

## L'uso di risorse online per lo studio sincronico e diacronico del lessico politico. Il caso di "democrazia rappresentativa"

<sup>1</sup>Federico Zanettin, <sup>2</sup>Fausto Proietti

<sup>1,2</sup>Università di Perugia, Italia

<sup>1</sup>[federico.zanettin@unipg.it](mailto:federico.zanettin@unipg.it)

<sup>2</sup>[fausto.proietti@unipg.it](mailto:fausto.proietti@unipg.it)

### Abstract

In questo articolo forniamo una panoramica delle risorse documentali reperibili online e utilizzabili per la ricerca sulla storia dei concetti politici, e discutiamo alcune "buone pratiche" volte a ottimizzare l'uso di collezioni di testi digitali nel pubblico dominio nella ricerca storiografica. Il caso di studio utilizzato per illustrare le nostre considerazioni riguarda la ricerca delle occorrenze, in ampie banche dati, archivi e biblioteche digitali, del sintagma "democrazia rappresentativa" (e dei suoi equivalenti in inglese e francese), tra il 1778, primo anno in cui se ne registra l'occorrenza nei nostri dati, e il 1799. La maggior parte delle occorrenze rinvenute tramite un attento processo di selezione e controllo non erano ancora state portate all'attenzione degli studiosi. Nella parte successiva dell'articolo prendiamo in considerazione il contributo che possono apportare alla ricerca sul lessico politico in una prospettiva diacronica strumenti di analisi testuale come *Google Books Ngram Viewer*, *HathiTrust + Bookworm* e *Sketch Engine*. A questo scopo discutiamo dati relativi alla frequenza e al profilo lessicale di "democrazia" e "democrazia rappresentativa" in corpora storici e contemporanei.

In this article we wish to provide a survey of online documentary resources for the historical study of political thought and illustrate a series of "good practices" which could be adopted to optimize the use of publicly available digital collections in historiographical research. As a case study, we first report on the search, in very large digital collections of historical documents, for occurrences of the phrase "representative democracy" (and its dictionary equivalents in Italian and French) between 1778, the first occurrence of the term recorded in our data, and 1799. Most of the occurrences retrieved through a careful process of selection and scrutiny have not previously been discussed in the literature. In the last part of the article we discuss the contribution of text analysis tools to diachronic research, looking at frequency data from resources such as *Google Books Ngram Viewer* and *HathiTrust + Bookworm*, and comparing findings about the lexical profiles of "democracy" and "representative democracy" in historical and contemporary corpora.

## Testi digitali e ricerca storica

Uno dei fenomeni che caratterizzano l'era digitale è la cosiddetta “rimediazione” ([4]), cioè la trasposizione/traduzione di oggetti culturali in un medium diverso da quello in cui originariamente erano stati pensati. In particolare, stanno assumendo un peso sempre maggiore importanti progetti che prevedono la trasposizione in formato digitale del patrimonio dei testi a stampa pubblicati negli ultimi cinque secoli. La possibilità concreta di avere accesso immediato a un enorme numero di testi, e in prospettiva all'intero patrimonio librario del passato, inevitabilmente influenza il lavoro dello storico, al quale fino a pochi anni fa quelle risorse non erano accessibili con la stessa facilità e immediatezza.

Questo articolo si propone di indagare le opportunità e i limiti offerti da strumenti, risorse e metodologie elaborati negli ambiti della linguistica computazionale, dell'informatica umanistica e della linguistica dei corpora ai fini della ricerca in ambito storico, e in particolare per quanto riguarda la storia delle idee e del lessico politico. A tale scopo si prenderà in considerazione come caso di studio il sintagma “democrazia rappresentativa” e i suoi equivalenti comunemente forniti in dizionari ed enciclopedie per le lingue inglese e francese, per indagare l'utilità delle risorse disponibili e le “buone pratiche” metodologiche implementabili per ottimizzare la ricerca in ambito storiografico. La scelta di prendere in esame questo termine nasce dalla constatazione di un paradosso: nonostante lo sviluppo, negli ultimi decenni, di una letteratura piuttosto ampia dedicata all'analisi storica del lessico politico,<sup>1</sup> e benché i contributi sulla democrazia e la sua “storia” abbiano assunto ormai una dimensione addirittura fluviale, sono assai poche a tutt'oggi le ricerche specificamente dedicate all'indagine delle origini storiche di un sintagma ideologicamente rilevante come “democrazia rappresentativa” (si vedano però [30], [18], [39] e [15]). Alcuni studi, come [33], [12], [17] e [14], si sono concentrati su specifici termini politici, mentre altri hanno dedicato un'attenzione particolare alla parola “democrazia” ([8];[13];[22];[32]), senza, però, soffermarsi sulla locuzione che qui ci interessa, trattata implicitamente come un semplice sottoinsieme della più ampia categoria semantica individuata dal sostantivo. Gli studiosi della democrazia, dal canto loro, hanno quasi sempre trascurato la dimensione semantica del problema, andando alla ricerca della 'cosa' e non del nome: secondo questa prospettiva, sono stati classificati come fautori della “democrazia rappresentativa” tutti quegli autori che, nelle più varie epoche e nei più diversi contesti linguistici, hanno proposto teorie articolate su uno o più degli elementi (ad esempio la rappresentanza, la divisione dei poteri o l'esistenza di un controllo da parte dei governati sull'azione del governo) identificati (dagli interpreti) come consustanziali a ciò che oggi chiamiamo democrazia rappresentativa. Le ricerche di questo tipo, in altre parole, normalmente partono da un dato – la democrazia rappresentativa come oggi la conosciamo, nel caso specifico – andandone a cercare nel passato le 'origini' concettuali; così facendo, mettono la verifica storica al servizio di una prospettiva che è, di fatto,

---

1 All'interno dell'ampio dibattito metodologico che ha sottolineato la rilevanza degli elementi linguistici, semantici e discorsivi per la ricostruzione storica delle idee politiche si vedano in particolare [34]; [40];[19];[47].

normativa.<sup>2</sup> La prospettiva da noi adottata è sostanzialmente speculare rispetto a quella appena richiamata. Il nostro scopo è infatti quello di verificare come le risorse, gli strumenti e le metodologie digitali contribuiscano ad indagare il contesto storico in cui un sintagma oggi comune come "democrazia rappresentativa" è stato introdotto – più o meno in contemporanea – in alcune delle principali lingue europee, e vedere in opposizione a quali altri termini e locuzioni sia stato dibattuto, anche in una prospettiva di lungo termine che pone a confronto le costellazioni terminologiche e concettuali con cui si è relazionato il sintagma "democrazia rappresentativa" in prospettiva diacronica, dalla fase del suo emergere come neologismo fino ai giorni nostri. Anche solo l'affermazione duratura di questa locuzione indica indubbiamente un fenomeno al quale vale la pena dedicare, in sede storica, seria attenzione. I risultati dell'indagine qualitativa che abbiamo condotto in ottica sincronica su questo tema in relazione alla fase settecentesca del dibattito politico sono stati pubblicati in altra sede ([36]); in questo saggio intendiamo mettere in luce alcune problematiche di carattere metodologico emerse durante la ricerca, e soffermarci sulle analisi di tipo diacronico.

Due ambiti di ricerca sulla storia del lessico politico sembrano in particolare poter beneficiare delle risorse disponibili online, che permettono l'analisi di una quantità di testi impensabile fino a non molti anni fa, e altrimenti spesso difficilmente accessibili. Un primo ambito è quello della ricerca dei dati, l'equivalente digitale della ricerca e consultazione dei testi presso biblioteche sparse su un'ampia area geografica, che però, rispetto ai tradizionali strumenti biblioteconomici, permette di individuare i testi rilevanti ai fini dell'indagine effettuando non soltanto le classiche ricerche nei campi 'autore', 'titolo' e 'soggetto' (campo questo in cui i contenuti sono fortemente condizionati dalla soggettività di chi ha materialmente realizzato la catalogazione), ma anche attraverso la ricerca di determinate espressioni linguistiche contenute nel testo. Le ricerche condotte sulle banche dati testuali necessitano di precise cautele e attenzione nella sintassi di ricerca, che può variare e raggiungere diversi livelli di raffinatezza a seconda della risorsa utilizzata. I risultati delle ricerche vanno poi naturalmente verificati esaminando i singoli testi, prima di poter procedere all'analisi e all'interpretazione dei contesti discorsivi. Un secondo ambito riguarda lo studio della nascita e dell'evoluzione di uno o più termini lungo un determinato arco di tempo. Tali studi sono resi possibili dall'analisi statistica delle occorrenze in banche dati testuali di grandi dimensioni, per mezzo di specifici strumenti di analisi lessicometrica che permettono di tracciare il variare delle frequenze d'uso di una parola o espressione nel corso dei secoli. Si tratta di un comparto di ricerca denominato "culturomics",<sup>3</sup> che ha lo scopo di investigare come determinati fenomeni culturali si riflettono nell'uso linguistico. Li et al. ([24]), ad esempio, hanno esaminato la frequenza della parola *risk* e delle sue co-occorrenze durante un periodo di 150 anni, per verificare i cambiamenti semantici del discorso sul rischio, mentre

---

2 Tra i più noti – e riusciti – esempi di questo approccio si colloca uno studio come quello di Nadia Urbinati, che individua preliminarmente i "principi" della democrazia rappresentativa, andandone poi a cercare la "genealogia" all'interno del pensiero di autori che, con l'eccezione di Condorcet, non hanno mai utilizzato la formula verbale corrispondente ([46]).

3 Il termine è stato introdotto in riferimento al tipo di analisi che è possibile svolgere a partire dal progetto di lessicologica computazionale *Google Books Ngram Corpus* ([29]) (vedi sezione 3).

Younes & Reips ([46]) hanno indagato l'uso di parole che si riferiscono alla religiosità nel XX secolo, per constatare il graduale declino nell'uso di tale terminologia con occasionali crescite di importanza in periodi di crisi come la seconda guerra mondiale. Infine, l'opportunità di mettere a confronto corpora di lingua contemporanea con corpora storici in prospettiva diacronica offre alla ricerca storica la possibilità di trarre vantaggio da metodologie e risorse principalmente utilizzate in ambito della linguistica dei corpora, della lessicografia e della terminologia.

### **Descrizione e uso delle risorse testuali**

Nella seconda sezione di questo articolo presentiamo dapprima una rassegna degli strumenti e delle risorse disponibili online che permettono l'accesso a testi in pubblico dominio. Poi illustriamo brevemente le modalità e i risultati di una ricerca riguardante l'uso del termine "democrazia rappresentativa" nell'arco di tempo che va dal suo primo apparire nei dati da noi verificati, il 1778 (anno in cui il sintagma compare per la prima volta in un testo stampato a Londra in lingua francese ([38]), fino al 1799, data che rappresenta un'importante cesura storica non solo perché segna la fine del XVIII secolo ma anche per le ripercussioni a livello internazionale del colpo di Stato napoleonico del 9 novembre 1799.

