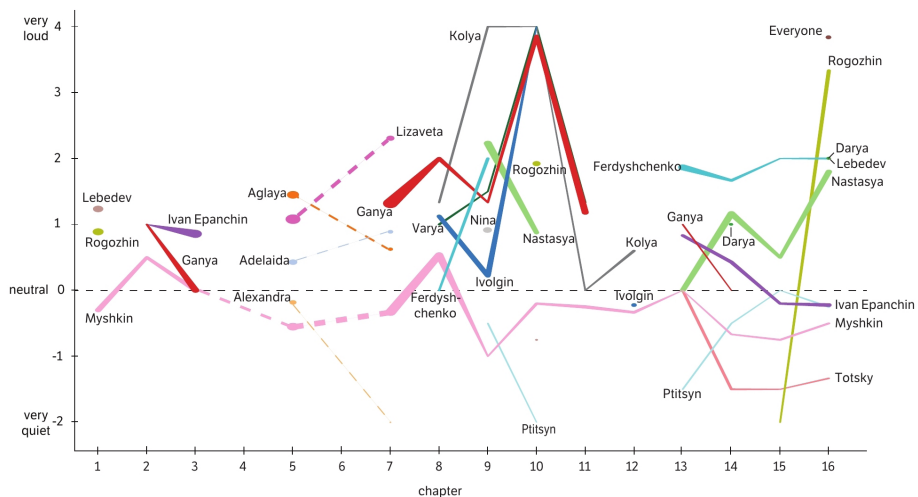
“Interazioni di elementi strutturali”

Più sopra, abbiamo visto come le serie temporali delle *digital humanities* abbiano una spiccata preferenza per un arco di tempo secolare. Nella visualizzazione di tratti morfologici, però – il flusso di coscienza, il ritmo delle commedie americane, il ruolo del protagonista nel dramma, la direzione dello sguardo nel ritratto europeo – l’arco diventa molto più vario: 30 anni, 60, 350, 500.¹⁴ E negli studi che non fanno uso di serie temporali, la varietà aumenta ancora, dai 3 anni delle canzoni di successo, ai 10 dei “romanzi teologici” inglesi, ai duemila (e più) delle *Pathosformeln* di Aby Warburg, fino alla durata del tutto indeterminata – “all’incrocio tra le narrazioni orali del folklore e i contesti letterari” – delle favole europee di magia.¹⁵

A che si deve tale disinvoltura storiografica? Al fatto che l’oggetto di studio è cambiato. Nel descrivere la differenza tra biologi “evoluzionisti” e “funzionali”, Ernst Mayr ha osservato come i secondi – che lavorano sulle “operazioni e interazioni di elementi strutturali” – non cerchino di seguire quegli “elementi” lungo il corso del tempo, come avverrebbe in uno studio evoluzionistico, ma si sforzino quasi di immobilizzarli al fine di “eliminare, o almeno controllare, tutte le variabili”.¹⁶ “Il termine ‘morfologia’ significa studio delle forme” – scrisse Vladimir Propp in apertura della *Morfologia della fiaba* – “studio delle parti componenti [...] e delle loro relazioni reciproche e con il tutto.”¹⁷ Per questo tipo di lavoro, più che la storia serve un tavolo anatomico.

Rispetto all’inizio di questo articolo, abbiamo virato di 180 gradi. Le serie temporali ignoravano completamente le “parti componenti” e “relazioni con il tutto”, e trattavano i testi come tanti “sacchi di parole” che “rivelavano” in modo del tutto trasparente il mondo circostante. La visualizzazione morfologica capovolge quest’impostazione: prende le “operazioni e interazioni” di una struttura, e le colloca *tra l’osservatore e il mondo*. Se le immagini fin qui riprodotte non offrivano alcuna idea di come funzionassero i testi, nel diagramma dell’*Idiota* della figura 18 l’organizzazione interna del romanzo è letteralmente *tutto*: l’autore della ricerca – un pianista che stava laureandosi in letteratura – era interessato al “volume” sonoro delle opere narrative, e aveva scoperto un modo assai elegante di operazionalizzarlo, suddividendo i *verba dicendi* in alti, neutrali e bassi,¹⁸ e ponendoli in rapporto con l’andamento dell’intreccio. Il lavoro analizzava anche “ciò che” i personaggi dicevano, beninteso, però soprattutto *come* lo dicevano: dal diagramma si capisce che, al termine di quei sedici capitoli, Rogozhin sta gridando e Myshkin bisbigliando – ma non si ha la minima idea di *che cosa* stiano dicendo. È una riduzione audace, che ha permesso di misurare, alla lettera, la metafora bachtiniana della “polifonia romanzesca”. *L’idiota*, era divenuto uno spartito.

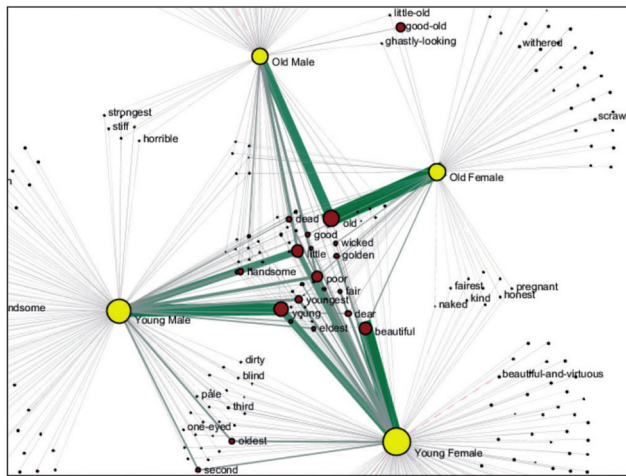
Fig. 18. “Il rumore nel romanzo” (2014)



Benché l’intensità sonora sia creata da un gran numero di individui distinti, [questa immagine] non è il grafico di una cacofonia. È possibile riscontrare [...] prima, un leggero picco nel capitolo 7, poi il picco drammatico nel capitolo 10, e infine una biforcazione del dialogo in due estremi: Rogozhin, Lebedev, Ferdyschenko, Dar’ja, Nastas’ja [...] riempiono la stanza di schiamazzi; Myshkin, Ivan Epanchin, Ptitsyn, e Tockij creano un sottofondo di bisbigli. È un vero e proprio grafico polifonico di voci “artisticamente organizzate” (Bachtin).

