Politics. Rivista di Studi Politici
www.rivistapolitics.eu
n. 22 (6), 2/2024, 133-144
DOI: 10.69116/22/09

ISSN 2785-7719
Guida editori s.r.l.

Digital Humanities, database testuali e storia del pensiero politico: potenzialità, limiti e nuovi scenari epistemologici

Digital Humanities, Textual Databases and the History of Political Thought: potentials, limitations and new epistemological scenarios

Fausto PROIETTI

Abstract

In the field of the so-called digital humanities, lexicometrics based on the use of new, gigantic online databases represents perhaps the most interesting aspect with respect to the potential for application to the history of political thought. This article presents some examples of recent studies based on the use of these resources, and lists the main problems reported so far by critics. The conclusion reflects on the new methodological scenarios, in particular with respect to historical semantics and contextualist history of political thought, which may open up following the advent of new digital tools.

Keywords

Online Databases - Lexicometry - Digitized Texts - Political Lexicon - Method

1. Agli albori di una rivoluzione metodologica

In questo intervento proverò a interrogarmi sul rapporto che può esistere tra l'ambito disciplinare in cui si colloca la mia attività di ricerca – la Storia del pensiero politico – e quell'ampio insieme di pratiche e strumentazioni che si usa ricomprendere sotto la denominazione di *digital humanities* o informatica umanistica, nonché sulle implicazioni che l'uso di questi strumenti comporta sul piano epistemologico; temi sui quali esiste già una bibliografia critica, peraltro non molto ampia, che richiamerò nel corso della trattazione. Inevitabilmente il mio ragionamento si muoverà a un livello generale e introduttivo, e si concentrerà, nella seconda e più ampia parte del testo, su uno specifico aspetto della questione: l'uso di alcuni giganteschi database testuali online per la ricerca.

Tra gli ambiti riconducibili alle *digital humanities* (Ciotti 2024), l'analisi quantitativa su testi digitalizzati condotta attraverso strumenti informatici è particolarmente rilevante per chi, come lo storico del pensiero politico, realizza le proprie ricerche in primo luogo

– e a volte esclusivamente – a partire da testi, e si interessa necessariamente a questioni di storia del lessico politico. È, dunque, su questo tipo di strumenti che mi soffermerò; mi pare opportuno chiarire preliminarmente che il carattere 'metodologico' della rivoluzione richiamata nel titolo del presente paragrafo intende riferirsi, in prima battuta, non alla dimensione teorico-epistemologica delle questioni evocate, ma al metodo inteso, in senso lato, come procedura adatta al raggiungimento dello scopo scientifico prefissato tramite l'utilizzo di idonei strumenti.

In effetti la linguistica dei *corpora*, ossia l'analisi testuale applicata a insiemi più o meno ampi di documenti, e la lessicometria, vale a dire la ricerca quantitativa realizzata, di norma, con strumenti informatici, da vari decenni hanno incrociato le loro traiettorie con quella della storia delle idee, delle culture e del pensiero politico. Ciò è avvenuto in due distinti momenti; in questo primo paragrafo richiamerò alcuni esempi relativi alla prima fase, che si può far coincidere con la seconda metà del Ventesimo secolo, periodo in cui specifici studiosi e/o *équipes* hanno iniziato a sperimentare le potenziali applicazioni della digitalizzazione di testi al fine di renderli più rapidamente disponibili per ricerche testuali raffinate e complesse.

Tra i risultati più significativi di questa fase 'pionieristica' è doveroso citare quelli ottenuti da uno studioso italiano, il gesuita Roberto Busa, principale artefice dell'*Index Thomisticus*, all'interno del quale l'intera opera di Tommaso d'Aquino è trascritta e, con l'ausilio di computer IBM, resa interrogabile per singoli lemmi. La prima versione dell'*Index*, alla quale Busa aveva già iniziato a lavorare nei primi anni '50, è pubblicata a stampa, in 56 volumi, tra il 1974 e il 1980; nel 2005 il *corpus* viene reso disponibile su internet, in forma ipertestuale liberamente navigabile[1]. Sulla scorta di questo modello sono innumerevoli i cantieri di ricerca che, nel corso degli ultimi decenni, hanno dato vita a ipertesti contenenti l'*opera omnia* o determinati scritti di un singolo autore. In effetti, la digitalizzazione del *corpus* relativo a uno specifico pensatore politico rappresenta una delle applicazioni più immediate e intuitive delle tecnologie informatiche alla storia delle idee; come ha sostenuto Ioannis V. Evrigenis, ciò consente di ampliare in modo esponenziale l'accuratezza e il grado di completezza delle edizioni critiche, mettendo ad esempio a confronto, in un unico ipertesto navigabile online, i manoscritti e le varie edizioni a stampa di un testo, nonché le sue traduzioni in differenti lingue (Evrigenis 2015, 196). Rientrano in quest'ambito sperimentazioni come quella condotta dall'*équipe* dell'École Normale de Lyon sotto la guida di Séverine Gedzelman e Jean-Claude Zancarini (Gedzelman e Zancarini 2011) per lo sviluppo del *software* HyperMachiavel, in grado di comparare, parola per parola, il testo del *Principe* a quelli delle prime traduzioni francesi.

