Alice Perin

Università di Pisa

Matricola n. 626005

*Relazione – Metodi della Fisica per le Scienze Umane A.A. 2020/21*

Legge di Zipf: Mary Wollstonecraft e Virginia Woolf a confronto

Applicazione delle leggi di potenza al testo

1. Introduzione

La mia passione per la lingua inglese e la curiosità verso la cultura britannica, mi ha portato a scegliere “A Vindication of the Rights of Woman” e “Three Guineas” come cuore di questa relazione. Questi due capolavori della letteratura inglese non hanno in comune solo la lingua in cui sono stati scritti, ma anche il fatto che sono stati composti da donne. Le scrittrici Mary Wollstonecraft e Virginia Woolf nascono a più di un secolo di distanza, rispettivamente nel 1759 e nel 1882, in Inghilterra. Nella città di Londra, in particolare, le donne d’Europa iniziano a far valere le loro idee e i loro diritti a partire proprio da Mary Wollstonecraft (nota anche come Mary Shelley, moglie del poeta Percy Bysshe Shelley), considerata una delle prime attiviste del movimento Femminista anche grazie alla sua opera “Rivendicazione dei diritti della donna” del 1792. Per la prima volta nella storia, una donna scrive sulla condizione e i diritti delle donne, sostenendo che sono indispensabili alla nazione per il loro ruolo di madri e mogli, educatrici dei figli e quindi parte integrante della società. Da questo momento in poi, all’interno della propaganda Femminista nasceranno altre opere con scopo di disseminare idee di uguaglianza dei diritti e l’autodeterminazione del proprio corpo. L’espansione dell’ideologia inizia piano piano ad insinuarsi nella società Inglese ed Europea, soprattutto attraverso la letteratura. Wollstonecraft nel suo saggio, veicola con fervore che la donna non è “per natura” inferiore all’uomo, ma è la società che la relega ad un ruolo e a una condizione considerati infima e subordinata al marito. Per la prima volta in una pubblicazione, vengono messi sullo stesso piano i diritti dell’uomo e della donna, considerando quest’ultima come essere umano (British Library, bl.uk), prima dell’essere moglie e madre.

Il secondo romanzo in questione, “Three Guineas” di Virginia Woolf viene pubblicato nel 1938, alla vigilia della Seconda guerra mondiale. Nasce dalla lettera immaginaria di un avvocato alla scrittrice, che chiede come si può evitare il conflitto mondiale e come le donne possono contribuire a questo ideale. È un saggio epistolare complesso, ricco di spunti e di idee di azioni concrete. La sua risposta all’avvocato è in sintesi: le tre simboliche ghinee a disposizione dovrebbero essere date una al fondo per l’istruzione femminile, l’altra al fondo per garantire alle donne la possibilità di accedere alla libera professione, e la terza a un’associazione femminile pacifista chiamata “La Società delle Estranee”. L’autrice mette così sul piatto della bilancia il sistema patriarcale, il militarismo e l’esclusione delle donne dalla politica e dal potere (Zodiaco, 2018). Giudica la guerra e la violenza come un riflesso della società che ha a capo degli uomini, relegando la donna ad uno status di inferiorità perché esclusa dal potere. La donna ha, secondo la Woolf, il “vantaggio” di essere esclusa da questo “mondo”, ma per portare i suoi valori nella società le deve essere concesso di accedere alle libere professioni e prima di tutto essere istruita.

Cos’è cambiato tra la pubblicazione di “A Vindication of the Rights of Woman” e “Three Guineas” nella società? Quali parole vengono usate da Mary Wollstonecraft per esprimere i propri ideali di uguaglianza e quali quelle usate da Woolf più di un secolo dopo per veicolare le stesse idee femministe?

Questa relazione vuole confrontare i due testi dal punto di vista delle parole più frequenti, distribuzione di frequenza secondo la Legge di Zipf, e correlazione matematica tra le occorrenze delle parole nei due romanzi attraverso l’uso di strumenti computazionali e *tools* per l’analisi linguistica. Il secondo obiettivo, ugualmente importante, è riflettere sulle loro differenze ma soprattutto osservare le similarità semantiche tra il linguaggio della prima femminista inglese Mary Wollstonecraft del 1792 e l’introspettiva e tormentata attivista Virginia Woolf del 1938.