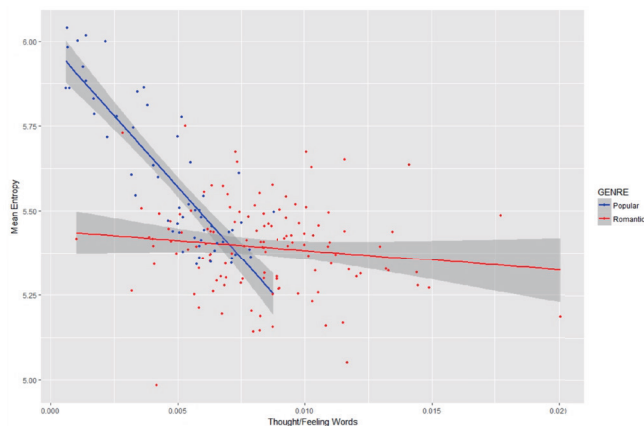
“Lo studio delle parti componenti [...] e delle loro relazioni reciproche”, si leggeva nel testo di Propp. In “Il rumore nel romanzo”, le parti componenti sono la voce e l’intreccio; in “Computational analysis of the body in European fairy tales” (2013), si tratta invece di una matrice di quattro attributi (giovane/vecchio, maschile/femminile: figura 19), che un fitto reticolo di connessioni collega ai principali aggettivi del corpus, mostrando come – in questo tipo di fiabe – l’opposizione tra maschile e femminile sia in realtà meno significativa di quella tra vecchio e giovane. In modo analogo, la figura 20 mette in evidenza la correlazione tra interiorità e ripetizione linguistica nella letteratura cinese e giapponese moderna, e la figura 21 quella tra sintassi e semantica in un corpus di romanzi inglesi dell’Ottocento.

Fig. 19. “Computational analysis of the body in European fairy tales” (2013)



Gli aggettivi all’interno della “stella”, come *poor* e *good*, sono condivisi da tre o quattro personaggi; altri sono condivisi da due di loro (come lo *small* che unisce i due personaggi giovani, o il *ghastly-looking* dei due vecchi); e infine alcuni (*beautiful-and-virtuous* per la donna giovane, o *scrawny* per quella vecchia) riguardano un solo personaggio. Presi nel loro insieme, tutti questi termini esemplificano il tessuto di giudizi di valore che soggiace a questo genere di fiabe.

Fig. 20. “Self-Repetition and East Asian Literary Modernity 1900-1930” (2018)



Rapporto tra termini legati al pensiero e al sentimento e entropia linguistica in Giappone e Cina, con rispettive regressioni lineari. In entrambi i casi, si può osservare come, all'aumento dei termini legati al pensiero e al sentimento lungo l'asse orizzontale, l'entropia media dei testi declina, indicando un aumento della ripetizione lessicale.

Fig. 21. “Lo stile al livello della frase” (2013)



Un esempio di quel che è divenuta la letteratura nel nuovo spazio dei laboratori critici, dove i romanzi vengono “preparati” per l’analisi in un modo che rescinde ogni legame con l’esperienza concreta della lettura. In questo diagramma, frasi e parole sono estratte dal contesto narrativo, combinate tra loro, e misurate, in modo da verificare l’esistenza di un rapporto tra sequenze sintattiche e campi semantici.

Tempo senza storia

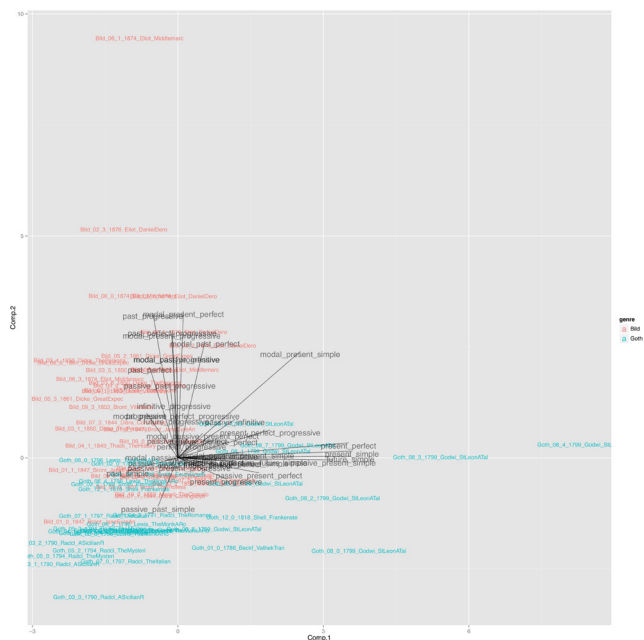
“Nella storia economica” ha osservato Ernest Labrousse “a differenza di quanto si osserva in altri ambiti della storia, tutto ciò che è importante è ripetuto.”¹⁹ Nella morfologia, vale lo stesso principio: i suoi oggetti d’indagine preferiti – i *patterns* – emergono appunto dalla regolare ripetizione di un processo nel corso del tempo.²⁰ Ma è un “tempo” diverso da quello della storia economica, o delle serie temporali che abbiamo visto più sopra. I 250 romanzi della figura 21 erano stati tutti pubblicati nel corso dell’Ottocento, e sarebbe stato facile aggiungere all’immagine le loro date di pubblicazione; ma eravamo così presi dai meccanismi interni di quelle frasi che la storia non ci venne proprio in mente. Il nostro lavoro dipendeva dal tempo – dal ripetersi delle stesse scelte lessicali a distanza di anni e decenni – ma non dal tempo *dello storico*: quel che ci serviva, *era il tempo del laboratorio*. Tempo astratto: poche variabili, processate infinite volte al fine di comprendere “operazioni e interazioni”. Era il tempo dell’esperimento, da tenere rigorosamente *separato* – da *proteggere* – dal tempo del mondo.²¹ Tempo, senza storia.