[1] Vedi https://www.corpusthomisticum.org/it/index.age.

Un secondo aspetto di grande rilievo che emerge fin dagli albori dell'applicazione della strumentazione lessicometrica allo studio della storia delle idee politiche è la possibilità di ricostruire e indagare in una prospettiva nuova il lessico politico utilizzato in contesti storici dati, non da singoli autori ma da gruppi ideologici. È il caso, in particolare, del lavoro portato avanti a partire dal 1964 dal Centre de lexicologie politique dell'École Normale Supérieure di Saint-Cloud, la cui denominazione è stata più volte modificata nel corso del tempo. Nel 1980 questo gruppo di studiosi, tra i quali spiccano le figure di Maurice Tournier, Jacques Guilhaumou e Raymonde Monnier, fonda *Mots*, rivista dedicata allo studio, come recita il sottotitolo, di «parole, computers, testi e società»; una grande novità nel panorama dell'editoria scientifica dell'epoca. Dedicati soprattutto al lessico politico della Rivoluzione francese, i principali studi di questa *équipe* comprendono il *Dictionnaire des usages socio-politiques (1770-1815)*, del quale sono usciti, tra il 1985 e il 2006, otto volumi. In molti, anche se non in tutti gli studi raccolti in quella serie, la strumentazione con cui vengono affrontati i vari lemmi politici è di tipo informatico e lessicometrico. Muovendo da una prospettiva diversa, e lavorando sulla *longue durée* anziché su contesti cronologici più ristretti, la possibilità di applicare tecniche computazionali e di *data mining* alla storia delle idee è stata, più recentemente, esplorata da Arianna Betti e Hein van den Berg (Betti e van den Berg 2016).

Per concludere questa rapida carrellata sulla fase che ho definito 'pionieristica' dell'applicazione di strumenti digitali allo studio della storia delle idee politiche, è utile ricordarne un terzo campo di applicazione. L'inferenza lessicometrica è stata impiegata per risolvere uno dei problemi più ricorrenti nell'ambito storiografico di cui stiamo trattando: l'attribuzione a uno specifico autore di un testo anonimo. L'esempio più noto di questo tipo di studi è costituito dal saggio di Frederick Mosteller e David Lee Wallace *Inference and disputed authorship: The Federalist* (Mosteller e Wallace 1964), nel quale strumenti di analisi quantitativo-lessicale sono utilizzati per individuare gli autori di alcuni degli articoli del *Federalist* la cui attribuzione era, fino a quel momento, dubbia; il metodo utilizzato dai due studiosi consiste nel mettere a confronto gli usi lessicali presenti in uno specifico testo con quelli derivati dall'analisi di opere il cui autore è noto, in modo di arrivare, utilizzando la statistica bayesiana, a un'attribuzione dotata di un elevato grado di attendibilità probabilistica.

2. Il XXI secolo, era dei database online

Tutti gli studi ricordati finora si basavano su una metodologia di tipo, per così dire, artigianale: la loro realizzazione ha comportato, cioè, la preliminare costruzione da parte dei ricercatori stessi di un *corpus* di fonti informatizzate e, in qualche caso, l'elaborazione di specifici *software* che ne consentissero l'indagine. Declinato in questa forma, il lavoro di ricerca richiede una lunga fase preparatoria: i testi da includere nel *corpus* devono

essere trascritti e sottoposti a una complessa operazione di *text mining*, comprendente, in particolare, la cosiddetta tokenizzazione, ossia la trasformazione di ogni parola del testo in un *token*, una unità riconoscibile come tale da parte del software di analisi. Si tratta, come è evidente, di operazioni che richiedono uno specifico *know-how* per essere realizzate; questa è la ragione principale per cui, in una prima fase, l'utilizzo di queste metodologie è stato limitato a singoli studiosi o, più spesso, a *équipes* di ricercatori in possesso di competenze di tipo informatico.