1. **Metodologia**

Per quanto riguarda la parte dell’analisi, ho usato alcuni strumenti computazionali per tokenizzare il testo (posizionare una parola per riga per permetterne il conteggio), nello specifico un programma scritto in linguaggio Python che usa la libreria NLTK per suddividere il testo in unità minime (token) sulla base di modelli probabilistici, in questo caso della lingua inglese. Una volta tokenizzato il testo, ho usato il prompt dei comandi e le sue funzioni: “wc” per il conteggio delle parole, “sort” per ordinare i token in ordine alfabetico, “uniq -c” per eliminare i doppioni della parola e contarne le occorrenze nel testo. L’output di questa operazione è formato da una colonna di frequenze, con il token corrispettivo nella seconda colonna.

Con l’ausilio di Excel, ho provveduto a creare la distribuzione di Zipf costruendo un grafico avente come primo parametro il LOG­10 del rango delle parole, e come secondo parametro il LOG­10 delle frequenze assolute. Attraverso un programma di analisi linguistica, ho calcolato la densità lessicale come il rapporto tra la quantità di parole “piene” (verbi, sostantivi, aggettivi) e il totale delle parole.

Ho poi personalmente scritto un programma Python ad hoc, che dà come output tutte le parole piene presenti nei due testi, insieme alle frequenze assolute. Calcolando la frequenza relativa di ogni parola, ho costruito un grafico a dispersione, inserendo una linea di tendenza per mostrare la correlazione semantica tra le parole piene dei due testi.

1. **Dati e analisi**

Il primo dato fondamentale per effettuare analisi, è verificare che la legge di Zipf sia applicabile al testo preso in considerazione. La legge di Zipf è una legge teorica che gode dell’invarianza di scala, che significa che è valida per qualsiasi fenomeno, non importa quanto “grande” sia. La formula è la seguente:

Nell’ evento , la frequenza dell’evento C è in funzione del rango *z*. L’apice *a*, del rango, è un indice inverso della ricchezza lessicale, che secondo Zipf è molto vicino e approssimabile a 1. Quel che mi ha fatto interessare a questa legge, è che è applicabile a numerosi fenomeni naturali e sociali: la distribuzione del reddito, del quoziente intellettivo, della popolazione nelle città, ecc. Applicando la legge di Zipf ad un testo, l’evento è una particolare parola e la sua distribuzione è data dal rapporto dalla sua frequenza assoluta (C) e il suo rango (numero assegnato partendo dalla parola più frequente che cresce man mano che si arriva alla parola meno frequente). Ad esempio, se la parola più frequente *x* occorre 100 volte in un testo, la seconda parola più frequente *y* (quindi con rango 2) secondo la legge teorica dovrebbe occorrere 50 volte ().

Qui di seguito, i grafici della distribuzione di Zipf relativi ai due testi, in doppia scala logaritmica, creati grazie ad Excel.

*Grafico 1. Distribuzione di Zipf in doppia scala logaritmica a confronto*

Come si può evincere dai grafici, entrambe le distribuzioni nei testi soddisfano la legge di Zipf, rispettando la classica retta disegnata dalla frequenza discendente dei vocaboli. Il secondo step di analisi riguarda in primo luogo l’esposizione delle prime 20 parole vuote e piene di ciascun testo, allo scopo di evidenziare in qualche modo il “cuore” semantico del testo. Ecco i due testi a confronto:

*A Vindication of the Rights of Woman*

|  |  |  |  |
| --- | --- | --- | --- |
| Frequenza | Parole Grammaticali | Frequenza | Parole Lessicali |
| 4905 | the | 428 | women |
| 3752 | of | 384 | will |
| 3304 | to | 344 | man |
| 2402 | and | 293 | should |
| 1787 | a | 292 | men |
| 1360 | that | 252 | reason |
| 1215 | in | 232 | mind |
| 882 | not | 202 | woman |
| 821 | their | 186 | virtue |
| 817 | by | 164 | life |
| 781 | they | 161 | love |
| 762 | for | 147 | same |
| 748 | It | 142 | nature |
| 733 | I | 137 | character |
| 656 | which | 135 | own |
| 602 | with | 132 | children |
| 569 | As | 130 | never |
| 556 | or | 129 | sex |
| 467 | her | 129 | respect |
| 434 | on | 126 | heart |

*Tabella 1. Prime 20 parole grammaticali e lessicali del testo A Vindication of the Rights of Woman*