Nella prossima sezione vedremo che la morfologia non è inevitabilmente prigioniera di questo tempo astratto; qui osserviamo intanto come la svolta quantitativo-computazionale abbia al tempo stesso offerto una grandissima occasione allo studio morfologico, ma ne abbia anche ingigantito i problemi di fondo. Un’occasione, perché l’operazionalizzazione delle categorie estetiche le ha rese più vive che mai: “polifonia”, *Pathosformel*, “personaggio minore”: quantificare queste astrazioni le ha rese finalmente *visibili* – e solo il tempo “protetto” del laboratorio poteva riuscire nell’impresa. Ma le condizioni stesse di tale successo – il drastico, artificiale *isolamento* delle forme dalla vita – ha al tempo stesso aperto una voragine tra la ricerca morfologica e la riflessione storica. La forma come l’aspetto più profondamente sociale dell’opera d’arte: quest’idea, che sembrava così vicina mezzo secolo fa, è rimasta estranea alle *digital humanities*, divise come sono tra il contenutismo semplicistico descritto più sopra, e l’astratto formalismo da laboratorio di queste ultime pagine. Una vecchia maledizione è tornata sotto nuove vesti.

“Lo studio delle parti componenti [...] e delle loro relazioni reciproche e con il tutto”, scrisse Propp nella *Morfologia*. Le figure 18, 19, 20 e 21, con le loro copie di variabili espresse lungo i due assi cartesiani, o distribuite su un grafico a dispersione del componente principale, sono un’ottima illustrazione delle “relazioni *reciproche*” tra le parti componenti; ma nessuna di loro – e, in realtà, nessuno studio di cui siamo a conoscenza – è in grado di visualizzare “le relazioni *con il tutto*”. A volte si è riusciti a “condensare” una struttura entro un unico tratto, come nelle forme verbali dei romanzi gotici e dei romanzi di formazione nella figura 22; altrove, la base della comparazione è divenuta più complessa, come nel convergere di diversi tipi di misurazione entro la “focalizzazione tematica” della figura 23, e il “protagonismo” della figura 24. Al fondo, però, anche queste

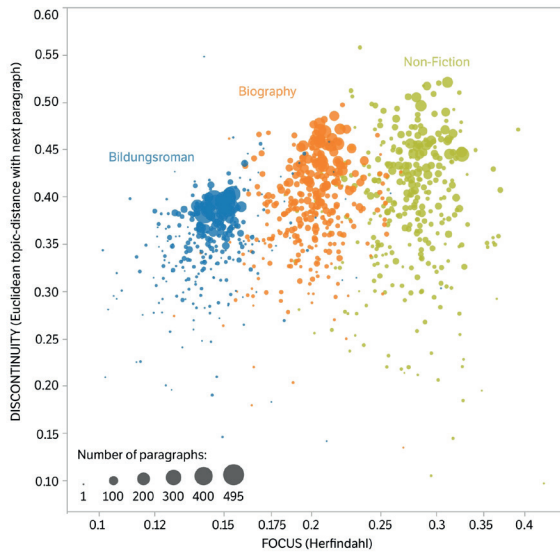
immagini hanno eluso il problema di come le “parti componenti” interagiscano fino a formare una struttura complessa. Nel ridurre il tutto ad alcune sue parti, questo tipo di visualizzazione offre un’utile *semplificazione* delle strutture in esame, più che una vera e propria analisi. Per visualizzare la *Morfologia* resta ancora molta strada da fare.

Fig. 22. “Lo stile al livello della frase” (2013)



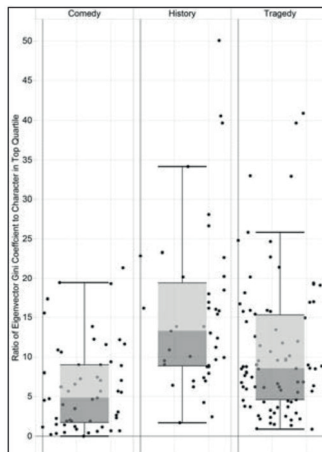
In blu i romanzi gotici, raggruppati nel quadrante in basso a sinistra, e in rosso i *Bildungsromane*, presenti in quello superiore [...] Le unità presenti nel diagramma (Goth_03_0_1790_Radcl_AsicilianR nell’angolo sinistro in basso della 6.1 o Bild_06_1_1874_Eliot_Middlemarc in quello in alto a sinistra) sono segmenti dei romanzi inclusi nel nostro database, ognuno dei quali contiene 200 frasi narrative. Le forme verbali cui è dovuta la separazione fra i due generi sono il *present perfect* e il *passive past simple* per il gotico, e le forme modali e progressive per il *Bildungsroman*.

Fig. 23. “Sui paragrafi. Scale, temi e forme della narrazione” (2015)



L'asse delle x misura la focalizzazione tematica dei paragrafi (vale a dire, quanto spazio sia riservato a un singolo tema) e quello delle y la loro discontinuità (vale a dire, la differenza tra i temi presenti tra paragrafi spazialmente contigui). La separazione tra i tre tipi di discorso – e soprattutto tra la narrativa e la saggistica – è inconfondibile.