#### ***Banche dati testuali e archivi digitali online***

La principale risorsa per quanto riguarda la ricerca sui testi digitali nel pubblico dominio è costituita dal progetto *Google Books*, iniziato nel 2005 dall'omonima azienda proprietaria del noto motore di ricerca con il proposito di catalogare e indicizzare l'intero patrimonio librario mondiale. Utilizzando un'interfaccia simile a quella del motore di ricerca generale dell'azienda, qualsiasi utente ha accesso ad un database testuale contenente i testi di decine di milioni di libri. La consistenza del corpus, quantificata nel 2009 in 12 milioni di volumi – stimati corrispondere a circa il 15% di tutti i libri prodotti fino ad allora ([29]: 177) – ha raggiunto nel 2019 i 40 milioni ([23]).<sup>4</sup> I testi contenuti in questo database rappresentano indubbiamente il più ampio corpus di libri digitalizzati a livello internazionale e costituiscono una considerevole frazione di tutti i libri mai pubblicati.

Il progetto *Google Books* è stato accolto in modo critico sotto tre aspetti: controversie sul diritto d'autore ([43]: 48-52), dibattito sulla privatizzazione delle conoscenze<sup>5</sup> e dubbi relativi alla qualità del prodotto. Per quanto riguarda in particolare quest'ultimo aspetto, diversi studiosi hanno criticato l'utilità del progetto ai fini della ricerca accademica, lamentando una scarsa

---

4 Il database contiene attualmente testi in 400 lingue. Non sembrano essere però disponibili dati relativi alla distribuzione quantitativa dei testi per lingua e per anno di pubblicazione.

5 Ad esempio, il progetto *Gallica* e il successivo progetto *Europeana* (vedi oltre), il primo istituito e il secondo promosso con forza dalle istituzioni pubbliche francesi, sono spesso presentati come una risposta alla privatizzazione dalla memoria culturale da parte di *Google Books*, e alla colonizzazione americana della cultura europea.

qualità dei dati per quanto riguarda sia l'accuratezza dei testi digitalizzati<sup>6</sup> che l'affidabilità dei metadati associati ai testi stessi. In riferimento alla prima versione di *Google Books* (rilasciata nel 2009) Nunberg ([31]) sosteneva che il problema fosse così esteso, in particolare per quanto riguarda l'affidabilità dei metadati, da costituire un "disastro" per la ricerca. Altre critiche riguardano la presenza sproporzionata di testi scientifici, soprattutto nei decenni più recenti, e la possibilità che determinati autori influenzino in modo troppo pesante i dati con l'utilizzo di specifiche parole o sintagmi ([46]: 2). *Google Books*, di cui è di recente uscita una nuova versione, sostiene di avere aumentato l'accuratezza dei testi, sia per quanto riguarda ad esempio la standardizzazione dei caratteri in testi storici (vedi oltre), sia per quanto riguarda quella dei dati bibliografici. Pur presentando un margine di errore le cui dimensioni non sembrano essere stimabili, *Google Books* rappresenta indubbiamente un enorme contributo alla digitalizzazione delle conoscenze e un'utile risorsa nell'ambito della nostra ricerca.

Accanto a *Google Books*, altre grandi banche dati testuali e archivi digitali di testi storici realizzati da enti bibliotecari pubblici e privati hanno messo a disposizione dei ricercatori il patrimonio librario di pubblico dominio. Per quanto riguarda gli Stati Uniti, il consorzio di biblioteche accademiche *HathiTrust*, fondato nel 2008, fornisce accesso a un patrimonio di circa 8,5 milioni di titoli,<sup>7</sup> mentre *Gallica*, un database sviluppato dalla Bibliothèque Nationale de France a partire dal 1997, contiene oltre un milione di titoli pubblicati tra il 1751 e il 2000.<sup>8</sup> Il contenuto di *Google Books* si sovrappone in parte ai contenuti di altri archivi digitali (compresa buona parte dei contenuti di *HathiTrust*) in quanto questi ultimi hanno aderito al progetto della multinazionale americana.<sup>9</sup> La British Library, ad esempio, ha in corso un progetto di digitalizzazione di circa 250.000 libri pubblicati tra il 1700 e il 1870 ([38]), mentre circa un milione di titoli del patrimonio presente nelle biblioteche nazionali di Roma e Firenze sono gradualmente inseriti nel database di *Google Books* a seguito di un accordo, risalente al 2010.<sup>10</sup>

---

6 Thylstrup nota come il compromesso tra accuratezza e accessibilità rappresenti una delle questioni più dibattute: "one of the central points of contestation in mass digitization [is] the trade-off between accuracy and accessibility" ([43]: 10).

7 Cfr. <https://www.hathitrust.org/about> per aggiornamenti relativi al numero di testi digitalizzati presenti in HathiTrust.

8 Cfr. <https://gallica.bnf.fr/GallicaEnChiffres> per aggiornamenti relativi al numero di testi digitalizzati presenti in Gallica.

9 Per un elenco, aggiornato al 2011, delle biblioteche partners del progetto cfr. <https://books.google.it/intl/it/googlebooks/partners.html>

10 Non è dato sapere a che punto si trovi attualmente il processo di digitalizzazione, iniziato nel dicembre 2012 (<https://www.beniculturali.it/comunicato/digitalizzazione-del-patrimonio-librario-parte-il-progetto-mibac-google-1>). In assenza di database nazionali paragonabili a *HathiTrust* o *Gallica*, *Google Books* rappresenta lo strumento d'elezione per quanto riguarda i testi in lingua italiana. Alcuni testi utili alla nostra ricerca, come le traduzioni italiane degli articoli di Mallet du Pan nella rivista *Mercurio Britannico*, sono accessibili online presso la Teca digitale della Biblioteca Nazionale Centrale di Roma (<http://digitale.bnc.roma.sbn.it/tecadigitale/>). Questi documenti non sono, però, ricercabili a testo libero.

Dato che la BnF non aderisce al progetto di *Google Books*, la sovrapposizione tra *Gallica* e quest'ultimo è minore, e si limita sostanzialmente a testi conservati in altre biblioteche francesi,<sup>11</sup> mentre *Google Books* contiene anche testi depositati presso biblioteche in altre aree francofone. Nella nostra ricerca si è visto, ad esempio, che una parte notevole delle occorrenze di “*démocratie représentative*” provenivano da testi stampati in Svizzera. Il database di *Gallica* contiene una frazione considerevole del patrimonio bibliotecario francese; il periodo quantitativamente più rappresentato nel corpus è quello che va dal 1851 al 1900, mentre il cinquantennio di nostro interesse (1751-1800) contiene un numero decisamente minore di testi, rappresentativi però, verosimilmente, di una percentuale maggiore sul totale dei testi editi in quel periodo.

Per quanto riguarda il patrimonio di libri storici in lingua inglese, una notevole risorsa è costituita dai documenti digitali conservati nell'*Oxford Text Archive* della Bodleian Library dell'università di Oxford. A differenza di *Google Books* e delle collezioni presenti nelle biblioteche digitali, che permettono di visualizzare i documenti scannerizzati e di ricercare e scaricare i documenti in formato PDF e spesso anche in formato solo testo, i testi contenuti nell'*Oxford Text Archive* sono codificati secondo gli standard TEI XML e sono scaricabili individualmente in vari formati, ma non sono ricercabili a testo libero. Un corpus contenente le collezioni *Early English Books Online (EEBO) – Phase I* (libri stampati tra il 1473 e il 1700), *ECCO - Eighteenth Century Collections Online* (libri stampati tra il 1701 e il 1800) e *Early American Imprints, Series I: Evans* (libri stampati in America tra il 1640 e il 1821), per un totale complessivo di circa 33.000 testi (dattiloscritti piuttosto che scannerizzati), è però consultabile sulla piattaforma di servizi lessicografici *Sketch Engine*<sup>12</sup> (vedi oltre).

Negli ultimi tempi si registra lo sviluppo di progetti relativi alla digitalizzazione di un altro tipo di documenti storici, ovvero le pubblicazioni periodiche. Il progetto *Chronicling America* della Library of Congress di Washington fornisce l'accesso a un corpus di giornali storici americani contenente oltre 2,5 milioni di copie da oltre 3.400 diversi quotidiani pubblicati negli Stati Uniti tra il 1777 e il 1963. *The British Newspaper Archive* è un progetto che prevede la digitalizzazione della collezione di quotidiani della British Library, il cui corpo centrale è costituito da giornali pubblicati principalmente nel XIX secolo, ma contenente anche documenti pubblicati a partire dal 1700 e fino al 1950.<sup>13</sup> Alla fine del 2020 il corpus consisteva in oltre 40 milioni di pagine.<sup>14</sup> Un'altra importante risorsa è costituita dalla collezione di quotidiani storici *Retronews*, creata dalla Bibliothèque Nationale de France e contenente la digitalizzazione di circa 1000 periodici

---

11 Ad esempio la biblioteca comunale di Lione, ma anche altre biblioteche europee e americane che conservano libri in lingua francese.

12 L'accesso a *Sketch Engine* ([21]) è fornito gratuitamente per scopi non commerciali alle istituzioni accademiche europee nell'ambito del Progetto ELEXIS fino all'aprile 2022.

13 A differenza di *Chronicling America* e *RetroNews*, il corpus di quotidiani britannici non contiene nessuna occorrenza dei termini cercati per il periodo considerato.

14 Cfr. <https://www.britishnewspaperarchive.co.uk/>. Si può accedere alle singole pagine in cui compare il termine o locuzione cercati e scaricare il PDF della pagina (non ricercabile) oppure fare una copia e incolla di tutto il testo.

pubblicati tra il 1631 e il 1950. L'accesso a *Chronicling America* è gratuito, mentre *The British Newspaper Archive* e *Retroneus* sono accessibili dietro pagamento di una sottoscrizione mensile.