*Three Guineas*

|  |  |  |  |
| --- | --- | --- | --- |
| Frequenza | Parole Grammaticali | Frequenza | Parole Lessicali |
| 4313 | the | 254 | men |
| 2668 | of | 222 | women |
| 2448 | to | 218 | must |
| 1928 | and | 214 | educated |
| 1358 | that | 211 | will |
| 1289 | a | 171 | war |
| 1233 | in | 137 | own |
| 796 | it | 132 | man |
| 664 | for | 127 | daughters |
| 628 | we | 124 | help |
| 599 | not | 121 | woman |
| 500 | you | 116 | money |
| 497 | by | 116 | fact |
| 459 | as | 110 | education |
| 457 | which | 97 | time |
| 412 | with | 91 | profession |
| 385 | or | 80 | professions |
| 360 | her | 79 | work |
| 350 | us | 79 | great |
| 343 | The | 76 | public |

*Tabella 2. Prime 20 parole grammaticali e lessicali del testo Three Guineas*

Dal punto di vista della percentuale di parole lessicali significative, i due testi sono simili. In particolare, il testo di Mary Wollstonecraft ha una minima percentuale in più di parole lessicali significative rispetto a Three Guineas. Ho ricavato questi dati grazie a un programma Python che attraverso il Part Of Speech tagging (POS-tag), seleziona solo le parole etichettate come “verbo”, “sostantivo” e “aggettivo”.

*Grafico 2. Percentuale di parole piene (lessicali) e di parole vuote (grammaticali)*

Di seguito i dati di densità lessicale nello specifico, divisi in percentuali di parte del discorso, comprendendo però anche avverbi, preposizioni, pronomi, verbi ausiliari. La densità lessicale complessiva di un testo, è il rapporto tra la quantità di occorrenze di parole lessicali, e il numero di parole totali. Questo concetto misura di fatto la complessità linguistica del parlante o in una composizione scritta attraverso le parole funzionali (grammaticali) e le parole contenuto (lessicali). La densità lessicale di un individuo si evolve con l’età, la sua educazione, contesto sociale, stile di scrittura ed influisce sulla leggibilità del testo stesso (“facilità” con cui si legge e si comprende una composizione).

Le leggere discrepanze tra il grafico 2 e la tabella 3 in quanto a percentuale di parole piene sono dovute al fatto che per costruirli ho utilizzato due strumenti diversi che potrebbero quindi essere diversamente sensibili alle parole e “smistarle” in categorie diverse, in base al tipo di addestramento che è stato usato per creare lo strumento stesso. Inoltre, il grafico 2 comprende solo le categorie: sostantivi, verbi e aggettivi.

*A Vindication of the Rights of Woman*

|  |  |
| --- | --- |
| Parti del discorso | Percentuali |
| Sostantivi | 23.86% |
| Aggettivi | 8.71% |
| Verbi | 11.26% |
| Avverbi | 5.92% |
| Preposizioni | 14.28% |
| Pronomi | 5.6% |
| Verbi ausiliari | 5.65% |

|  |  |
| --- | --- |
| Densità Lessicale totale | 49.75% |

*Three Guineas*

|  |  |
| --- | --- |
| Parti del discorso | Percentuali |
| Sostantivi | 25.76% |
| Aggettivi | 7.18% |
| Verbi | 10.75% |
| Avverbi | 5.19% |
| Preposizioni | 14.22% |
| Pronomi | 6.32% |
| Verbi ausiliari | 6.02% |

|  |  |
| --- | --- |
| Densità Lessicale totale | 48.88% |

*Tabella 3. Percentuali di parti del discorso dei due testi*

Dopo aver raccolto le parole piene di ciascun testo, ho stilato la lista dei vocaboli che appaiono in entrambi i saggi, unendo le due liste. Successivamente, ho creato una tabella che riportasse nella prima colonna la lista dei vocaboli totali, nella seconda colonna le occorrenze di quella parola nel primo testo e nella terza colonna le occorrenze nel secondo testo. Mediante l’uso di Excel, ho creato il grafico a dispersione riportato di seguito prendendo come input le due colonne che rappresentano le occorrenze.

*Grafico 3. Linea di tendenza.*

La correlazione è un valore tra 1 (correlazione perfetta) e -1 (anti-correlazione), ricavata tramite la definizione matematica:

La linea di tendenza nel grafico mostra la correlazione tra un insieme di valori messi in corrispondenza diretta, e in questo caso possiamo vedere che la retta passante per l’origine è inclinata di quasi 45°, denotando una correlazione vicina al valore 1. Il motivo di questa correlazione “quasi” perfetta, la possiamo trovare nella semantica delle parole piene significative che occorrono più spesso nei due testi. Le due autrici compongono questi saggi femministi a più di un secolo di distanza. Nonostante il tempo trascorso, i due scritti presentano grandi similarità: innanzitutto perché parlano di diritti delle donne, mettendo a confronto i due sessi ed è per questo motivo che troviamo “women” e “men” in cima ad entrambe le liste.