Fig. 24. “Distributed Character: Quantitative Models of the English Stage, 1550-1900” (2017)



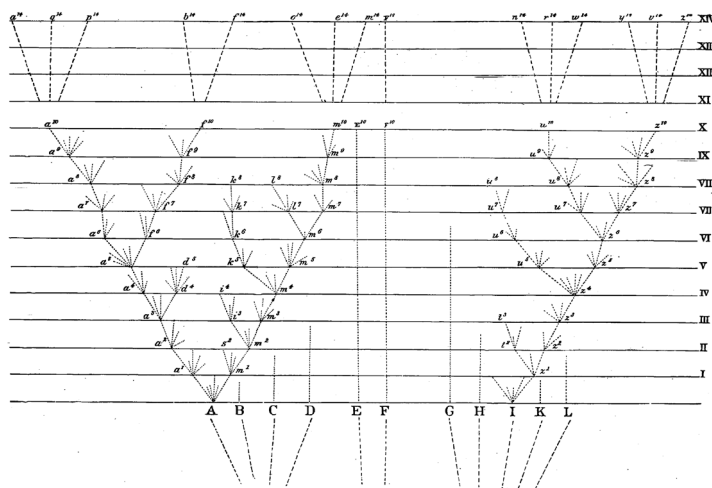
Il rapporto tra centro e periferia è simile in *history plays* e tragedie, e molto diverso per le commedie, dove la periferia tende a essere molto più ristretta.

III. Morfologia storica?

Una sezione sulla storia, una sezione sulla morfologia. Distinte, perché i loro scopi sono anch'essi distinti, e per certi aspetti anche opposti. La morfologia va in cerca di specificazioni e correlazioni: tratti dal cui interagire emerge una struttura particolare, via via fino a quel vero e proprio sistema di distinzioni che è una tassonomia. Gli studi storici puntano invece alla continuità e alla successione: più che "aprire" un oggetto per comprenderne il meccanismo, ne seguono le vicissitudini lungo il corso del tempo – nelle *digital humanities*, sintetizzandole di solito in un'unica *trend line*. E uno si chiede: che forma potrebbe mai prendere *l'unione di morfologia e storia*?

La figura 25 riproduce l'unica immagine presente nell'*Origine delle specie* (nel capitolo sulla "Selezione naturale"). Il "diagramma", scrive Darwin, aiuta a rendere intuitivamente visibile il rapporto tra lo scorrere del tempo (misurato in migliaia di generazioni lungo l'asse verticale), e la crescente differenziazione delle specie A e I indicate in calce all'immagine. Quel che ne emerge è insomma *l'inestricabile interconnessione di storia e morfologia che caratterizza il mondo della natura*. Una specie è quel che è perché è divenuta quel che è. Non esistono forme senza storia.

Figura 25: *L'origine delle specie*, Capitolo IV: "Selezione naturale"



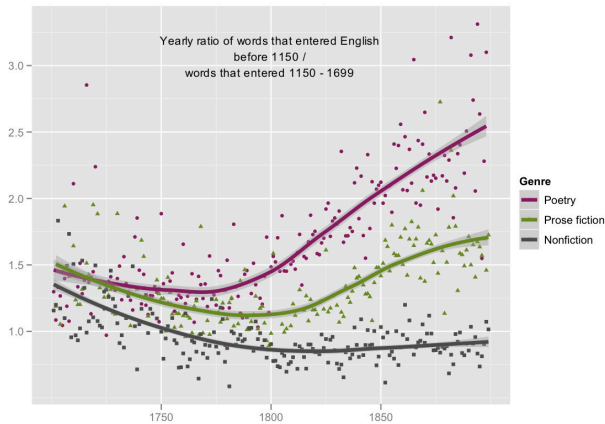
Troverete il Cap. IV strano e incomprensibile, senza l'ausilio del bizzarro diagramma, di cui accludo una vecchia & inutile bozza di stampa.

Charles Darwin a Charles Lyell, lettera del 2 settembre 1859.

Non esistono forme senza storia. Ma tra le centinaia di immagini del nostro corpus, ce ne è sì e no una manciata che abbia una qualche somiglianza col diagramma di Darwin. Una di queste, alla figura 26, misura la differenziazione del lessico

della poesia, narrativa, e saggistica inglesi tra il 1700 e il 1900, sulla base della percentuale di termini entrati a far parte della lingua prima o dopo il 1150 (vale a dire, prima o dopo la conquista normanna). Ne emerge un *pattern* che ricorda l'immagine dell'*Origine delle specie*, la separazione tra reati violenti e non-violenti dell'Old Bailey (figura 15, più sopra), e la divergenza nell'uso delle anacronie nel cinema degli ultimi 40 anni della figura 27.²²

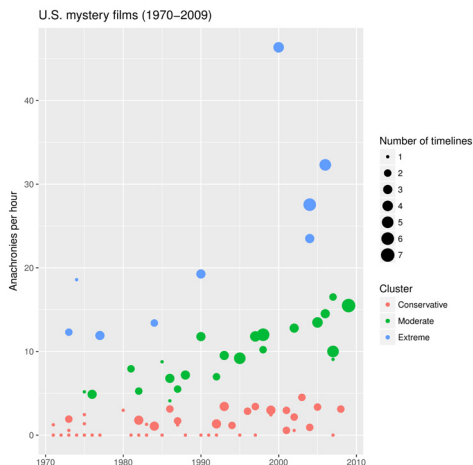
Fig. 26. "The Emergence of Literary Diction" (2012)



Variazioni annuali nell'uso dei termini presenti nella lingua inglese prima del 1150, calcolate su un corpus di oltre 4000 volumi pubblicati tra 1700 e 1900.