A conclusione di questa rassegna, citiamo alcune importanti banche dati e archivi digitali accessibili online che contengono pubblicazioni nel pubblico dominio, ovvero il *Project Gutenberg*, l'*Internet Archive*, il *Corpus of Historical American English* ed *Europeana*. Queste iniziative non sono risultate utili alla nostra ricerca per diversi motivi: il *Project Gutenberg*, progetto precursore iniziato nel 1971 all'alba di Internet, attualmente ospita oltre 60.000 testi, la maggior parte dei quali in formato solo testo e relativi alla lingua inglese. I testi contenuti in *Project Gutenberg* non contengono però metadati ([41]) e non è possibile limitare la ricerca a determinati periodi. L'*Internet Archive*, un progetto no-profit iniziato nel 1996 con lo scopo di conservare siti e altri materiali pubblicati su Internet, dal 2005 ha iniziato a digitalizzare libri e altri documenti. All'inizio del 2021 il sito conteneva oltre 5 milioni di testi, di cui solo una piccola parte (apparentemente circa 130.000 documenti) risalenti a prima del XIX secolo. Le ricerche nell'*Internet Archive* non hanno prodotto dati utili alla nostra ricerca. Il *Corpus of Historical American English* (COHA), creato da Mark Davies ([10]) ed accessibile come parte delle risorse linguistiche messe a disposizione dallo studioso, contiene circa 107.000 testi (oltre 400 milioni di parole) in inglese americano pubblicati tra il 1810 e il 2009, suddivisi per fasce decennali e in 4 componenti, fiction (circa metà del corpus), riviste e quotidiani (a partire dal 1860), e libri di non fiction.<sup>15</sup> Il corpus, interrogabile per fasce temporali e separatamente per ciascuna componente, è annotato linguisticamente e permette elaborate ricerche, anche se nel caso della nostra indagine non è stato utile per via dell'arco cronologico preso in esame. Infine, il progetto *Europeana*, lanciato nel 2008 dall'Unione Europea con il proposito di costruire un'infrastruttura di rete che colleghi il patrimonio culturale conservato nelle biblioteche nazionali europee. Il progetto si propone di rendere accessibili attraverso un'unica interfaccia contenuti digitali di oltre 1000 istituzioni europee; tuttavia, l'interfaccia di ricerca di *Europeana* non permette la ricerca a testo libero, in quanto si basa su un diverso approccio culturale che privilegia l'esplorazione delle collezioni ospitate per temi e argomenti.<sup>16</sup>

### ***Ricerca e analisi dei risultati***

Le banche dati e gli archivi digitali online contengono riproduzioni digitali di diversi oggetti, testi, suoni e immagini, e l'esplorazione di una biblioteca digitale può essere affascinante e complessa come quella di una biblioteca fisica. Per quanto riguarda la ricerca nei testi storici, soprattutto se riguardante diverse lingue, il ricercatore si deve confrontare con le opportunità e restrizioni offerte dalle diverse risorse consultate, che variano per quanto riguarda l'interfaccia utilizzata e le modalità di visualizzazione e accesso ai dati.

---

15 Una versione "ripulita" da errori è stata recentemente resa disponibile ([3]). Per l'elenco completo dei testi (in buona parte presi da fonti online come *Project Gutenberg*) vedi <https://www.english-corpora.org/coha/help/texts.asp>

16 Per una presentazione delle diverse problematiche legate a *Europeana* cfr. [43]: 57-77.

Non è chiaro quanti dei testi pubblicati fino alla fine del XVIII secolo siano contenuti nei database e negli archivi digitali. Secondo una stima<sup>17</sup> i titoli pubblicati in lingua inglese tra il 1473 e il 1800 ammontano a circa 480.000, il 7% circa dei quali contenuti nelle tre *English Historical Book Collections* disponibili presso l'*Oxford Text Archive*. È evidente perciò che molti testi contenenti occorrenze di “representative democracy”, anche importanti dal punto di vista storiografico, potrebbero non essere presenti in nessuno dei database analizzati, oppure non essere recuperabili in base ai criteri temporali utilizzati.<sup>18</sup> In secondo luogo, è inevitabile che le copie digitali non restituiscano tutte le occorrenze contenute negli originali cartacei, in quanto il processo di acquisizione automatizzato dei testi comporta problemi di riconoscimento dei caratteri (ad esempio la “s” in parole come “rappresentativa”, che nel Settecento è spesso visualizzata come “P”, parole ed espressioni che vanno a capo di riga o di pagina, spazi tra le parole, ecc.) e problemi di indicizzazione che non sempre vengono risolti. È possibile, ad esempio, che una ricerca di “representative democracy” (limitatamente a un determinato arco di tempo) produca risultati diversi se effettuata su copie digitalizzate dello stesso documento provenienti però da fonti diverse. Inoltre, è inevitabile che una certa percentuale di dati sia corrotta cioè, ad esempio, che non tutte le occorrenze del termine “democrazia rappresentativa” presenti nei testi originariamente riprodotti in formato immagine e/o testuale dalle copie depositate nelle diverse biblioteche e musei siano richiamate in queste copie digitali. Il fatto che la maggior parte delle numerose occorrenze infine rinvenute non fossero mai state richiamate all’attenzione della ricerca storica conferma però sicuramente la validità dell’indagine.

Sia le biblioteche digitali pubbliche come *HathiTrust* e *Gallica* che *Google Books* permettono di ricercare parole ed espressioni a testo libero, di accedere ai singoli risultati di una ricerca (in molti casi consentendo di visualizzare direttamente le singole occorrenze di un’espressione all’interno di un testo) e di scaricare il testo in formato PDF ricercabile (e in molti casi anche in formato testo). Essendo state progettate secondo criteri più specificamente biblioteconomici rispetto a *Google Books*, *HathiTrust* e *Gallica* permettono ricerche più puntuali e un migliore accesso ai testi e ai metadati (consentendo ad esempio di scaricare i dati aggregati di una ricerca) rispetto a *Google Books*, più ‘fluttuante’ e opaco in termini di accuratezza e di verifica dell’esatta consistenza del corpus. Il corpus di testi delle *English Historical Book Collections* dell’*Oxford Text Archive* disponibile su *Sketch Engine* è invece annotato per lemma e *part-of-speech* e l’interfaccia del software di interrogazione permette sofisticate ricerche. L’espressione di ricerca evidenziata in Figura 1, ad esempio, è stata utilizzata per individuare tutte le occorrenze dell’espressione “representative democracy”. La sintassi di ricerca specifica tutte le occorrenze di “representative” iniziati con qualsiasi carattere (allo scopo di comprendere sia maiuscole che minuscole), che comprendano una “s” o qualsiasi altro carattere (allo scopo di includere anche parole contenenti

---

17 Si tratta di una stima basata sullo Short title catalogue della British Library, cfr. [http://estc.bl.uk/F/?func=file&file\\_name=login-bl-estc](http://estc.bl.uk/F/?func=file&file_name=login-bl-estc)

18 Ad esempio, un importante testo di Condorcet del 1790 è reperibile online solo in un’edizione successiva ([9]).



la "s" settecentesca) seguite da tutte le forme del lemma "democracy" (cioè in forma singolare e plurale, contenenti lettere maiuscole o minuscole).



Figura 1: Ricerca avanzata con *Sketch Engine*.

Attraverso le ricerche condotte sulle diverse risorse abbiamo infine individuato occorrenze del sintagma "democrazia rappresentativa" e equivalenti nelle altre due lingue, per l'arco cronologico 1778-1799, in 43 testi in lingua inglese, in 89 in lingua francese e in 16 in lingua italiana.<sup>19</sup>

A questo risultato si è arrivati attraverso un considerevole processo di controllo e scrematura dei risultati delle ricerche, che hanno riguardato tutte le forme singolari e plurali del termine nelle tre lingue, tenendo conto per quanto possibile delle diverse grafie. Inoltre, si è operata anche la ricerca di altri termini ("democrazia diretta", "repubblica rappresentativa" ecc.), suggeriti sia dalla frequentazione e familiarità con i testi dell'epoca sia da evidenze lessicometriche (vedi sezione successiva). Dato che, a parte per i corpora storici contenuti in *Sketch Engine*, i software di ricerca non permettevano ricerche con espressioni regolari, le stringhe di ricerca utilizzate nelle interfacce delle diverse risorse hanno generato notevoli sovrapposizioni in termini di risultati. Ad esempio, occorrenze della forma singolare e plurale possono essere contenute nello stesso testo. Inoltre, come si è detto i contenuti di database e archivi possono sovrapporsi, per cui la stessa copia digitale può essere contenuta in vari database, con uguale o diversa classificazione. La presenza di duplicati si verifica non solo quando lo stesso libro, proveniente dalla stessa biblioteca, è accessibile da due interfacce diverse, ma anche quando due interfacce diverse (o anche la stessa interfaccia) rimandano a copie digitali dello stesso testo a stampa provenienti da biblioteche diverse. Per quanto riguarda le ricerche condotte su *Google Books*, in particolare, va segnalato il fatto che, in molti casi, uno stesso testo è presente più volte; non solo in diverse, e a volte parecchie, edizioni, ma anche, in qualche caso, con più copie dello stesso volume digitalizzate da istituzioni diverse. Si sono verificati, inoltre, alcuni casi in cui un documento reperito attraverso

<sup>19</sup> A questi ultimi vanno aggiunti altri 10 testi già segnalati da Guerci ([18]: 177-222) e non reperibili online.

la ricerca del sintagma nel testo si rivelava, a una più attenta analisi, non contenere l'espressione cercata; in altri casi non è stato possibile verificare se il documento la contenesse effettivamente. Nella nostra ricerca abbiamo incluso nelle analisi successive solo le occorrenze per le quali è stato possibile reperire il testo completo che le contiene e verificarne il contesto ideologico, escludendo ad esempio quelle per le quali non era possibile la visualizzazione o era possibile solo una visualizzazione di frammenti (*Google Books snippet view*).

Invece, pur non approfondendole sotto il profilo qualitativo, non abbiamo considerato come falsi positivi le citazioni letterali all'interno di altri contesti testuali di documenti già presi in considerazione (l'esempio più tipico è quello delle recensioni di volumi pubblicate all'interno di periodici), così come le nuove edizioni di testi già analizzati apparse nell'arco cronologico indagato, dal momento che si tratta, in entrambi i casi, di testimonianze significative della circolazione del sintagma.

La consultazione di una così ampia quantità di testi in diverse lingue ci ha permesso di gettare luce sui diversi significati attribuiti al termine “democrazia rappresentativa” dagli attori sociali dell'epoca in diversi contesti geografici e culturali. Si è evidenziato, ad esempio, come all'espressione “representative democracy” nel contesto americano fosse spesso preferita la più neutra “republic”, mentre in quello britannico essa fosse utilizzata principalmente da radicali filogiacobini e/o in riferimento al contesto rivoluzionario francese. “*Démocratie représentative*” fu usata in modo sporadico, prima di diventare notevolmente diffusa nel periodo del Direttorio (1795-1799), in Francia come in Svizzera; in Italia, il sintagma “democrazia rappresentativa” era utilizzato con una certa frequenza durante il triennio rivoluzionario 1797-1799, con accezione tanto apprezzativa quanto derogatoria (cfr. [36]). È ipotizzabile, quindi, che l'aggettivazione possa aver consentito la riappropriazione in chiave moderna di un termine – democrazia – identificato con una forma di governo antica e ritenuta inadatta a Stati di grande estensione territoriale.<sup>20</sup> Inoltre, si è verificato come, perlomeno in questa fase iniziale, il termine “democrazia rappresentativa” (o meglio, i suoi equivalenti in inglese e francese) fossero in competizione con altri termini che non attecchirono, come “*démocratie représentée*”, o scompariranno dall'uso nel corso degli anni, come “representative republic” e come si riscontrassero, prima della sua affermazione, espressioni alternative come “democracy by representation” o “*démocratie par représentation*”. È interessante infine notare come le traduzioni di alcuni testi, in particolare dall'inglese al francese ([1];[2]), e viceversa ([27];[28];[6];[7]),<sup>21</sup> possano aver contribuito all'affermazione del termine “democrazia rappresentativa”.

---

20 La storiografia ha da tempo sottolineato come al sostantivo “democrazia” fossero universalmente associati significati derogatori fino all'ultimo scorcio del XVIII secolo; per una recente focalizzazione del problema e per una sintesi bibliografica cfr. [35]: 21-36.

21 Inoltre, il *Mercurio Britannique* di Jacques Mallet Du Pan venne interamente pubblicato anche in traduzione italiana.