Questo scenario si è radicalmente trasformato dopo l'avvento di database online, che, unendo a un *corpus* di testi digitalizzati un'interfaccia di ricerca più o meno avanzata, consentono a qualsiasi utente di effettuare analisi testuali. Rispetto ai *corpora* interrogabili online come, ad esempio, quelli delle opere di Tommaso d'Aquino e di Machiavelli cui si è fatto riferimento in precedenza, questi database hanno, inoltre, la caratteristica distintiva di comprendere al loro interno un'enorme quantità di testi. Un altro elemento rilevante, che rende questi archivi informatici proficuamente utilizzabili da parte degli storici delle idee, è il fatto che in essi la ricerca testuale si accompagna alla disponibilità dei pdf riproducenti, anche in formato immagine, i documenti originali scansionati. In tal modo, com'è ovvio, alcuni aspetti (dimensioni del foglio, qualità della carta, legatura) non saranno apprezzabili, ma l'esperienza del contatto diretto con il testo nella sua 'materialità' – la «bookishness», come l'ha definita Stephen H. Gregg (Gregg 2020, 36) – sarà almeno in parte preservata, contrariamente a quanto accade se si lavora su trascrizioni. Le potenzialità maggiori per la ricerca sono legate, a mio avviso, ad alcuni tra i più grandi, in termini di testi in essi contenuti, di questi database; nella trattazione che segue, dunque, mi limiterò a questi, nella consapevolezza che l'uso di altri database, più ristretti e specifici (come, ad esempio, la Biblioteca Italiana[2] che contiene, ad oggi, poche migliaia di titoli), può servire a integrare l'analisi condotta sui primi.

Il primo, in ordine cronologico, tra questi giganteschi database è Gallica. Sviluppato dalla Bibliothèque Nationale de France a partire dal 1997, contiene attualmente 857.00 volumi digitalizzati e si presenta come il più grande tra i database testuali ad accesso gratuito gestiti da enti pubblici; ad esso si è affiancato nel 2016 Retronews, database gemello – ma con accesso a pagamento – dedicato esclusivamente alla stampa periodica (contiene quasi tre milioni di fascicoli di giornali e riviste in lingua francese). Eighteenth Century Collections Online (ECCO)[3], database lanciato nel 2003 dalla Gale Publishing e comprendente, ad oggi, circa 200.000 volumi editi in Gran Bretagna tra il 1701 e il 1800 (Gregg 2020), è accessibile solo attraverso un'iscrizione a pagamento, e

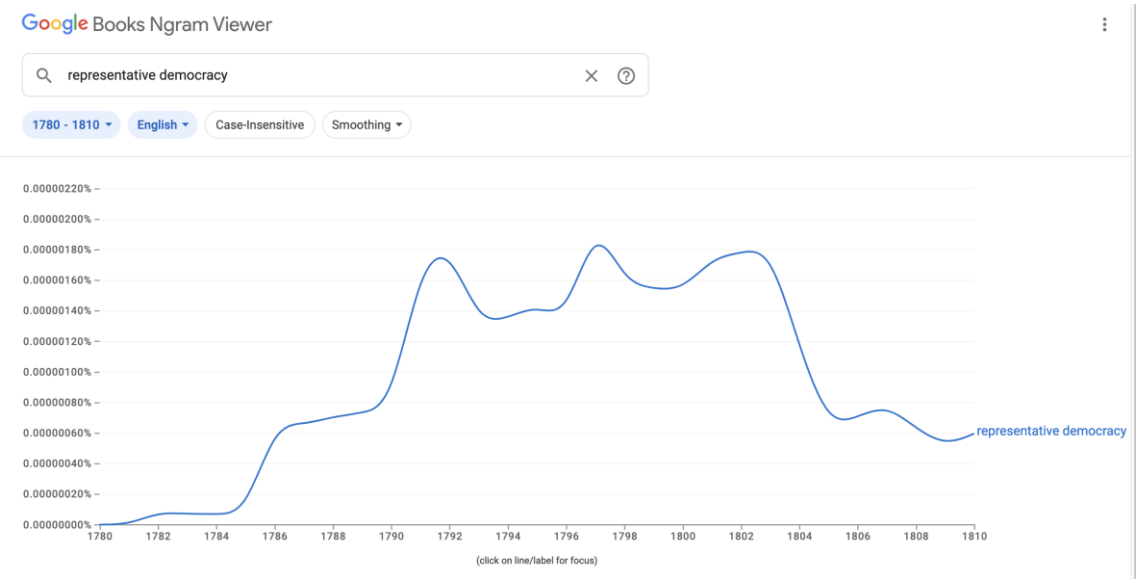
[2] http://www.bibliotecaitaliana.it
[3] https://www.gale.com/intl/primary-sources/eighteenth-century-collections-online