"The Emergence of Literary Diction", "The civilizing process in London's Old Bailey", e "Il tempo a pezzi" hanno tre tratti in comune. L'attenzione alla morfologia li distingue dalla corrente "contenutistica" delle *digital humanities*; la correlazione tra il modificarsi degli elementi strutturali e il passare degli anni permette di evitare il "tempo astratto" della morfologia da laboratorio; e, infine, tutti e tre si ispirano a ciò che Ernst Mayr ha chiamato "population thinking". "Per comprendere l'origine della biodiversità" scrive Mayr "non basta studiare 'verticalmente' una sola popolazione in momenti diversi, ma è necessario confrontare tra loro diverse popolazioni di una specie che siano contemporanee tra loro".²³ Popolazioni contemporanee: invece di studiare una sola "popolazione" di, poniamo, assassini, "The Civilizing Process in London's Old Bailey" segue il corso simultaneo di due distinti tipi di reati; "The Emergence of Literary Diction", di tre generi di discorso; "Broken Time", di tre forme cinematografiche. Mettendo a confronto tra loro queste diverse "popolazioni culturali" a intervalli regolari, prende forma davanti ai nostri occhi tutto un sistema di analogie e divergenze. Con un po' di fortuna, ci si trova ad assistere alla nascita di una nuova specie.²⁴

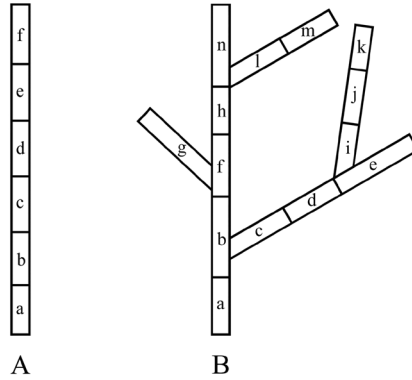
Fig. 27. “Il tempo a pezzi. Anacronie nel cinema contemporaneo” (2017)



In questo studio, il processo di speciazione fu trovato per caso: l'ipotesi di ricerca era infatti quello di un aumento regolare – un trend: l'ipotesi-base delle *digital humanities* – nell'uso delle anacronie (*flashback* e *flashforwards*). Via via, divenne chiaro che mentre alcuni film (i puntini blu e verdi) mostravano appunto le tracce di tale aumento, un altro gruppo (i puntini rossi) non mostrava invece alcun aumento di rilievo. La cosa più significativa era che la differenza tra i tre gruppi non si limitava al numero delle anacronie, ma investiva anche la loro posizione all'interno dell'intreccio e la loro funzione narrativa. Nel gruppo rosso, *flashback* e *flashforwards* erano quasi interamente localizzati all'inizio e/o alla fine del film, e servivano quasi sempre a spiegare un mistero rendendo visibile la scena originaria. Nei gruppi blu e verde, per contro, le anacronie erano distribuite piuttosto regolarmente lungo tutto l'arco del film, e sembravano aver assunto la funzione, del tutto diversa, di moltiplicare gli enigmi disseminandoli un po' ovunque.

Nella figura 28, Mayr chiarisce le conseguenze di questo modo di ragionare. Nell'evoluzione filetica (A), una specie evolve nel tempo, passando da *a* a *f* nel suo adattarsi ai mutamenti dell'ambiente; nonostante tali trasformazioni rimane però sempre *una sola specie*. Non c'è aumento di biodiversità. Per questo “developmental thinking”, come l'ha chiamato Robert O'Hara, “la storia [è] un processo di sviluppo al singolare – una ‘evoluzione’ nel senso originale del termine.”²⁵ Un solo filo lega fra loro le fasi che si susseguono – *Australopithecus*, *Homo habilis*, *Homo erectus*, ... Classicismo, Romanticismo, Realismo, Simbolismo Niente rami, ma un tronco diviso in diverse sezioni – un po' come nei *trends* discussi più sopra, dove un'unica linea veniva ritenuta sufficiente a riassumere la storia di tutta una popolazione. Nel caso della speciazione (B), invece – dove la specie iniziale *a* dà vita a *g*, *n*, *m*, *k*, *e* ... – la presenza dei diversi rami rende immediatamente visibile un aumento della diversità. Le popolazioni della poesia e della narrativa divergono da quella della saggistica; i reati violenti, da quelli non violenti. La storia diventa un processo di diversificazione creativa.²⁶

Fig. 28. Evoluzione filetica vs. speciazione



Rami, tronchi, reticoli ... ma alla fin fine, che interesse può mai avere la “forma” della storia? Questo: che essa è di fatto *un’ipotesi su quali siano le forze che agiscono all’interno della storia*; è l’inizio, per quanto ancora solo intuitivo, di un impianto concettuale. Quando “vediamo” la storia della cultura come la Scala della ragione, sappiamo che dovremo spiegare sia il *trend* rettilineo che gli archi sbilenchi – e che è difficile siano stati prodotti da un unico meccanismo. Se vediamo dei “rami”, dobbiamo capire come e perché siano emersi. La “forma” della storia pone degli interrogativi che vanno al di là delle spiegazioni di questo o quel risultato: degli interrogativi *teorici*. “Ancor più degli strumenti di laboratorio” scrisse Thomas Kuhn più di mezzo secolo fa “le teorie sono gli attrezzi essenziali del mestiere di scienziato”;²⁷ se solo gli avessimo dato retta! Anche se, per fortuna, le forme più grevi di anti-intellettualismo – come i proclami della rivista *Wired*: “la correlazione basta e avanza”; “il metodo scientifico è obsoleto”²⁸ – sono rimaste un’eccezione, con l’aumento dei dati quantitativi è successo che le spiegazioni “forti” si sono fatte sempre più rare, sopraffatte da un’ondata di dilazioni, false modestie, ragionamenti *ad hoc*, e via svicolando. Abituati come siamo a vedere i dati empirici e le spiegazioni teoriche come le due facce della stessa medaglia, a prima vista, la cosa sembra impossibile. Purtroppo, è la realtà delle *digital humanities*. Il giorno che l’abbondanza di dati aprirà la strada a concetti ambiziosi, invece di scoraggiarli – quel giorno, la storia quantitativa della cultura avrà davvero inizio, e il confronto con le “altre” discipline umanistiche diventerà finalmente possibile.