## **Analisi lessicometriche**

Un secondo approccio allo studio del lessico politico in una prospettiva storica a partire da corpora costituiti dalle raccolte di testi contenuti in database e archivi testuali deriva dall'elaborazione statistica dei dati quantitativi in prospettiva diacronica. A tale scopo sono determinanti i dati sulla frequenza d'uso (relativa) di una o più parole ed espressioni in diversi momenti storici e le variazioni di tale frequenza nel corso del tempo. Ad esempio, per quanto riguarda la nostra ricerca, abbiamo verificato che l'espressione "democrazia rappresentativa" conosce due momenti principali di istituzionalizzazione, il primo nell'ultimo scorcio del XVIII secolo, il secondo nella seconda metà del XIX secolo, trasformandosi da neologismo in categoria politica a pieno titolo, che designa in modo tecnicamente preciso una forma di governo funzionante.

In questa sezione prenderemo dunque in esame la caratterizzazione dell'espressione "democrazia rappresentativa" in un'ottica diacronica e contrastiva, analizzando dapprima l'evoluzione del termine a confronto con altri termini correlati, e successivamente le sue caratteristiche lessicografiche a confronto con quelle derivabili da testi contemporanei.

### *Democrazia rappresentativa nel corso del tempo*

Il computo delle frequenze relative delle parole rapportate alle frequenze assolute può rivelare importanti informazioni se relativo a un grande numero di testi e lungo un ampio arco cronologico. Uno strumento che permette analisi di questo tipo è *HathiTrust + Bookworm*, che permette di visualizzare le frequenze relative lungo l'asse del tempo di una o più parole all'interno del corpus di *HathiTrust*.<sup>22</sup> Il grafico in Figura 2 rappresenta l'andamento dell'uso delle parole "republic", "democracy", "monarchy", "aristocracy" e "anarchy" in testi stampati (prevalentemente) negli Stati Uniti dal 1760 fino al presente. Si vede, ad esempio, che gli anni a cavallo tra XVIII e XIX secolo hanno segnato un momento di particolare vivacità nell'uso del lessico relativo alle forme di governo. Si nota in particolare una maggiore frequenza d'uso nei due termini già più frequenti, ovvero "republic" e "monarchy", in quest'ordine. Questo secondo termine inizia però un lento declino che lo accompagnerà uniformemente fino al tempo presente, portandolo di poco sopra i livelli di "aristocracy" e "anarchy". Un altro termine che vede un aumento nella frequenza intorno al volgere del secolo è "democracy", il cui uso diminuirà negli anni successivi per aumentare sempre più verso la metà XIX secolo, e poi quasi arrivare a competere per frequenza con "republic" intorno alla metà del secolo successivo.

---

<sup>22</sup> È possibile effettuare ricerche tramite questa applicazione all'indirizzo <https://bookworm.htrc.illinois.edu/>

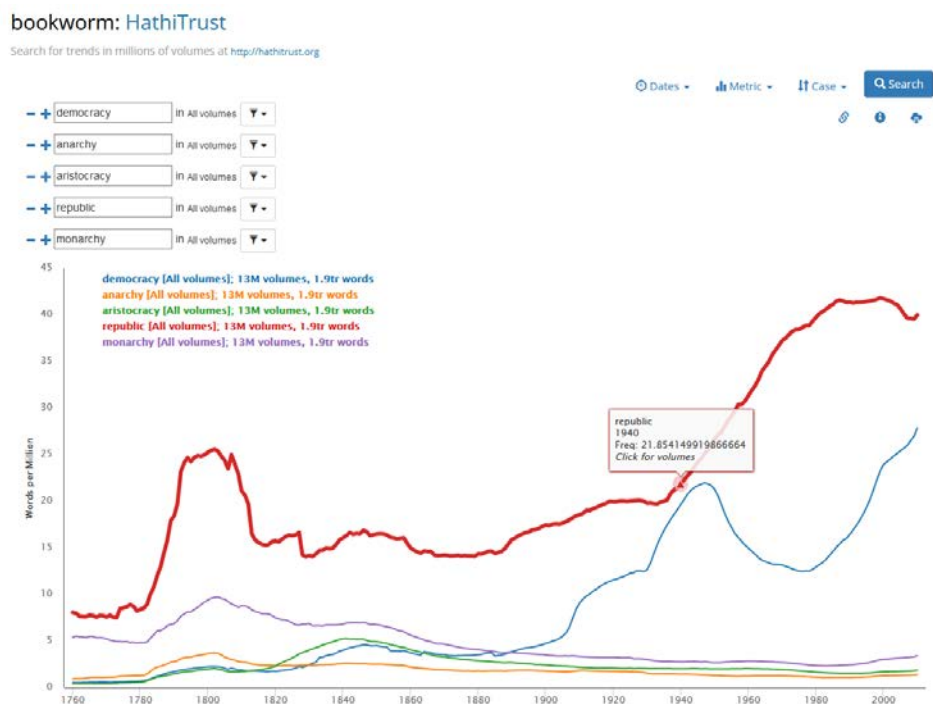


Figura 2: andamento di parole in testi stampati (prevalentemente) negli Stati Uniti dal 1760 fino al presente.

*HathiTrust* + *Bookworm* permette esclusivamente la ricerca di parole singole. Uno strumento simile, ma che permette ricerche più complesse e su corpora di maggiori dimensioni, è *Google Books Ngram Viewer*.<sup>23</sup> Si tratta di un progetto sviluppato dal gruppo di ricerca *Cultural Observatory* presso l'università di Harvard in collaborazione con Google. Come *Bookworm*, *Ngram Viewer* permette la produzione di grafici che evidenziano la frequenza relativa di parole o gruppi di parole.<sup>24</sup> *Ngram Viewer* utilizza come base dati un corpus contenente una parte dei testi di *Google Books*, suddiviso in diversi subcorpora in 8 lingue diverse. Le seguenti tabelle illustrano le dimensioni dichiarate di questi corpora per quanto riguarda l'inglese (americano e

23 A differenza di *Google Books Ngram Viewer*, *HathiTrust+Bookworm* permette però di risalire a e visualizzare i testi che contengono le occorrenze di una determinate parola in un dato anno. Da *Google Books Ngram Viewer* è possibile lanciare una ricerca in *Google Books* relativa a un'espressione cercata, ma i risultati della ricerca non corrispondono ai dati su cui si basa *Google Books Ngram Viewer*.

24 Per "ngram" si intende una o più parole ortografiche in sequenza. Il corpus *Google Books Ngram* contiene dati di frequenza calcolati per ngram contenenti fino a 5 parole e con una frequenza minima nel corpus di 40 occorrenze.

britannico),<sup>25</sup> l'italiano e il francese. In Tabella 1 vengono riportati, per ciascun anno corrispondente a una delle tre versioni di *Google Books Ngram Viewer*, i dati (ove reperibili, cfr. [23];[25];[29];[11]) relativi alle dimensioni di *Google Books* in numero di titoli (e percentuale stimata rispetto al totale dei titoli pubblicati) e alle dimensioni del corpus *Google Books Ngram* in numero di documenti, percentuale stimata sul totale e numero di parole.

Anno	Google Books (N. di documenti)	GB Ngram Corpus (N. di documenti)	GB Ngram Corpus (N. di parole)
2009	15 milioni (12% totale)	5,2 milioni (4% totale)	500 miliardi
2012		8,1 milioni (6% totale)	
2019	40 milioni		

Tabella 1: Dimensioni totali dei corpora.

In Tabella 2 sono riportati i dati (disponibili solo per il 2009 e il 2012)<sup>26</sup> relativi alle dimensioni dei tre subcorpora di *Google Books Ngram* per l'inglese, il francese e l'italiano (numero di documenti e di parole). I testi in lingua italiana non erano compresi nella prima versione di *Google Books Ngram Viewer* (in cui c'erano solo 3 lingue). La versione del 2012 contiene non solo una maggior quantità di testi rispetto alla versione precedente, ma i testi sono stati anche annotati linguisticamente, consentendo così di svolgere ricerche relativamente complesse su tutte le parole ed espressioni che ricorrono nel corpus. La versione più recente (rilasciata nel 2020) è stata migliorata in termini di qualità dei testi scannerizzati, arricchita di ulteriori annotazioni linguistiche e aggiornata all'anno 2019, anche se non sembra siano ancora disponibili informazioni esatte sulla sua consistenza.

	N. di documenti inglese	N. di parole inglese	N. di documenti francese	N. di parole francese	N. di documenti italiano	N. di parole italiano
2009		361 miliardi		45 miliardi	---	----

<sup>25</sup> Per "americano" e "britannico" si intende nei corpora di *Google Books Ngram* fare riferimento a documenti stampati rispettivamente nelle due aree geografiche.

<sup>26</sup> I dati sono derivati da Davies ([11]), che offre un accesso al corpus *Google Book Ngram* (versione 2009) attraverso l'interfaccia standard utilizzata per gli altri corpora da lui creati, consentendo l'utilizzo dei dati di frequenza per ricerche linguistiche complesse.

<b>2012</b>	4,5 milioni	468 miliardi	792.118	102 miliardi	305.763	40,3 miliardi
-------------	-------------	--------------	---------	--------------	---------	---------------

Tabella 2: Dimensioni totali dei corpora inglese, francese e italiano utilizzate da Google Books Ngram Viewer.

L'interfaccia di *Google Books Ngram Viewer* permette di effettuare ricerche selezionando l'intero corpus o subcorpora di specifiche lingue, per un determinato arco temporale e impostando i settaggi di visualizzazione.<sup>27</sup> Questo corpus multilingue non è costituito tanto da un insieme di testi, quanto da un insieme di dati relativi alla frequenza per anno di ciascun ngram. Essendo i dati di *Google Books Ngram Viewer* a disposizione della comunità scientifica,<sup>28</sup> è possibile verificare il numero preciso delle occorrenze su cui si basano i grafici relativi a “democrazia rappresentativa” (nelle tre lingue e, per quanto riguarda l'inglese, suddivisi tra dati americani e dati britannici) nei testi fino al 1799 (Tabella 3).

	1648	1773	1778	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	O	D
<b>GB</b>	1/1									3/3	20/7	7/5	7/7	4/4	11/11		27	10
<b>US</b>						13/3					4/1	2/2		5/1	1/1	2/2	52	37
<b>FR</b>			1/1	1/1	2/2	4/2			23/15	2/2	3/1	8/6	11/6	45/28	47/33	24/16	167	109
<b>IT</b>		2/1				1/1	1/1	1/1					2/2	20/8	13/13	6/6	44	33

Tabella 3: Occorrenze e titoli per anno di “representative democracy”, “démocratie représentative” e “democrazia rappresentativa” in Google Books Ngram Viewer.