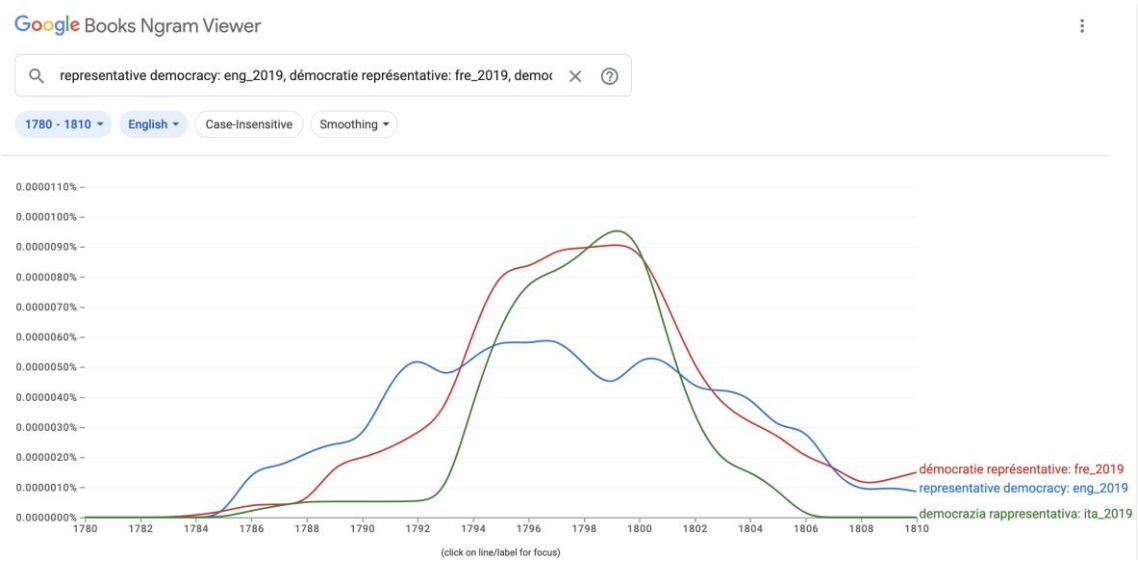
rappresenta la risorsa principale per gli studiosi impegnati in ricerche sul pensiero politico inglese del XVIII secolo.

Gratuito è, invece, l'accesso al più ampio tra i database testuali attualmente esistenti: Google Books, varato nel 2005 da Google LLC con l'ambizioso obiettivo di catalogare e indicizzare l'intero patrimonio librario mondiale[4]. Attraverso un motore di ricerca che si presenta graficamente del tutto simile a quello generale dell'azienda, qualsiasi utente ha accesso a un database contenente i testi digitalizzati di decine di milioni di libri. La consistenza del *corpus*, quantificata nel 2011 in 15 milioni di documenti – stimati corrispondere a circa il 12% di tutti i libri stampati fino ad allora (Michel *et al.* 2011, 177) – ha superato già nel 2019 i 40 milioni, stando a quanto dichiarato dalla stessa azienda[5]; al momento non sono disponibili stime più aggiornate. Google Books si presenta, inoltre, come l'unico archivio di testi digitalizzati in un numero molto ampio di lingue; i gestori ne dichiarano oltre 500, ma sono 46 quelle effettivamente selezionabili nella maschera di ricerca avanzata. La straordinaria ampiezza di questo database è determinata dal fatto che, pur essendo il progetto di proprietà di una compagnia privata come Google, hanno collaborato e collaborano ad esso alcune delle principali biblioteche del mondo occidentale, compresa la Biblioteca Nazionale Centrale di Roma[6].

Un altro strumento interessante sviluppato da Google a partire dal database Google Books è il generatore di grafici Ngram Viewer[7], che permette agli utenti di creare una rappresentazione grafica della presenza nel *corpus* di Google Books di un certo termine o di una certa espressione, in una specifica lingua (in questo caso, le lingue selezionabili sono 8) lungo un arco cronologico dato. Negli esempi riprodotti qui sotto, sono visualizzati il numero di occorrenze dell'espressione "representative democracy" nel *corpus* in lingua inglese tra il 1780 e il 1810, e la comparazione, per lo stesso intervallo cronologico, tra la frequenza di questo lemma e gli analoghi "démocratie représentative" e "democrazia rappresentativa" nei rispettivi *corpora*[8].