APPENDICE

Lista per anno di pubblicazione degli articoli, interventi e materiali citati

2011

James E. Cutting et al, "Quicker, Faster, Darker: Changes in Hollywood Film over 75 years", *i-Perception*, vol. 2

Jean-Baptiste Michel et al, "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science*, vol. 331, n. 6014

2012

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

Ted Underwood e Jordan Sellers, "The Emergence of Literary Diction", *Journal of Digital Humanities*, I, 2

2013

Sarah Allison et al, "Style at the Scale of the Sentence", *Stanford Literary Lab*, Pamphlet 5

Olivier Morin, "How Portraits Turned Their Eyes upon Us: Visual Preferences and Demographic Change in Cultural Evolution", *Evolution and Human Behavior*, vol. 34

Scott Weingart e Jeana Jorgensen, "Computational Analysis of the Body in European Fairy Tales", *Literary and Linguistic Computing*, XXVIII, 3

2014

Jordan E. DeLong, Kaitlin L. Brunick e James E. Cutting, "Film Through the Human Visual System: Finding Patterns and Limits", in J. C. Kaufman e D. K. Simonton, a cura di, *The Social Science of Cinema*, Oxford University Press, Oxford

Andrew Goldstone e Ted Underwood, "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us", *New Literary History*, XLV, 3

Holst Katsma, "Loudness in the Novel", *Stanford Literary Lab*, Pamphlet 7

Sara Klingenstein, Tim Hitchcock e Simon DeDeo, "The Civilizing Process in London's Old Bailey", *Proceedings of the National Academy of Sciences of the United States of America*, CXI, 26

2015

Mark Algee-Hewitt, Ryan Heuser e Franco Moretti, "On Paragraphs: Scale, Themes and Narrative Form", *Stanford Literary Lab*, Pamphlet 10

James E. Cutting e Aysel Candan, "Shot Durations, Shot Classes and the Increased Pace of Popular Movies", *Projections*, IX, 2

Anne Dewitt, "Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press", *Victorian Periodicals Review*, XLVII, 2

Franco Moretti e Dominique Pestre, "Bankspeak: The Language of World Bank Reports", *New Left Review* 92 (March–April)

2016

Mark Algee-Hewitt et al, "Canon/Archive: Large-scale Dynamics in the Literary Field", *Stanford Literary Lab*, Pamphlet 11

James F. English, "Now, Not Now: Counting Time in Contemporary Fiction Studies", *Modern Language Quarterly*, LXXVII, 3

Marissa Gemma, "The Making of Middle American Style: Narrative Talk in the 19th Century Novel", presentation at the Stanford Literary Lab, 15 February

Hoyt Long e Richard Jean So, "Turbulent Flow: A Computational Model of World Literature", *Modern Language Quarterly*, LXXVII, 3

Olivier Morin e Alberto Acerbi, "Birth of the Cool: A Two-Centuries Decline in Emotional Expression in Anglophone Fiction", *Cognition and Emotion*, XXXI, 8

Ted Underwood e Jordan Sellers, "The *Longue Durée* of Literary Prestige", *Modern Language Quarterly*, LXXVII, 3

2017

Mark Algee-Hewitt, "Distributed Character: Quantitative Models of the English Stage, 1550–1900", *New Literary History*, XLVIII, 4

Marissa Gemma, Frédéric Glorieux e Jean-Gabriel Ganascia, "Operationalizing the Colloquial Style: Repetition in 19th Century American Fiction", *Digital Scholarship in the Humanities*, XXXII, 2

Leonardo Impett e Franco Moretti, "Totentanz: Operationalizing Aby Warburg's *Pathosformeln*", *New Left Review* 107 (September–October)

Maria Kanatova et al, "Broken Time, Continued Evolution: Anachronies in Contemporary Films", *Stanford Literary Lab*, Pamphlet 14

Thomas Lansdall-Welfare et al, "Content Analysis of 150 Years of British Periodicals", *Proceedings of the National Academy of Sciences of the United States of America*, CXIV, 4

Grace Muzny, Mark Algee-Hewitt e Dan Jurafsky, "Dialogism in the Novel: A Computational Model of the Dialogic Nature of Narration and Quotations", *Digital Scholarship in the Humanities*, XXXII, 2

2018

Taylor Arnold e Lauren Tilton, "Distant Viewing: Analyzing Large Visual Corpora" [under review] [Ora pubblicato: *Distant Scholarship in the Humanities*, fqz013, <https://doi.org/10.1093/digitalsh/fqz013>, 16 March 2019]

Jonah Berger e Grant Packard, "Are Atypical Things More Popular?", *Psychological Science*, XXIX, 7

Jo Guldi, "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora", *Journal of Cultural Analytics*, 20 dicembre

Hoyt Long, Anatoly Detwyler e Yuancheng Zhu, "Self-Repetition and East Asian Literary Modernity, 1900–30", *Journal of Cultural Analytics*, 21 maggio

Ted Underwood, David Bamman e Sabrina Lee, "The Transformation of Gender in English-Language Fiction", *Journal of Cultural Analytics*, 13 febbraio

NOTE

* Franco Moretti ha insegnato in varie università italiane e statunitensi ed è autore di molti studi di critica letteraria quasi tutti pubblicati da Einaudi. Attualmente è Permanent Fellow presso il Wissenschaftskolleg di Berlin. Il suo ultimo volume è *Un paese lontano*, uscito nel 2019 da Einaudi. Su *Ácoma* ha recentemente pubblicato "Giorno e notte. Sull'antitesi tra western e film noir" (*Ácoma* n.s., n. 14, 2018).

Oleg Sobchuk è ricercatore presso il Max Planck Institute for the Science of Human History di Jena. Si occupa di evoluzione dell'arte su larga scala, servendosi di metodi computazionali.

Questo saggio è apparso con il titolo "Hidden in Plain Sight" sulla *New Left Review* n. 118 (July/August 2019).