In ogni casella sono indicati per ciascuna lingua/paese il numero totale di occorrenze e numero di testi in cui occorrono, tranne nelle ultime due colonne a destra in cui le somme sono indicate separatamente (O = numero di occorrenze singole, D = numero di documenti). La tabella contiene anche i falsi positivi, come il testo apparentemente pubblicato in Inghilterra nel 1648 (in realtà nel 1873) e quelli pubblicati in Italia nel 1773, 1789 e 1790 (non rinvenuti nelle ricerche). Anche se i numeri delle occorrenze di “democrazia rappresentativa” e dei testi in cui l'espressione compare in *Google Books Ngram Viewer* sono diversi da quelli delle pubblicazioni effettivamente rintracciate nella nostra ricerca, essi sembrano evidenziare una certa corrispondenza (vedi Tabella 4). Si tratta d'altra parte di dati non direttamente paragonabili tra loro, in quanto riferiti a basi dati diverse e ricerche non completamente omogenee (ad esempio, i numeri verificati qualitativamente nei database sono riferiti anche a testi che contengono l'espressione in forma plurale, a differenza di quelli in *Google Books Ngram Viewer*).

27 È possibile visualizzare i dati sulla base dei numeri per ciascun anno, oppure considerando una media calcolata su più anni (smoothing). Inoltre, vi è la possibilità di distinguere o meno tra termini che utilizzano maiuscole e minuscole.

28 È possibile scaricare i file contenenti questi dati all'indirizzo <https://storage.googleapis.com/books/ngrams/books/datasetsv3.html>

Lingua	Google Books Ngram Corpus	Verifica qualitativa nei database
Inglese	47	43
Francese	109	89
Italiano	33	16

Tabella 4: Numero di testi fino al 1799 in cui compaiono le espressioni "representative democracy", "démocratie représentative" e "democrazia rappresentativa".

Come si diceva, *Google Books Ngram Viewer* permette di verificare le tendenze relative alla frequenza di termini in diversi subcorpora. Nella Figura 3 è evidenziata la differenza nell'uso della locuzione "democrazia rappresentativa" (1750-2019) nei subcorpora di testi in italiano, francese e inglese rispettivamente.<sup>29</sup> Come si può vedere, in tutte e tre le lingue il termine compare alla fine del XVIII secolo, tende a regredire nell'uso nei decenni successivi e ricompare verso la metà del XIX secolo, anche se apparentemente più tardi in Italia rispetto a America/Gran Bretagna e Francia. L'uso del termine aumenta decisamente nel XX secolo, anche se nell'ultima parte di quel secolo e all'inizio di quello successivo si registra una relativa regressione in francese e italiano, lingue in cui peraltro l'espressione è più comune che in inglese.

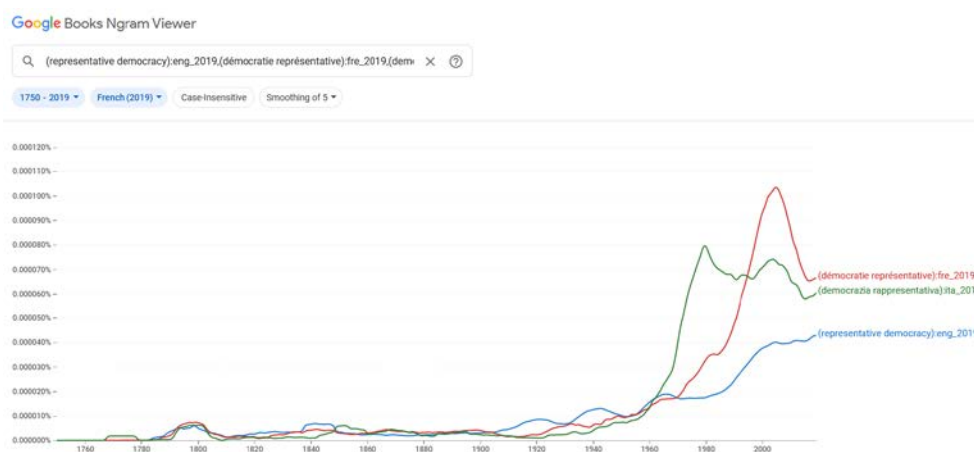


Figura 3: Frequenza di "representative democracy", "democrazia rappresentativa" e "démocratie représentative" (1750-2019).

È interessante vedere i termini a cui "representative democracy" si contrappone. Uno di questi è sicuramente "republic", termine la cui frequenza, a differenza di "democracy", è andata diminuendo nel corso degli ultimi due secoli. Allo scopo di individuare possibili termini

<sup>29</sup> La sintassi di ricerca è la seguente: (representative democracy):eng\_2019,(démocratie représentative):fre\_2019,(democrazia rappresentativa):ita\_2019.

alternativi o in competizione con “democrazia rappresentativa” abbiamo condotto delle ricerche in *Google Books Ngram Viewer* per identificare, per ciascuna delle lingue, due categorie di espressioni: tutti i sintagmi contenenti il sostantivo “democrazia” seguiti (o preceduti, per l’inglese) da un aggettivo, e tutti i sintagmi contenenti l’aggettivo “rappresentativo/a” preceduti (o seguiti, nel caso dell’inglese) da un sostantivo. In Figura 4 si vede il risultato di una ricerca per la parola “*démocratie*” seguita da un aggettivo,<sup>30</sup> che restituisce i 10 risultati più frequenti nell’arco di tempo considerato. Come si può vedere, l’aggettivo che più spesso occorre accanto a “*démocratie*” è “*représentative*”, seguito da “*directe*”, “*libérale*”, “*parlementaire*”, “*moderne*”, “*chrétienne*”, “*politique*”, “*française*”, “*populaire*” e “*socialiste*”. Si tratta di una rappresentazione di dati relativi alle frequenze di queste espressioni calcolate nel lungo periodo, da cui si evidenzia come, tranne “*démocratie représentative*” e “*démocratie française*”, che vengono introdotte alle fine del XVIII secolo, la maggior parte di queste espressioni inizia a diffondersi solo verso la metà del XIX secolo. Si può ad esempio cogliere la subitanea fortuna e il declino di espressioni come “*démocratie populaire*” negli anni immediatamente successivi la Seconda guerra mondiale, e l’andamento parallelo di “*démocratie chrétienne*” e “*démocratie socialiste*”, con un primo picco attorno al 1900 e poi un momento di maggiore frequenza negli anni ’70 del Novecento.

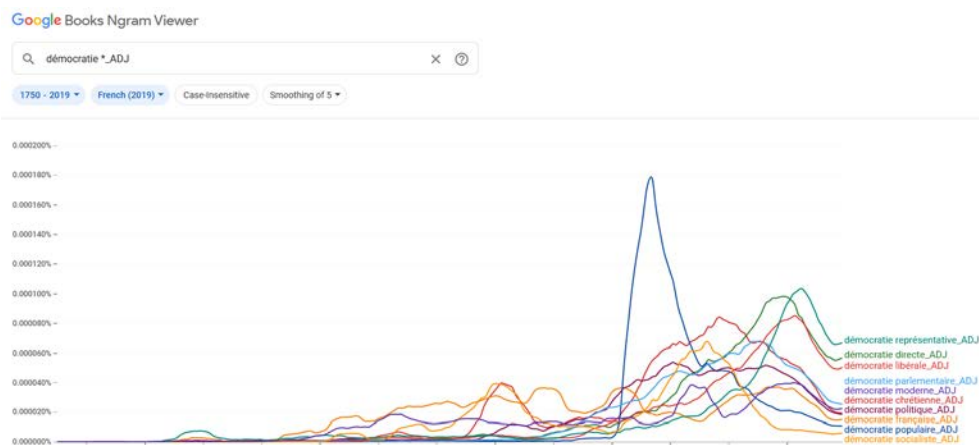


Figura 4: “*démocratie*” seguita da un aggettivo (1700-2019).

È ugualmente interessante, inoltre, analizzare il percorso storico del termine “*démocratie directe*”, che oggi si contrappone correntemente a “*démocratie représentative*”; sarebbe infatti fuorviante immaginare che questi due termini nascano in contemporanea, contrapponendosi l’uno all’altro. Come si può vedere nella Figura 5, la locuzione “*démocratie directe*” inizia a diffondersi solo circa un secolo dopo, quando cioè la categoria di “democrazia rappresentativa”

30 La sintassi di ricerca è la seguente: *démocratie \*\_ADJ*, in cui l’asterisco rappresenta qualsiasi numero di caratteri, mentre l’etichetta *\_ADJ* indica che la parola cercata deve essere un aggettivo.



era ormai stabilmente entrata nell'uso comune,<sup>31</sup> per poi seguire un percorso parallelo nel XX secolo. Queste evidenze lessicometriche trovano conferma anche nelle ricerche dettagliate compiute per il periodo fino al 1799, che hanno prodotto solo due occorrenze di "démocraties directes" fino a quell'anno ([26]: xv; [16]: 591).

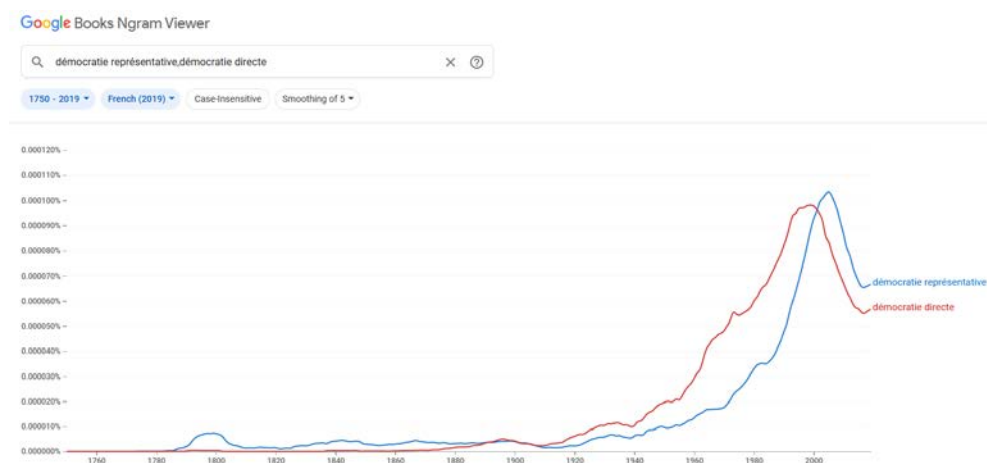


Figura 5: "démocratie directe" e "démocratie représentative" (1750-2019).

Mentre ricerche condotte su tutto il corpus permettono di evidenziare l'andamento della frequenza di uno o più termini nel corso del tempo, è possibile raffrontare la stessa ricerca riferita a due momenti temporali diversi per mettere in evidenza i differenti contesti verbali in cui termini correlati appaiono in contesti storici diversi. Così, ad esempio, è possibile vedere che nel ventennio 2000-2019 (Figura 6) gli aggettivi che più frequentemente appaiono come collocazioni di "democrazia", oltre a "rappresentativa" sono "liberale", "partecipativa", "Italiana" e "parlamentare".

31 Lo certifica il fatto che essa viene registrata dai dizionari (il primo che abbiamo reperito a dedicare uno specifico lemma alla "representative democracy" è la quinta edizione del *Law dictionary* di John Bouvier del 1855 ([5]:461) e compare nel titolo di fortunati trattati politici, in Francia ([20]) come negli Stati Uniti ([44]).