[4] https://books.google.it.
[5] https://blog.google/products/search/15-years-google-books.
[6] Per un elenco aggiornato al 2011 si veda il seguente link https://books.google.it/intl/it/googlebooks/partners.html.
[7] https://books.google.com/ngrams.
[8] Per ulteriori approfondimenti sull'utilizzo di questo strumento, che si presta alla realizzazione di grafici comparativi tra termini diversi, anche in più lingue, rimando a un altro studio di cui sono coautore (Zanettin e Proietti 2021).





Da segnalare che anche Gallica ha sviluppato nel 2021 uno strumento analogo, Gallicagram[9], esplicitamente presentato dai suoi creatori come ausilio alla ricerca nel campo delle scienze umane e sociali (Azoulay e De Courson 2021). Un database online per la lingua tedesca, simile ai precedenti ma di dimensioni più ridotte (contiene, ad oggi, poco meno di 100.000 documenti digitalizzati), è il *metacorpus* Historische Korpora[10], realizzato nell'ambito del progetto Digitales Wörterbuch der deutschen Sprache implementato dall'Accademia delle Scienze di Berlino-Brandeburgo a partire dal 1999;

[9] https://shiny.ens-paris-saclay.fr/app/gallicagram.
[10] https://www.dwds.de/d/korpora/dtaxl.

anch'esso consente la creazione di diagrammi sul modello Ngram Viewer, denominati Wortverlaufskurven (Curve di progressione delle parole).

Tutti questi database di nuovo tipo consentono all'utente di restringere, attraverso l'uso di apposite stringhe presenti nella maschera di ricerca, la sezione del *corpus* da interrogare (ad esempio, selezionando solo i documenti in una specifica lingua nel caso di un database plurilingue come Google Books e delimitando uno specifico arco cronologico di pubblicazione), e di cercare una determinata parola o espressione all'interno dei testi. Si prestano dunque egregiamente ad essere impiegati per indagini sul lessico politico del passato; la possibilità di condurre queste analisi su quantità di testi precedentemente inimmaginabili permette di reperire occorrenze ancora sconosciute di un dato termine o espressione, consentendo, in riferimento a intervalli temporali non troppo ampi, di ricostruire con un buon grado di attendibilità le 'origini' di specifici usi linguistici (Proietti e Zanettin 2021; Tuccari 2024). Se applicati ad intervalli cronologici e/o geografici più ampi, aiutano il ricercatore a verificare le fluttuazioni di significato di singoli termini nel tempo e nello spazio (Greenfield 2013; Proietti 2024) o, addirittura, a tentare di ricostruire le complesse interazioni che portano alla formazione di una nuova cultura politica (De Bolla 2013). In altri casi, l'uso dei database e dei grafici si affianca ad analisi storico-linguistiche condotte con strumenti più tradizionali (McQueen 2018).

3. Potenzialità, limiti e prospettive

Conducendo le proprie ricerche su insiemi così vasti di documenti, lo storico del pensiero politico farà l'esperienza di imbattersi in testi di cui ignorava l'esistenza e che, in qualche caso, si riveleranno fondamentali per l'esito delle ricerche stesse, molto più spesso di quanto potesse accadergli visionando documenti cartacei in biblioteche e archivi. È, questo, il principale vantaggio che l'uso di metodologie informatico-quantitative presenta rispetto alla ricerca bibliografica condotta con i tradizionali criteri biblioteconomici (selezione dei testi per autore, titolo eccetera). In particolare, la possibilità che si offre oggi agli studiosi di verificare – selezionando opportune parole-chiave da ricercare all'interno dei testi del *corpus* indagato – la presenza o meno di documenti riguardanti determinate tematiche rende del tutto obsoleta la classificazione degli stessi, nei cataloghi delle biblioteche, per "argomento" o "soggetto": una classificazione che, per quanto accurata possa essere, risulta invariabilmente viziata da *bias* (Howard e Knowlton 2018) o, perlomeno, dalla scarsità di informazioni in possesso di chi la realizza.