1 Il corpus su cui abbiamo lavorato rispondeva a criteri diversi. Innanzitutto, ci siamo limitati alle discipline umanistiche dove la svolta quantitativa è un fatto recente, escludendo dunque sia la linguistica sia la storia economico-sociale. Abbiamo anche escluso gli studi dedicati a un singolo testo o autore, e quelli di cui eravamo gli unici autori. Su questa base, abbiamo passato in rassegna tutte le maggiori riviste di *digital humanities* dal 2010 a oggi, ricavandone un campione di 62 saggi, la metà dei quali è discussa in questo articolo (e, oltre i necessari riferimenti in nota, è elencata nell'appendice bibliografica), mentre l'altra metà si può trovare sul sito della *New Left Review*.

Abbiamo cercato di ragionare ad ampio raggio, coinvolgendo diversi campi disciplinari; poiché siamo però entrambi specialisti di letteratura, è di lì che viene la maggior parte dei nostri dati, il che naturalmente indebolisce alcune delle nostre tesi. Per di più, le scelte su cui si basa il nostro corpus sono chiaramente soggettive: criteri diversi avrebbero prodotto un corpus e delle conclusioni altrettanto diverse. Infine, quando la versione inglese di questo articolo era già finita, *Critical Inquiry* XLV, 3 (spring 2019) ha pubblicato un altro panorama delle *digital humanities* – "The Computational Case against Computational Literary Study" di Nan Z. Da (pp. 601-39). Visto che gli scopi e i metodi dei due articoli sono completamente diversi, non abbiamo ritenuto di dover modificare quanto già scritto.

2 Chi legge deciderà se questo sia effettivamente il caso delle *digital humanities*; per parte nostra, ne siamo convinti. Le discipline umanistiche sono state prese completamente alla sprovvista dall'arrivo degli archivi digitali e dei metodi di analisi computazionali, che non conoscevano, e di cui, soprattutto, non avevano alcun bisogno. Si dice spesso che i nuovi metodi possono svolgere all'interno delle nostre discipline un ruolo paragonabile a quello del telescopio per l'astronomia della prima età moderna; ma si trascura il fatto, decisivo, che la teoria astronomica *necessitava* di qualcosa come il telescopio – e le nostre discipline no. Dato un simile punto di partenza, era logico che l'uso concreto dei nuovi strumenti – cioè appunto la pratica – prendesse rapidamente il sopravvento sulla loro giustificazione teorica.

3 Volume 331, numero 6014, pp. 176-82.

4 In questo articolo, i titoli in italiano si riferiscono ai lavori dello Stanford Literary Lab tradotti in *La letteratura in laboratorio*, a cura di Giuseppe Episcopo, Federico II University Press, Napoli 2019.

5 Andrew Goldstone e Ted Underwood, "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us", *New Literary History* XLV, 3 (summer 2014), pp. 359-384, p. 359.

- 6 Daniel Rosenberg e Anthony Grafton, *Cartografie del tempo*, Einaudi, Torino 2012 [2010], p. 13.
- 7 Grace Muzny, Mark Algee-Hewitt e Dan Jurafsky, "Dialogism in the Novel: A Computational Model of the Dialogic Nature of Narration and Quotations", *Digital Scholarship in the Humanities*, XXXII, 2 (December 2017), pp. ii31-ii52, p. ii36.
- 8 James E. Cutting e Ayse Candan, "Shot Durations, Shot Classes and the Increased Pace of Popular Movies", *Projections*, IX, 2 (winter 2015), pp. 44-45.
- 9 Benché il canone e le medie statistiche siano naturalmente cose ben diverse, impongono entrambi, ognuno a suo modo, una drastica semplificazione al campo letterario. "Le medie" ha scritto Ernst Mayr nel 1959 "sono solo delle astrazioni statistiche: solo gli individui di cui è composta una popolazione sono reali". Si veda "Darwin and the evolutionary theory in biology", in B. J. Meggers, a cura di, *Evolution and Anthropology: A Centennial Appraisal*, Anthropological Society of Washington, Washington 1959, p. 29.
- 10 Goldstone e Underwood, "The Quiet Transformations of Literary Studies", cit., p. 359.
- 11 James E. Cutting e Ayse Candan, "Shot Durations, Shot Classes and the Increased Pace of Popular Movies", cit., p. 53.
- 12 Lo stesso vale per uno studio "parallelo" – "Quicker, faster, darker. Changes in Hollywood films over 75 years" (2011) – che rileva "un aumento graduale e fondamentalmente lineare" nel "visual activity index" cinematografico degli ultimi 75 anni. Visto che l'indice per *Toy Story* è 15 volte più alto di quello per *Barry Lindon*, e che film come *Toy Story* sono diventati immensamente più frequenti di quelli come *Barry Lindon*, è perfettamente possibile che la migliore spiegazione del fenomeno in esame debba prendere le mosse dall'aumento di film per bambini e "young adults".
- 13 Sara Klingenstein, Tim Hitchcock e Simon DeDeo, "The Civilizing Process in London's Old Bailey", *Proceedings of the National Academy of Sciences of the United States of America*, CXI, 26 (July 2014), pp. 9419-9424, p. 9421.
- 14 "Turbulent Flow: A Computational Model of World Literature" (2016); "Distant Viewing: Analyzing Large Visual Corpora" (2018: under review); "Distributed Character: Quantitative Models of the English Stage, 1550–1900" (2017); "How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution" (2013).
- 15 "Are Atypical Things More Popular?" (2018); "Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press" (2015); "Totentanz: Operationalizing Aby Warburg's *Pathosformeln*" (2017); "Computational analysis of the body in European fairy tales" (2013).
- 16 Ernst Mayr, *Toward a new philosophy of biology*, Harvard University Press, Cambridge-London 1988, p. 25.
- 17 Vladimir Propp, *Morfologia della fiaba*, Einaudi, Torino 2000 [1927], p. 3.
- 18 Tre esempi tratti da una famosa scena di *Alice nel paese delle meraviglie* chiariscono la logica seguita nel saggio. Registro alto: "'Tagliategli la testa!' urlò la Regina". Neutrale: "'Me lo immagino', disse Alice". Basso: "'È condannata a morte', bisbigliò".
- 19 Labrousse è citato in un saggio retrospettivo di Krzysztof Pomian, dove si legge anche che, con Braudel, "lo studio delle ripetizioni esce dal campo in cui pareva confinato [...] e occupa l'intero orizzonte dello storico, o quasi." ("Storia delle strutture", in Jacques Le Goff, a cura di, *La nuova storia*, Mondadori, Milano 1980 [1979], pp. 93, 95). Avere a loro volta ampliato l'orizzonte della storia della cultura è stato uno dei principali risultati delle *digital humanities*; se però si guarda alla novità dei risultati e alla robustezza dei concetti, non c'è paragone con quel che il gruppo delle *Annales* – con archivi molto più angusti dei nostri, e strumenti computazionali poco più che primitivi – riuscì a fare due-tre generazioni fa.
- 20 I *patterns* hanno svolto all'interno della morfologia un ruolo assai simile a quello dei *trends* nella ricerca storica, venendo spesso considerati dei risultati validi *in sé* e *per sé*, laddove la vera sfida consiste nel portarne alla luce le cause nascoste. Per una discussione critica, si veda "Interpretare pattern", in *La letteratura in laboratorio*, cit.
- 21 "La tecnica chiave del biologo funzionale è l'esperienza", aveva scritto Mayr in *Toward a new philosophy of biology*, "e il suo modo di lavorare è essenzialmente lo stesso di quello del fisico o del chimico", cit., p. 25.