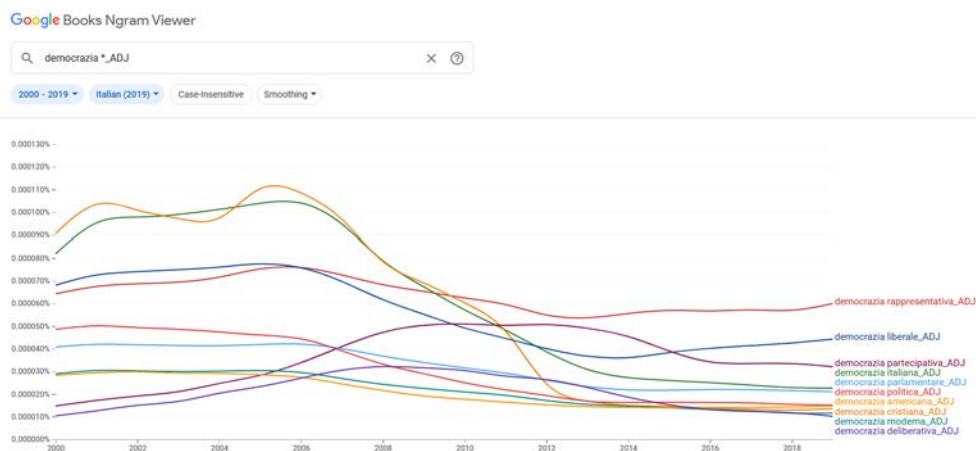


Figura 6: “democrazia + AGGETTIVO” (2000-2019).

Sintagmi letteralmente equivalenti si trovano eseguendo analoghe ricerche in francese e inglese (grafici qui non riportati); gli unici aggettivi che segnano delle particolarità per le tre lingue sono, rispettivamente, “Italiana” (in italiano), “locale” (in francese) e “American” (in inglese americano). Se invece conduciamo la stessa ricerca per il ventennio a cavallo dell’Ottocento (1790-1809) troviamo risultati diversi. Per l’inglese americano, ad esempio (Figura 7), le collocazioni più frequenti di “democracy” in quel periodo, in ordine di frequenza, sono “pure”, “perfect”, “simple”, “Athenian”, “French”, “wild”, “military”, “mere”, “representative” (unico aggettivo in comune tra quell’epoca e il momento attuale) e “tyrannical”. “Simple” è evidentemente, fino al 1799, l’aggettivo che in modo prevalente accompagna il sostantivo “democracy”.

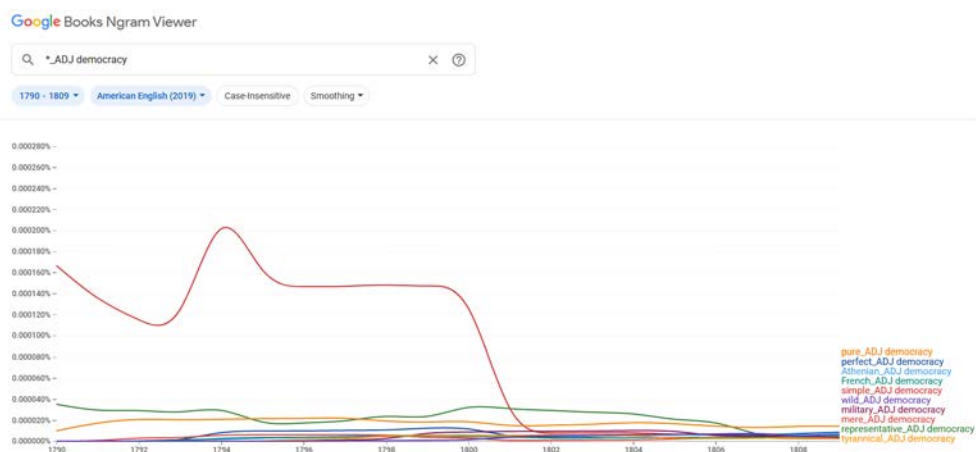


Figura 7: “AGGETTIVO + democracy” (1790-1809, inglese americano).

In inglese britannico (grafico qui non riportato) si registrano le medesime occorrenze, ad eccezione dell'aggettivo "absolute" in sostituzione di "perfect". In francese (grafico qui non riportato), per lo stesso periodo, gli aggettivi che più frequentemente occorrono accanto a "démocratie" sono, nell'ordine, "pure", "royale", "militaire", "absolue", "représentative", "parfaite", "simple", "française" e "génévoise". Risultati simili a quelli di queste due prime lingue si ottengono per l'italiano (grafico qui non riportato), in cui figura in aggiunta l'aggettivo "virtuosa".

La Tabella 5 riporta invece i sostantivi riferiti al lessico politico che occorrono più frequentemente precedendo "rappresentativo/a" / "représentatif/ive" e seguendo "representative" (per la totalità dei testi in inglese), per i due periodi considerati. La tabella condensa i risultati di diverse ricerche, dato che non è possibile in *Google Books Ngram Viewer* combinare le etichette riferite a parte del discorso (aggettivi, sostantivi, ecc.) con l'etichetta riferita all'inflessione (\_INF, nel nostro caso per individuare sia forme maschili che femminili, sia al singolare che al plurale). Inoltre, sono stati espunti tutti i sostantivi che si riferiscono ad altri domini, in primo luogo, in epoca contemporanea, alla statistica (in inglese, ad esempio, "sample" è il sostantivo che con più frequenza occorre accanto a "representative").

Periodo	2000-2019	1790-1809	2000-2019	1790-1809	2000-2019	1790-1809
	italiano	italiano	francese	francese	inglese	inglese
Sostantivi	<i>democrazia</i>	sistema	<i>démocratie</i>	<i>gouvernement</i>	<i>democracy</i>	body
	governo	costituzione	<i>gouvernement</i>	<i>ystème</i>	government	government
	potere,	forza	<i>ystème</i>	<i>assemblée</i>	institutions	system
	organo	governo	<i>régime</i>	corps	body	assembly
	funzione	consiglio	<i>conseil</i>	<i>constitution</i>		<i>democracy</i> <sup>34</sup>
	forza	corpo	<i>syndicat</i>	<i>démocratie</i>		
	regime	<i>democrazia</i>	<i>assemblée</i>	<i>régime</i>		
	monarchia		<i>institution</i>			
			<i>organisation</i>			

Tabella 5: "rappresentativo + SOSTANTIVO", in tre lingue e due diversi periodi storici.

Cosa ci dicono questi dati? Innanzitutto, che "rappresentativa" è oggi, insieme a "liberale" (e termini equivalenti in inglese e francese) l'aggettivazione principale di "democrazia" (e termini equivalenti in inglese e francese). In altre parole, per il lettore contemporaneo il concetto di "democrazia" è inteso prevalentemente come caratterizzato da questi due aggettivi, anche qualora essi non venissero esplicitati. Allo stesso modo l'aggettivo "rappresentativa" (e parole equivalenti in inglese e francese) richiama immediatamente, per il lettore contemporaneo, il concetto di "democrazia". Per gli attori storici della fine del Settecento, invece, il concetto di "democrazia" è principalmente qualificato come forma di governo legata al passato. Il neologismo "democrazia rappresentativa" è quindi caratterizzato in opposizione alla democrazia "pura" o "semplice",

ovvero la democrazia ateniese. Si tratta inoltre di una forma di governo la cui realizzazione nel presente (di allora) è legata a determinati ambiti territoriali (America, Francia, Italia) e ad aggettivi come “wild”, “military”, “tyrannical”, che al lettore di giornali del Ventunesimo secolo potrebbero apparire come lontani dal concetto di democrazia. Inoltre, si vede come l’idea di rappresentanza fosse a fine Settecento legata principalmente ad altri termini del lessico politico, e solo marginalmente alla parola “democrazia”.

### *Democrazia rappresentativa ieri e oggi*

Un’indagine condotta con diversi strumenti e risorse, ovvero i corpora consultabili su *Sketch Engine*, fornisce un’ulteriore conferma ai dati desunti da *Google Books Ngram Viewer*. Come si è detto, tra i corpora in lingua inglese disponibili sulla piattaforma lessicografica vi è anche un corpus contenente tre collezioni storiche dell’*Oxford Text Archive*, ed è possibile quindi mettere a confronto il profilo lessicale del termine “democrazia” in quel corpus con quello in un corpus di testi contemporanei. La Figura 8 e la Figura 9 mostrano una visualizzazione dei profili lessicali di questo termine rispettivamente in un corpus molto ampio di quotidiani pubblicati in tutto il mondo in lingua inglese tra il 2014 e il 2020<sup>32</sup> e in un subcorpus delle *English Historical Book Collections* relativo agli anni 1790-1809. I corpora presenti in *Sketch Engine* sono annotati linguisticamente e gli strumenti disponibili per l’analisi testuale permettono non solo una maggiore precisione nella ricerca delle occorrenze per mezzo di un linguaggio di ricerca molto flessibile (cfr. Figura 1), ma anche l’elaborazione di dati statistici sulla base delle annotazioni linguistiche. La Figura 8 e la Figura 9 mostrano le collocazioni più frequenti nell’immediato contesto di “democracy” non solo per quanto riguarda gli aggettivi (che confermano sostanzialmente le indicazioni desumibili dai grafici ricavati con *Google Books Ngram Viewer*), ma anche relativamente a verbi e sostantivi.<sup>33</sup> Si vede così come i sostantivi che con più frequenza sono associati a “democracy”<sup>34</sup> sono, nel XXI secolo (Figura 8), parole come “freedom”, “right”, “rule” e “justice”, mentre i verbi che più frequentemente hanno come oggetto “democracy” indicano il bisogno di rafforzare (“strengthen”, “consolidate”, “restore”) una forma di governo che appare minacciata (“threaten”, “subvert”) nel suo funzionamento (“function”). Per quanto riguarda invece i testi pubblicati tra il 1790 e il 1809 (Figura 9) “democracy” è contrapposta principalmente ad “aristocracy” (nelle sue varianti grafiche), “oligarchy”, “despotism”, “commonwealth” e “monarchy”, e può essere rovesciata (“overturn”) e odiata (“hate”), ma deve essere soprattutto realizzata (“erect”, “refine”, “introduce”). La particolare modalità di visualizzazione grafica proposta da questa funzione di *Sketch Engine* permette di percepire non soltanto la forza di attrazione tra due lemmi, espressa dalla vicinanza al centro del diagramma, ma anche la relativa frequenza, espressa dalle dimensioni dei caratteri e dei cerchi a cui le parole

---

32 Si tratta del corpus *Timestamped JSI web corpus 2014-2020 English*, contenente oltre 60 miliardi di parole.

33 Attraverso una serie di pulsanti, non visualizzati nelle figure, è possibile attivare o disattivare gruppi di parole in diverse relazioni sintattiche e grammaticali con il lemma cercato.