|  |  |
| --- | --- |
| A Vindication of the Rights of Woman | |
| 428 | women |
| 384 | will |
| 344 | man |
| 293 | should |
| 292 | men |
| 252 | reason |
| 232 | mind |
| 202 | woman |
| 186 | virtue |
| 164 | life |
| 161 | love |
| 147 | same |
| 142 | nature |
| 137 | character |
| 135 | own |
| 132 | children |
| 130 | never |
| 129 | sex |
| 129 | respect |
| 126 | heart |

|  |  |
| --- | --- |
| Three Guineas | |
| 254 | men |
| 222 | women |
| 218 | must |
| 214 | educated |
| 211 | will |
| 171 | war |
| 137 | own |
| 132 | man |
| 127 | daughters |
| 124 | help |
| 121 | woman |
| 116 | money |
| 116 | fact |
| 110 | education |
| 97 | time |
| 91 | profession |
| 80 | professions |
| 79 | work |
| 79 | great |
| 76 | public |

*Tabella 4. Prime 20 parole piene per frequenza*

Oltre a “uomo” e “donna”, le parole significative in comune (usate da entrambe), appartengono sia a una categoria che rimanda alla dimensione prettamente “umana”, come ad esempio “sex”, “daughters”, “life”, “love” (“genere”, “figlie”, “vita”, “amore”). Questi vocaboli definiscono l’universo del femminismo del quale le due autrici parlano, veicolandone i valori, avanzando richieste alla società, esprimendo desideri, lottando per una condizione di vita migliore. In particolare, entrambe le scrittrici utilizzano molto “woman” e “man”, per comparare l’uomo e la donna, nel loro stile di vita, nei diritti e nei doveri di fronte alla società. Selezionando tra le parole significative frequenti che vengono utilizzate da Mary Wollstonecraft troviamo “nature”, “character”, “reason” (“natura”, “carattere”, “ragione”). Di seguito qualche loro concordanza contenuta in frasi rappresentative del saggio:

**Nature**

* “…constitution is founded on the *nature* of man…”
* “…they must return to *nature* and equality.”
* “…should differ in respect to their *nature*.”
* “…arguments in favour of a state of *nature* are plausible, but unsound.”

**Character**

* “…laudable ambition is to obtain a *character* as a human being, regardless…”
* “…with respect to the female *character*…”
* “…wives, and mothers, their moral *character* may be estimated by their manner…”

**Reason**

* “…rooted prejudices have clouded *reason”*
* “…arguments may be drawn from *reason”*
* “…woman partake with him the gift of *reason*?”

A mio avviso, queste tre parole descrivono i temi *clou* del saggio della Wollstonecraft, e denotano già da sole il personaggio d’avanguardia che lei era nel XVIII secolo. La parola “Nature”, “Character” e “Reason”, sono usate per mettere descrivere la contrapposizione dei due sessi della società, con la replica dell’autrice: non importa quale sia il nostro carattere, la nostra personalità. La natura degli uomini e delle donne è Umana. Bisogna usare la ragione, sconfiggendo i pregiudizi. Mary si rifiutava di essere legata alla mentalità dell’epoca, si rifiutava di credere che le donne fossero ineducabili, inferiori, incapaci di pensare razionalmente (Scobie, 2017). Il suo saggio era di fatto una risposta allo scritto di Edmund Burke’s “Reflections on the French Revolution”, che sosteneva che la Rivoluzione Francese sarebbe stata un fallimento in quanto la società aveva bisogno di tornare alla sua struttura tradizionale (British Library). L’anno dopo, risponde con “La Rivendicazione dei Diritti delle Donne”, che ha un grande impatto sui movimenti femministi nel secolo successivo: una dimostrazione che le parole provocano effetti nel mondo “reale”.

Il saggio “Three Guineas” è stato pubblicato nel 1938, a scopo di protesta contro il conflitto mondiale e in forma epistolare come risposta a una lettera immaginaria che chiede il contributo alla risoluzione di tre diversi problemi: la prevenzione della guerra, la ricostruzione di un *college* femminile e la disoccupazione femminile. La Woolf risponde con delle soluzioni, iniziando con la denuncia al legame tra il genere maschile e il fascismo, come risultato di una società patriarcale. L’iniqua distribuzione del potere centrata sulla figura maschile, secondo lei, ha creato di conseguenza un’esclusione sistematica della donna dalle cariche politiche, di prestigio e di potere. Questo fenomeno crea una separazione della sfera pubblica e privata, relegando la donna alla seconda, togliendole non solo il potere in ambito pubblico, ma anche all’interno della famiglia (Gättens, 2001). I tre vocaboli significativi che l’autrice utilizza sono:

**War**

• “…the great majority of your sex are today in favour of *war*.”