22 Qualche altro articolo contiene degli indizi di speciazione, senza che l'ipotesi sia sviluppata dagli autori. In "Quicker, faster, darker", per esempio, le trasformazioni del "visual activity index" vengono interpretate come se costituissero un singolo *trend* (il consueto "aumento dell'attenzione"), benché i dati presentati nella figura 1b dell'articolo suggeriscano un inizio di divergenza assai simile a quello di "Broken Time"; James E. Cutting et al, "Quicker, Faster, Darker: Changes in Hollywood Film over 75 years", *i-Perception*, II (September 2011), pp. 569-76, pp. 571, 572. Qualcosa del genere accade anche in "Film through the Human Visual System" e in "Now, Not Now", a proposito dei bestseller, e dei romanzi insigniti dei più prestigiosi premi letterari; Jordan E. DeLong, Kaitlin L. Brunick e James E. Cutting, "Film Through the Human Visual System: Finding Patterns and Limits", in J. C. Kaufman e D. K. Simonton, a cura di, *The Social Science of Cinema*, Oxford University Press, Oxford 2014, pp. 123-37; James F. English, "Now, Not Now: Counting Time in Contemporary Fiction Studies", *Modern Language Quarterly*, LXXVII, 3 (September 2016), pp. 395-418.

Le divergenze morfologiche nel poliziesco intorno al 1890, e nello stile indiretto libero tra 1800 e 2000, erano state discusse in dettaglio in "The Slaughterhouse of Literature" (2000; ora in *Distant Reading*, Verso, London-New York 2013) e *La letteratura vista da lontano* (Einaudi, Torino 2005). In questi lavori mancava però la dimensione esplicitamente quantitativa che qui ci interessa.

23 Ernst Mayr, *What Evolution Is*, London: Phoenix, 2001. "Per coloro che hanno accettato l'idea del *population thinking*" aggiunge Mayr altrove "la variazione da individuo a individuo all'interno della popolazione è la realtà della natura, laddove il valore medio [il "tipo"] è solo un'astrazione statistica". Ernst Mayr, *Toward a New Philosophy of Biology*, cit., p. 15.

24 Va da sé che l'unione di morfologia e storia non porta inevitabilmente a trovare dei casi di speciazione: la figura 11, per esempio, che segue canone e archivio lungo un secolo e più, non mostra il minimo segno di divergenza tra le due popolazioni. Per di più, l'esatto opposto della speciazione – l'evoluzione "reticolata", in cui rami diversi finiscono con l'intrecciarsi tra loro – è a sua volta sempre possibile. La cosa è discussa in dettaglio da Oleg Sobchuk in *Charting Artistic Evolution: An Essay in Theory* (Tartu University Press, Tartu 2018); per riassumerne l'assunto centrale, lo scambio di tratti diversi da un ramo all'altro dell'albero della cultura avviene di norma tra rami vicini tra loro, e diventa via via più raro con l'aumentare della distanza morfologica. Il ramo del romanzo gotico può facilmente combinarsi con quello del romanzo storico, un po' meno con il romanzo sentimentale (anche se *Northanger Abbey* dimostra che la cosa è possibile); e le difficoltà aumentano man mano che ci si sposta – i romanzi *silver-fork*, quelli industriali, e così via. Per parafrasare il microbiologo Eugene Koonin: quando guardiamo ai dati da vicino vediamo un grande reticolo; da lontano, vediamo un albero. (Eugene Koonin, *The Logic of Chance: The Nature and Origin of Biological Evolution*, Pearson Education, New York 2011, p. 164).

25 Robert O'Hara, "Population thinking and tree thinking in systematics", *Zoologica Scripta*, XXVI, 4 (1997), pp. 325–327.

26 La tesi esposta in questo paragrafo torna, da una diversa angolatura, alla questione già discussa in merito ai *trends* storici: in entrambi i casi obiettiamo a un tipo di visualizzazione che presenta lo sviluppo storico come un percorso unico.

27 Thomas S. Kuhn, "The Function of Measurement in Modern Physical Science" [1961], in *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago University Press, Chicago 1977, p. 208.

28 Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/.