34 In questo caso si intende per “associazione” il fatto che tali parole compaiano in liste separate da una virgola o dalle congiunzioni “e” e “o”.

sono sovrapposte. Così, ad esempio, nel lessico giornalistico contemporaneo "democracy" è definita soprattutto dalle parole "freedom" e "liberal", mentre "representative" tende a definire "democracy" più fortemente di "parliamentary" o "western", nonostante la frequenza di co-occorrenza relativa di questi due ultimi aggettivi sia maggiore di quella di "representative". Tra la fine del Settecento e l'inizio dell'Ottocento invece la parola che indubbiamente meglio definisce "democracy" (per opposizione e contrasto) è "aristocracy", mentre "representative" si propone come uno di una serie di aggettivi che con "democracy" hanno una frequenza di co-occorrenza minore.

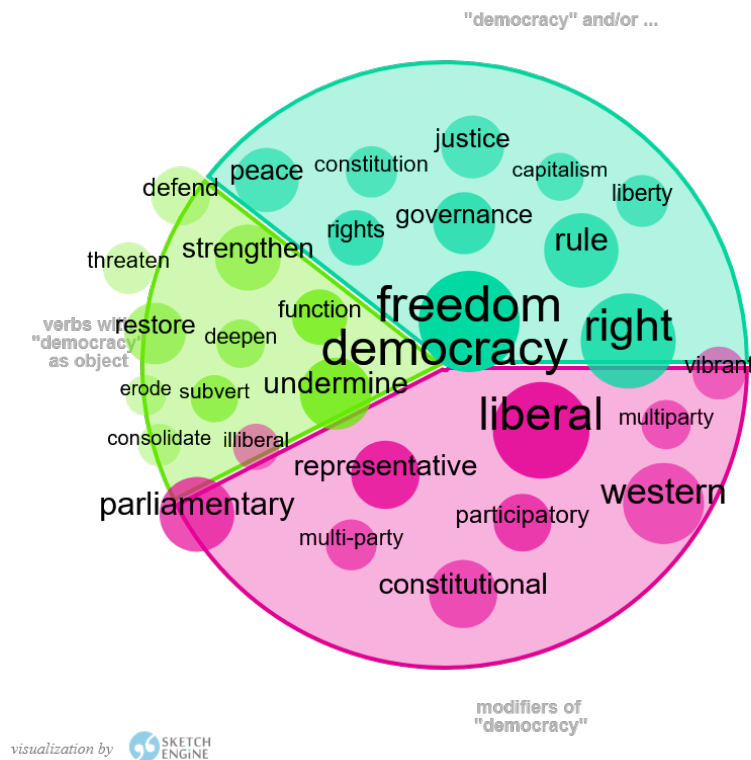
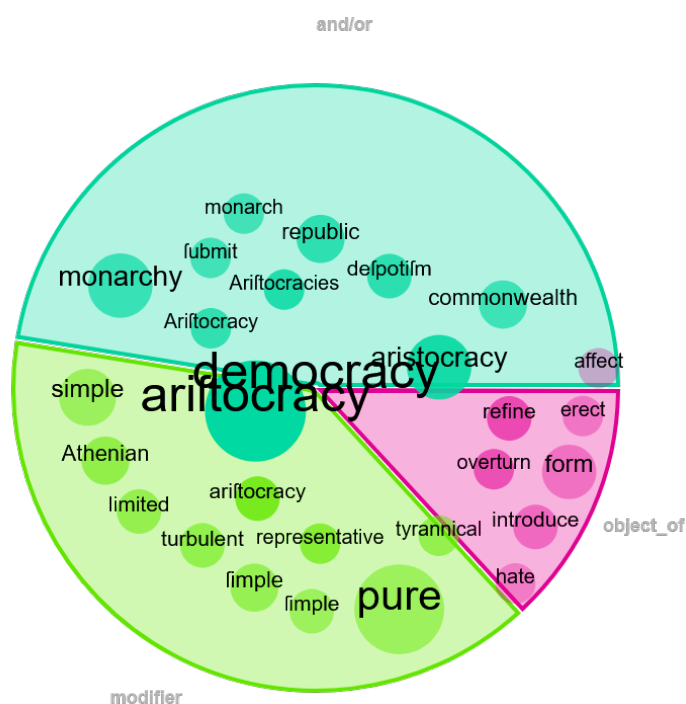


Figura 8: Profilo lessicale di "democracy" nel XXI secolo.



visualization by  SKETCH ENGINE

Figura 9: Profilo lessicale di “democracy” 1790-1809.

Nella Figura 9 la parola “aristocracy” appare in diverse varianti, ovvero viene conteggiata come diverso lemma a seconda che contenga la lettera “s” o la lettera “f”, sia in forma singolare o plurale, e abbia o meno l’iniziale maiuscola. Il termine “aristocracy” viene anche segnalato come appartenente al gruppo degli aggettivi che modificano il sostantivo. Questi dati si sovrappongono solo in parte a quelli rilevati con le ricerche in *Google Books Ngram Viewer* (Figura 7). Compaiono infatti “simple”, “Athenian”, “representative”, “tyrannical” e “pure” ma non “perfect”, “French”, “wild”, “military” e “mere”. Oltre a “turbulent” (che sembra rimandare alla sfera semantica di “wild” e “military”) troviamo qui invece “limited”. È importante sottolineare a questo punto che il corpus delle *English Historical Book Collections* contiene solo una parte (o una sovrapposizione parziale) dei testi contenuti nel corpus *Google Books Ngram*, e che per quanto riguarda il lemma “democracy” i numeri sono molto piccoli e quindi soggetti a un margine di errore più ampio. Ad esempio, i dati su cui si basa la Figura 9 sono calcolati su 803 occorrenze del lemma “democracy” (su un totale di 2.029 occorrenze nell’intero corpus), mentre le occorrenze del termine “representative democracy” sono 7, di cui solamente 3 con la grafia che appare in Figura 9 (di cui due da un testo duplicato).

Questa relativa scarsità di dati non rende possibile effettuare su questo corpus ricerche sulle collocazioni del sintagma "representative democracy", come è invece possibile fare sul corpus di testi giornalistici contemporanei (le cui dimensioni sono maggiori per diversi numeri di grandezza) (Figura 10). Come si può vedere la "representative democracy" è ora caratterizzata come multipartitica e parlamentare e come qualcosa che (non) è o (non) deve essere minato ("subvert", "undermine"), indebolito ("weaken") o eroso ("erode"), ma anche consolidato ("consolidate", "underpin").

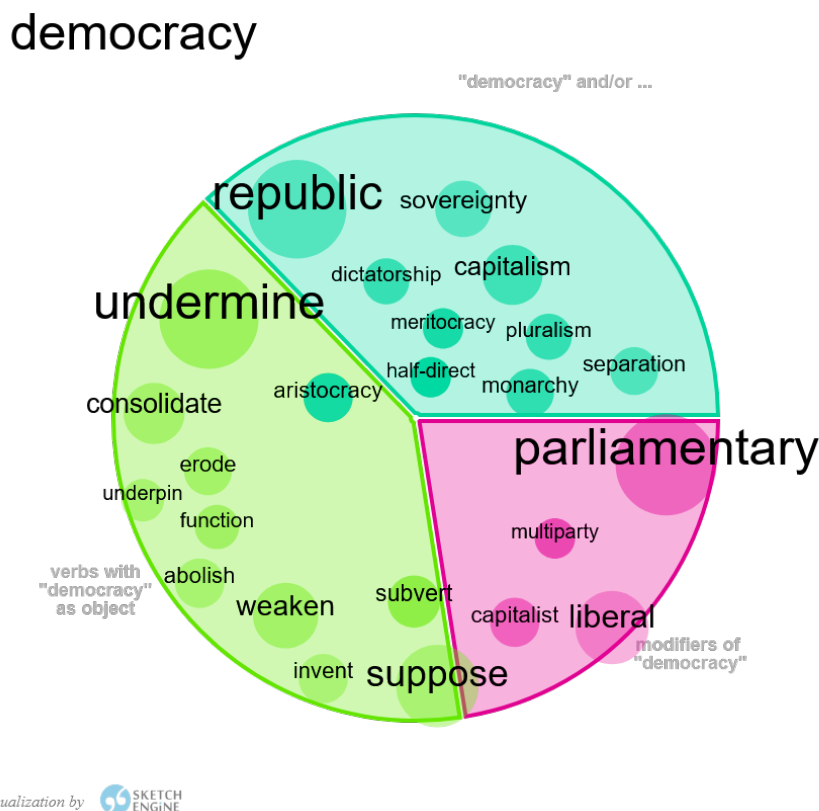


Figura 10: Profilo lessicale di "representative democracy" nel XXI secolo.

Le problematiche sopra evidenziate testimoniano da una parte dell'esigenza di tenere conto del livello di accuratezza dei testi e dall'altra della necessità di una verifica puntuale di singoli testi e occorrenze. Per quanto riguarda il primo aspetto, oltre al problema della standardizzazione grafica, è necessario contemplare il margine di errore nell'annotazione linguistica. Secondo il gruppo di ricerca *Google Ngram Viewer*, nei testi storici il livello di accuratezza nell'annotazione è minore che nei testi contemporanei "likely above 90% for part-of-speech tags and above 75% for dependencies. This implies a significant number of errors, which should be taken into account when drawing conclusions" ([42]); non c'è ragione di ritenere che questo non valga anche per i testi annotati presenti in *Sketch Engine*. Inoltre, il fatto che problemi di accuratezza

nel riconoscimento dei caratteri siano rilevabili solo caso per caso e non sull'intero corpus deve indurre a cautela rispetto all'assolutizzazione dei risultati rilevati, soprattutto su piccoli numeri di occorrenze e nel breve periodo, a causa della presumibile presenza di falsi positivi. Nondimeno, la flessibilità, versatilità e completezza delle ricerche effettuabili su corpora storici attraverso strumenti di analisi lessicografica come quelli presenti in *Sketch Engine* o nel *Corpus of Historical American English*, insieme alle diverse modalità di visualizzazione dei risultati, offrono un prezioso ausilio alla ricerca linguistica storica e alla ricerca storiografica effettuata sulla base di analisi testuali mettendo in luce evidenze lessicografiche altrimenti difficilmente rilevabili.

## Conclusioni

Seguendo il “filo rosso” di una ricerca sul sintagma “democrazia rappresentativa” si è cercato di mostrare che le risorse e gli strumenti digitali sviluppati negli ambiti dell'informatica umanistica, della linguistica computazionale e della linguistica dei corpora possono essere utili in due sfere di ricerca, quella sincronica/qualitativa (ristretta, nel nostro caso, a testi della parte finale del XVIII secolo) e quella diacronica (relativa allo studio comparativo delle occorrenze).

Nel primo ambito abbiamo fornito una descrizione delle principali risorse digitali disponibili per la ricerca storiografica, che sono state utilizzate allo scopo di rinvenire le occorrenze del sintagma “democrazia rappresentativa” dal 1778, primo anno in cui si registra l'occorrenza del sintagma, al 1799, scelto come *terminus ad quem* non solo perché è l'anno con cui si chiude il secolo, ma anche perché il colpo di Stato napoleonico del 9 novembre 1799 rappresenta un fondamentale momento di cesura politica. L'analisi delle occorrenze, dei testi e dei contesti d'uso ha permesso di verificare somiglianze e differenze nel significato attribuito al termine “democrazia rappresentativa” dagli attori sociali nel periodo in cui questo importante concetto nella storia del pensiero politico veniva introdotto nel dibattito culturale. Si è visto così, ad esempio, come questo termine potesse essere inteso sia in accezione positiva che negativa, a seconda dell'orientamento politico degli autori e del contesto nazionale di riferimento, ovvero Stati Uniti, Inghilterra, Francia, Svizzera e Italia (si rimanda a [36] per un'analisi approfondita delle occorrenze rinvenute).