• “*War*, you say, is an abomination; a barbarity; *war* must be stopped at whatever cost.”

• “If the working women of the country were to say: 'If you go to *war*, we will refuse to make munitions or to help in the production of goods,' the difficulty of *war*-making would be seriously increased.”

**Education /educated**

• “What sort of *education* will teach the young to hate war?"

• “Yet the part that *education* plays in human life is so important.”

• “The daughters of *educated* men have no direct influence […] but they possess the greatest power of all; that is, the influence that they can exert upon educated men.”

• “…and that influence of the kind that can be exerted by the daughters of *educated* men is very low in power, very slow in action, and very painful in use.”

**Profession**

• “Marriage, the one great *profession* open to our class since the dawn.”

• “…because it is an unpaid *profession*…”

• “…that to depend upon a *profession* is a less odious form of slavery than to depend upon a father.”

Queste tre parole, “Guerra”, “Professione” e “Educazione”, vengono usate per comunicare le soluzioni della scrittrice ai problemi delle tre lettere: come possiamo fermare la guerra? Come possiamo istituire l’educazione femminile? Come si può introdurre la donna nel mondo del lavoro? Woolf riconduce i problemi ad un unico male: il sistema patriarcale.

Al contrario di Wollstonecraft, che propone e pretende un cambiamento nella società, che in qualche modo agita le mani dicendo “siamo umane, abbiamo dei diritti, siamo necessarie alla società”, Woolf descrive in modo più concreto cosa una donna può fare per usare la sua condizione di inferiorità per riscattarsi. Essendo le donne sempre state escluse dalle dinamiche di potere della società, relegate a certi compiti, hanno sviluppato nei secoli una mentalità differente dall’uomo. Per questo, sono utili al cambiamento della società e alla risoluzione dei tre quesiti se venisse riconosciuto il loro diritto di accedere alle libere professioni, per guadagnarsi l’indipendenza attraverso il lavoro e così agire “lateralmente” per esercitare la loro influenza su coloro che detengono il potere: gli uomini.

1. **Conclusioni**

Con l’ausilio di alcuni strumenti computazionali e metodi della fisica, ho potuto prendere in esame due testi di circa cento mila parole e analizzarne i contenuti attraverso i dati ricavati. Comparando i due saggi che parlano dello stesso tema, scritti a più di un secolo di distanza, ho potuto notare molte somiglianze partendo dalla distribuzione di Zipf, alla densità lessicale, alle parole vuote. Le parole lessicali significative invece, presentano sostanziali differenze, nonostante la correlazione tra i due insiemi di dati sia vicina al valore 1. Le tre parole piene che ho scelto di descrivere nei dettagli del loro uso sono sostanzialmente diverse tra i due testi, perché nonostante trattino di fatto gli stessi temi, l’esposizione del pensiero delle autrici segue spunti e flussi diversi. Gli avvenimenti storici che hanno influenzato la scrittura sono diversi, che hanno ispirato racconti diversi. Vedi la parola “war” in “Three Guineas”. Occorre molte volte perché Woolf parla esplicitamente della guerra, ma non appare mai nel saggio di Mary Wollstonecraft, ovviamente perché la guerra mondiale era ben lontana dal succedere nel 1790. Un fatto interessante, a mio avviso, è che entrambi i saggi sono stati scritti in “risposta” a qualcuno. I due testi si mostrano conformi distribuzione teorica della legge di Zipf per quanto riguarda la frequenza delle parole ed è interessante compararli ottenendo il valore di correlazione. Quest’ultima, nasconde un’analisi semantica che spiega un intero universo di storie intrecciate che iniziano nel 1792 e continuano ad essere attuali ancora oggi.

1. **Bibliografia**

* Zodiaco I. (2018). “Le tre ghinee”: l’attualità del saggio femminista di Virginia Woolf
* British Library, Mary Wollstonecraft, A Vindication of the Rights of Woman
* Scobie R. (2017). An Analysis of Mary Wollstonecraft’s A Vindication of Rights of Woman
* Gättens M.L. (2001). Three Guineas, Fascism, and the Construction of Gender