A fronte delle nuove possibilità sin qui descritte, l'uso di questi strumenti presenta, però, alcuni limiti di cui occorre tener conto. Tra le critiche che i detrattori di questo approccio tendono a formulare va menzionata in primo luogo – non perché appaia particolarmente fondata, ma perché viene spesso ripetuta o tacitamente ritenuta valida – quella secondo

cui l'uso sistematico dei database informatizzati comporterebbe la sostituzione di un approccio 'qualitativo' alla ricerca storica con uno meramente 'quantitativo'. Questo tipo di critica, avanzato di solito da chi non ha familiarità con l'utilizzo degli strumenti lessicometrici, si basa sull'idea – del tutto erronea – che lo scandaglio quantitativo possa rappresentare *tutta* la ricerca, o buona parte di essa. In realtà, come ben sa chi si cimenta in questo tipo di studi, ogni occorrenza di un dato termine o espressione va analizzata nel suo contesto: ossia, nell'ambito dell'argomentazione sviluppata nel testo che la contiene, e in rapporto agli altri testi con i quali il primo è storicamente correlato. La ricostruzione del contesto discorsivo delle occorrenze è un lavoro fondamentale, che proprio i database menzionati consentono di effettuare più facilmente: individuati i testi contenenti il lemma di nostro interesse, infatti, potremo allargare l'indagine al testo nella sua completezza e – in molti casi – sarà possibile scaricarlo e conservarlo in formato pdf. Ad aumentare esponenzialmente, in altri termini, è la quantità di fonti che dovranno poi essere analizzate con la consueta strumentazione critica; il che rende, semmai, il lavoro dell'interprete più lungo e complesso, e certo non, come vorrebbe una certa *vulgata*, più rapido e superficiale.

Altre obiezioni, maggiormente fondate, riguardano le modalità di formattazione dei testi; in questa sede non potremo entrare nel dettaglio dei problemi segnalati dalla critica (Hill 2017; Zanettin e Proietti 2021), ma questi possono essere ricondotti, in sintesi, a un errato riconoscimento di caratteri, sintagmi e altre sezioni del testo da parte dei *software* OCR utilizzati per digitalizzare i testi presenti in questi database. Tali errori sono all'origine, in molti casi, di falsi positivi. Per fare un esempio basato sulla mia esperienza personale (Proietti 2024), ho verificato che i termini *aristocracy* e *aristocratie* vengono spesso erroneamente identificati, soprattutto nei testi pre-ottocenteschi, come *autocracy* e *autocratie*, tanto in Google Books quanto in Gallica. Certo, si può sostenere che i falsi positivi non costituiscano un problema se non in termini di aggravio del lavoro di 'ripulitura' qualitativa che segue necessariamente, come si è detto, alla ricerca quantitativa; i falsi negativi, però, di cui si deve supporre l'esistenza, corrispondono ad altrettante occorrenze sfuggite all'analisi, e dunque inficiano, in una percentuale difficilmente quantificabile, la completezza dei risultati della ricerca. Il fatto è che la precisione nella formattazione dei testi, assicurata dal lavoro manuale delle *équipes* che digitalizzano i documenti in caso di *corpora* di dimensioni limitate e autoprodotti secondo le modalità descritte nel primo paragrafo, è del tutto impossibile da ottenere in rapporto agli immensi database di cui disponiamo oggi. In un interessante intervento che mette a confronto la qualità dei dati presenti in Google Books e in Gallica, il secondo archivio viene definito più attendibile rispetto al primo in quanto elaborato da bibliotecari in carne e ossa, e non da algoritmi (Azoulay e De Courson 2021, 4). Questa superiorità è evidente, in effetti, per quanto riguarda l'accuratezza di metadati quali autore, titolo, editore e data di edizione; è stato calcolato, ad esempio, che in Google

Books il margine di errore in relazione alla corretta datazione dei testi può arrivare al 20%, il che rappresenta un problema rilevantissimo per la ricerca storica (James e Weiss 2012). Ma per quanto riguarda gli errori di formattazione del testo dovuti a un'errata scansione OCR, i due database sono sostanzialmente equivalenti tra loro.

Un secondo problema, ancora più rilevante del precedente, è costituito dall'opacità dei database precostituiti come ECCO, Google Books e Gallica. Per essere realmente attendibile, l'analisi lessicometrica necessita la preliminare conoscenza, da parte del ricercatore, di *cosa* esattamente sia presente nel *corpus*. Ciò è molto difficile da determinare nei casi di ECCO (Tolonen, Mäkelä e Lahti 2022) e Gallica, e del tutto impossibile nel caso di Google Books, sia per ragioni strettamente correlate alle modalità di digitalizzazione del *corpus* – operazione affidata a una grande quantità di istituzioni diverse, *in primis* biblioteche, senza una supervisione centralizzata sui contenuti, con la conseguente presenza di un gran numero di duplicati – sia, ancor più, per il fatto che quello di Google Books è un *corpus* fluttuante, instabile, in cui ogni giorno è possibile reperire nuovi documenti, ma al tempo stesso capita di constatare che alcuni di quelli precedentemente indagati non sono più visibili; basta operare la stessa ricerca sulla maschera avanzata di questa piattaforma a distanza di qualche settimana per rendersi conto di tale dato di fatto.