Le risorse utilizzate per l'individuazione delle singole occorrenze in precisi testi si sono però dimostrate utili anche per analisi di tipo quantitativo. *Google Books Ngram Viewer* si è rivelato un insostituibile strumento di rilevazione delle tendenze linguistiche di medio-lungo periodo e l'immensa consistenza del corpus, la sua entità plurilingue (che consente analisi comparative) e la sua costituzione su base sostanzialmente casuale hanno permesso di contrapporre in una prospettiva temporale la frequenza d'uso di “democrazia rappresentativa” a quella di altri termini ad esso collegati. L'analisi puntuale del profilo lessicale della parola “democrazia” in un corpus di testi storici annotato linguisticamente, a confronto con il profilo lessicale della stessa parola in un corpus di testi contemporanei, ha permesso di porre in prospettiva l'orizzonte interpretativo di questo concetto nel momento in cui iniziava a definirsi in relazione al concetto di rappresentanza e in opposizione a termini che rimandano a differenti forme di governo.



La prospettiva sincronica e quella diacronica possono utilmente integrarsi, in quanto le osservazioni derivanti dalle ricerche per individuare il corpus ristretto di testi da sottoporre all'analisi (e lo studio puntuale dei testi stessi) possono indicare elementi utili alla ricerca condotta con strumenti per l'analisi diacronica (come *Google Ngram Corpus Viewer* e *HathiTrust + Bookworm*) e per quella sincronica (come *Sketch Engine*), mentre queste ultime possono fornire elementi utili per la ricerca di fonti testuali e per l'analisi "manuale" dei testi e dei contesti discorsivi individuati. Malgrado i problemi segnalati riguardo all'affidabilità delle fonti disponibili online, non c'è dubbio sul fatto che l'uso di queste risorse consenta di raggiungere risultati che con metodologie tradizionali di ricerca sarebbe impossibile ottenere.

## References

- [1] Adams, John. 1788. *A Defence of the Constitution and Government of the United States of America, Vol. III*, London: C. Dilly and J. Stockdale.
- [2] Adams, John. 1792. *Défense Des Constitutions Américaines, t. II*. Paris: Buisson.
- [3] Alatrash, Reem, Dominik Schlechtweg, Jonas Kuhn, and Sabine Schulte Im Walde. 2020. "CCOHA: Clean Corpus of Historical American English." *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, no. May: 6958–66.
- [4] Bolter, Jay David, and Richard Grusin. 1999. *Remediation: Understanding New Media*. Cambridge, MA: MIT Press.
- [5] Bouvier, John. 1855. *A Law Dictionary, Adapted to the Constitution and Laws of the United States of America, Vol. II*. Philadelphia: Printed for the estate of John Bouvier.
- [6] Brissot, Jacques Pierre. 1791. *Nouveau Voyage Dans Les États-Unis de l'Amérique Septentrionale Fait En 1788, t. I*. Paris: Buisson.
- [7] Brissot, Jacques Pierre. 1792. *New Travels in the United States of America, Performed in 1788*. London: J.S. Jordan.
- [8] Christophersen, Jes Andreas. 1968. *The Meaning of "Democracy" as Used in European Ideologies from the French to the Russian Revolution*. Oslo: Universitetsforlaget Trykningsentral.
- [9] Condorcet, Jean-Antoine-Nicolas de Caritat. 1847. *Ouvres*. Paris: Firmin Didot Frères.
- [10] Davies, Mark. 2012. "The 400 Million Word Corpus of Historical American English (1810–2009)." In *English Historical Linguistics 2010. Selected Papers from the Sixteenth International Conference on English Historical Linguistics (ICEHL 16), Pécs, 23-27 August 2010*, edited by Irén Hegedűs and Alexandra Fodor, 231–62. Amsterdam & Philadelphia: John Benjamins. <https://doi.org/10.1075/cilt.325.11dav>.
- [11] Davies, Mark. 2014. "Making Google Books N-Grams Useful for a Wide Range of Research on Language Change." *International Journal of Corpus Linguistics* 19 (3): 401–16. <https://doi.org/10.1075/ijcl.19.3.04dav>.

- [12] Deleplace, Marc. 2001. *L'anarchie de Mably à Proudhon (1750-1850). Histoire d'une Appropriation Polémique*. Lyon: ENS.
- [13] Dupuis-Déri, Francis. 2013. *Démocratie. Histoire Politique d'un Mot Aux Etats-Unis et En France*. Montréal: Lux Editeur.
- [14] Ferreira, Oscar. 2013. «La Démocratie Dans Toute Sa Pureté». Une Longue Histoire de La Sortie En Politique Du Concept d'ochlocratie (1780-1880)." *Revue de La Recherche Juridique – Droit Prospectif* 2: 605–41.
- [15] Gainot, Bernard. 2001. *1799, Un Nouveau Jacobinisme? La Démocratie Représentative, Une Alternative à Brumaire*. Paris: CHTS.
- [16] "Gazette Nationale Ou Le Moniteur Universel, 28 Pluviôse an V [16 Febbraio 1797]." n.d.
- [17] Grandjonn, Jacques. 2013. *Communisme/Kommunismus/Communism. Origine et Développement International de La Terminologie Communautaire Prémarxiste Des Utopistes Aux Néo-Babouvistes 1785-1842*. Editions des Malassis.
- [18] Guerci, Luciano. 1999. *Istruire Nelle Verità Repubblicane. La Letteratura Politica per Il Popolo Nell'Italia in Rivoluzione (1796-1799)*. Bologna: Il Mulino.
- [19] Guilhaumou, Jacques. 2000. "De l'histoire Des Concepts à l'histoire Linguistique Des Usages Conceptuels." *Genèses* XXXVIII (1): 105–18.
- [20] Hayem, Armand. 1869. *La Démocratie Représentative*. Paris: Le Chevalier.
- [21] Kilgarrieff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1 (1): 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- [22] Kurunmäki, Jussi, Jeppe Nevers, and Henk te Velde, eds. 2018. *Democracy in Modern Europe. A Conceptual History*. New York: Berghahn Books.
- [23] Lee, Haimin. 2019. "15 Years of Google Books." The Keyword - Google. 2019. <https://www.blog.google/products/search/15-years-google-books/>.
- [24] Li, Ying, Thomas Hills, and Ralph Hertwig. 2020. "A Brief History of Risk." *Cognition* 203 (January): 104344. <https://doi.org/10.1016/j.cognition.2020.104344>.
- [25] Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman, and Slav Petrov. 2012. "Syntactic Annotations for the Google Books Ngram Corpus." In *Proceedings Of the 50th Annual Meeting Of the Association for Computational Linguistics*, 169–74.
- [26] Mallet Du Pan, Jacques. 1796. *Correspondance Politique Pour Servir à l'histoire Du Républicanisme Français*. Hambourg: P.F. Fauche.
- [27] Mallet Du Pan, Jacques. 1798. *Essai Historique Sur La Destruction de La Ligue et de La Liberté Helvétiques*. Londres: W. et C.
- [28] Mallet Du Pan, Jacques. 1799. *The History of the Destruction of the Helvetic Union and Liberty*. Boston: Manning & Loring.

- [29] Michel, Jean-Baptiste, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, et al. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–82. <https://doi.org/10.1126/science.1199644>.
- [30] Monnier, Raymonde. 2001. "«Démocratie Représentative» Ou «République Démocratique»: De La Querelle Des Mots (République) à La Querelle Des Anciens et Des Modernes." *Annales Historiques de La Révolution Française* 3: 1–21.
- [31] Nunberg, Geoffrey. 2009. "Google's Book Search: A Disaster for Scholars." *The Chronicle of Higher Education*, 2009.
- [32] Palmer, Robert Roswell. 1953. "Notes on the Use of the Word «Democracy» 1789-1799." *Political Science Quarterly* 68 (2): 203–26.
- [33] Piguet, Marie-France. 1996. *Classe. Histoire Du Mot et Genèse Du Concept Des Physiocrates Aux Historiens de La Restauration*. Lyon: Presses Universitaires de Lyon.
- [34] Pocock, John Greville Agard. 1990. *Politica, Linguaggio e Storia*. Milano: Edizioni di Comunità.
- [35] Proietti, Fausto. 2020. *L'invenzione Della Democrazia. Pensiero Politico e Istituzioni Nella Seconda Repubblica Francese (1848–1852)*. Canterano (RM): Aracne.
- [36] Proietti, Fausto, and Federico Zanettin. 2021. "«Democrazia rappresentativa». Indagine sulle origini di una categoria politica (1778-1799)." *Storia del pensiero politico* 1: 41–66. <https://doi.org/10.4479/100746>.
- [37] Rickman, Ollie. 2011. "The British Library and Google to make 250,000 books available to all," The British Library. 2011. <https://www.bl.uk/press-releases/2011/june/the-british-library-and-google-to-make-250000-books-available-to-all>
- [38] Sérionne, Jacques Accarias de. 1778. *La Richesse de La Hollande*. Londres: aux dépenses de la Compagnie.
- [39] Serna, Pierre. 1997. *Antonelle, Aristocrate et Révolutionnaire (1747-1817)*. Paris: Éditions du Félin.
- [40] Skinner, Quentin. 2001. *Dell'interpretazione*. Bologna: Il Mulino.
- [41] Sperberg-McQueen, Michael. 1994. "Textual Criticism and the Text Encoding Initiative." Archived from the Original on 4 March 2016. 1994. <http://xml.coverpages.org/sperb-mla94.html>.
- [42] The Google Ngram Viewer Team. n.d. "What Does the Ngram Viewer Do?" Google Books Ngram Viewer. Accessed January 12, 2021. <https://books.google.com/ngrams/info>.
- [43] Thylstrup, Nanna Bonde. 2018. *The Politics of Mass Digitization*. Cambridge, MA & London: MIT Press.
- [44] Ulrich, Bartow Adolphus. 1880. *Comparison Between the Forms of Government of the Representative Democracy, or Republic of the Unites States, and Those of Other Nationalities*. Chicago: The Chicago Legal News Company.
- [45] Urbinati, Nadia. 2006. *Representative Democracy. Principles & Genealogy*. Chicago & London: University of Chicago Press.

- [46] Younes, Nadja, and Ulf Dietrich Reips. 2019. “Guideline for Improving the Reliability of Google Ngram Studies: Evidence from Religious Terms.” *PLoS ONE* 14 (3): 1–17. <https://doi.org/10.1371/journal.pone.0213554>.
- [47] Zancarini, Jean-Claude. 2007. “Une Philologie Politique. Les Temps et Les Enjeux Des Mots (Florence, 1494-1530).” *Laboratoire Italien* 7: 61–74.