Un ulteriore problema da segnalare, collegato al precedente, consiste nel fatto che la maggior parte delle piattaforme testuali online, essendo in mano a compagnie private, non garantiscono, oltre alla trasparenza, nemmeno la necessaria continuità nel tempo che si associa a una strumentazione di tipo scientifico. Paradossalmente, se Google e altre aziende decidessero di abbandonare questo settore non ritenendolo, ad esempio, sufficientemente remunerativo sul piano commerciale, la conseguenza sarebbe che ricerche attualmente rese possibili proprio da quegli strumenti non sarebbero più replicabili. La rivoluzione digitale, in altri termini, non va necessariamente considerata irreversibile.

Infine, un ultimo aspetto da tenere in conto è l'illusione di completezza che l'analisi di una quantità così ampia di testi porta con sé. Questo punto richiede la massima attenzione da parte degli studiosi, e la massima cautela nel presentare i risultati delle proprie ricerche. Certamente, se si lavora sulla storia di precisi lemmi o locuzioni, sarà possibile ampliare o correggere in parte le interpretazioni di chi si è interessato agli stessi temi in passato (Proietti e Zanettin 2021, 44); ma i risultati ottenuti saranno sempre suscettibili di revisione, e documenti in grado di dare una svolta all'indagine potrebbero restare, forse per sempre, relegati nella sfera del 'non digitalizzato'. Anche di questo occorre essere consapevoli.

3. Conclusioni: *linguistic turn revisited*

Reversibile o meno che sia, la svolta epistemologica che la disponibilità di immensi database imprime alle modalità di lavoro nel campo della storia del pensiero politico è tale da indurre a interrogarsi sull'integrazione di questi strumenti rispetto ai principali orientamenti metodologici degli ultimi decenni. Come ho cercato di argomentare, è soprattutto nel campo della storia delle idee orientato alla ricostruzione del lessico, dei linguaggi e del dibattito politico che lo scenario appare radicalmente modificato dai database online, in particolare rispetto al dato fondamentale della quantità di fonti esplorabili da parte di un singolo ricercatore. Si tratta, in qualche modo, di un nuovo *linguistic turn* che segue quello di qualche decennio fa. Sembra abbastanza evidente che la semantica storica, sul modello della scuola della *Begriffsgeschichte*, non possa che giovarsi delle nuove potenzialità offerte dagli archivi di testi informatizzati, pur con tutte le cautele ricordate sopra; esistono già esempi in questo senso (Ihalainen e Holmila 2022). Inevitabilmente, è più nel campo dell'analisi *semasiologica* che in quello dell'analisi *onomasiologica* (Scuccimarra 2022, 31) che questi strumenti aprono prospettive feconde; ma anche l'onomasiologia può giovarsi della possibilità, offerta dai database online, di sottoporre a verifica le mutazioni, nel tempo e nello spazio, del campo semantico dei diversi termini riconducibili a uno stesso concetto.

Per quanto riguarda la metodologia della *Cambridge School*, intesa in senso lato come approccio che enfatizza la ricostruzione delle intenzioni degli autori politici indagati in relazione ai loro contesti di appartenenza, una riflessione è già stata avviata in merito al ripensamento di quel paradigma in conseguenza dell'avvento di strumenti lessicometrici così estesi. Jennifer London ha immaginato, ad esempio, la possibilità di un fecondo rinnovamento del «circuito ermeneutico» grazie alle nuove opportunità di dialogo tra studiosi «orientati in senso ermeneutico» e studiosi «orientati in senso empirico» favorite dagli strumenti informatici di cui stiamo trattando (London 2016, 370). Certo, è ancora presto per capire fino a che punto il rinnovamento della prassi di ricerca, su cui impatta l'uso di una strumentazione in precedenza inesistente, possa trasformarsi in novità epistemologica in senso pieno, costringendoci a ripensare in profondità le categorie storiografiche che usiamo abitualmente, e fino a che punto, invece, archivi cartacei e digitali – con le relative prassi di analisi critica – continueranno, anche in futuro, a coesistere.

Bibliografia

Azoulay, Benjamin e De Courson, Benoît. 2021. "Gallicagram, un outil de lexicométrie pour la recherche". Ultimo accesso 21 marzo 2025. https://doi.org/10.31235/osf.io/84bf3.

Betti, Arianna e van den Berg, Hein. 2016. "Towards a Computational History of Ideas". In *Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age: Luxembourg, Luxembourg, December 5-6, 2013*, a cura di Lars Wierneke, Catherine Jones, Marten Düring, Florentina Armaselu, René Leboutte. Ultimo accesso 21 marzo 2025. https://ceur-ws.org/Vol-1681/

Ciotti, Fabio (a cura di). 2023. *Digital Humanities. Metodi, strumenti, saperi*. Roma: Carocci.

De Bolla, Peter. 2013. *The Architecture of Concepts: The Historical Formation of Human Rights*. New York: Fordham University Press.

Evrigenis, Ioannis D. 2015. "Digital Tools and the History of Political Thought: The Case of Jean Bodin". *Redescriptions: Political Thought, Conceptual History and Feminist Theory* 18/2: 181-201.

Gedzelman, Séverine e Zancarini, Jean-Claude. 2011. "HyperMachiavel. Un outil de comparaison de traductions". *Lingua e stile* 2011/2: 247-266.

Greenfield, Patricia M. 2013. "The Changing Psychology of Culture From 1800 Through 2000". *Psychological Science* 24/9: 1722-1731.

Gregg, Stephen H. 2020. *Old Books and Digital Publishing: Eighteenth-Century Collections Online*. Cambridge: Cambridge University Press.

Hill, Mark J. 2016. "Invisible Interpretations: Reflections on the Digital Humanities and Intellectual History". *Global Intellectual History* 2016/2: 130-150.

Howard, Sara A. e Knowlton, Steven A. 2018. "Browsing through Bias: The Library of Congress Classification and Subject Headings for African American Studies and LGBTQIA Studies". *Library Trends* 67/1: 74-88.

Ihalainen, Pasi e Holmila, Antero (a cura di). 2022. *Nationalism and Internationalism Intertwined: A European History of Concepts Beyond the Nation State*. New York: Berghahn Books.

James, Ryan e Weiss, Andrew. 2012. "An Assessment of Google Books' Metadata". *Journal of Library Metadata* 12/1: 15-22.

London, Jennifer A. 2016. "Re-imagining the Cambridge School in the Age of Digital Humanities". *Annual Review of Political Science* 19: 351-373.

McQueen, Alison. 2018. "Mosaic Leviathan: religion and rhetoric in the philosophy of Hobbes". In *Hobbes on Politics and Religion*, a cura di Laurens van Apeldoorn e Robin Douglass, 116-134. Oxford: Oxford University Press.

Proietti, Fausto. 2024. "*Autocratie*: histoire politique d'un mot du XVII^{e} au XX^{e} siècle". *The Tocqueville Review/La Revue Tocqueville* 45/1: 127-146.

Proietti, Fausto e Zanettin, Federico. 2021. "«Democrazia rappresentativa» Indagine sulle origini di una categoria politica (1778-1799)". *Storia del pensiero politico* 1/2021: 41-66.

Scuccimarra, Luca. 2022. "Semantiche della temporalità e conoscenza storica: il contributo della *Begriffsgeschichte*". *Quaderno di storia del penale e della giustizia* 4: 29-50.

Tolonen, Mikko, Mäkelä, Eetu e Lahti, Leo. 2022. "The Anatomy of Eighteenth Century Collections Online (ECCO)". *Eighteenth-Century Studies* 56/1: 95-123.

Tuccari, Francesco. 2024. "La «democrazia diretta» nel pensiero politico del XIX secolo". *Il pensiero politico* 57: 53-84.

Zanettin, Federico e Proietti, Fausto. 2021. "L'uso di risorse online per lo studio sincronico e diacronico del lessico politico. Il caso di 'democrazia rappresentativa'". *Umanistica digitale* 10: 165-192.

Fausto PROIETTI is full professor of History of Political Thought at the University of Perugia. The main focus of his research consists in the historical reconstruction of the ideological debate that accompanied the birth of representative democracy in Europe, as well as the discussion on "radical" alternatives to that model (direct popular legislation, federative self-government on a municipal basis). He has published several essays on the European political exile during the revolutions of 1848, on the European political lexicon in the 19th century and on the epistemological dimension of the history of political thought.

Email: fausto.proietti@unipg.it