

GIORGIA MELONI, MELONIO LA MELONI?

LA CODIFICA DEGLI ANTROPONIMI FEMMINILI IN BIOGRAFIE GENERATE DA CHATGPT E PUBBLICATE SU WIKIPEDIA

*Anna-Maria De Cesare*¹

1. OBIETTIVO DEL LAVORO

I testi generati automaticamente da modelli linguistici di grandi dimensioni (dall'ingl. *Large Language Models* o LLM²) basati su intelligenza artificiale, come la serie GPTn di OpenAI, costituiscono un nuovo e importante oggetto di studio della linguistica, nei suoi vari indirizzi: teorico, applicato, sociolinguistico ecc. (per alcune prime osservazioni relative all'italiano, cfr. Cicero 2023 e De Cesare 2023). Questi testi sono prodotti da sistemi addestrati su campioni estremamente ampi di dati: GPT-3.5, il sistema operativo di ChatGPT (dell'interfaccia che abbiamo usato), per esempio, è stato addestrato su un corpus di 300 miliardi di token raccolti su internet via webcrawling. Secondo Nissim & Pannitto (2022: 118), i LLM riproducono fedelmente la realtà presente nei dati su cui sono stati addestrati. Da una prospettiva linguistica, ma anche socioculturale, questi testi possono quindi essere utilizzati come specchio per rilevare le caratteristiche delle comunità linguistiche che rappresentano, inclusi i pregiudizi e le asimmetrie tra diversi gruppi sociali.

L'obiettivo generale del nostro studio è duplice: vogliamo analizzare la rappresentazione linguistica delle donne nei testi generati da ChatGPT e vogliamo capire se ci sono differenze – e quali – rispetto a testi redatti da esseri umani. Più precisamente, vogliamo indagare la presenza, la forma e la frequenza di asimmetrie linguistiche legate al genere, inteso qui in senso tradizionale, come categoria dicotomica (donna vs. uomo). Per capire come sono rappresentate le donne nei testi generati rispetto a quelli redatti da esseri umani, proponiamo un'analisi delle forme linguistiche impiegate per riferirsi a 34 personalità femminili note, indagando il loro uso in un corpus comparabile di biografie creato appositamente per lo studio. Gli aspetti da considerare sono naturalmente tanti: la forma dei sostantivi che si riferiscono alle professioni svolte dalle donne menzionate nei

¹ Technische Universität Dresden; <https://orcid.org/0000-0003-0548-2240>; <https://ror.org/042aqky30>
Ringrazio i due revisori anonimi per i preziosi suggerimenti relativi a una stesura anteriore del contributo.

² Cfr. la descrizione del neologismo fornita da Treccani ([https://www.treccani.it/vocabolario/neo-modello-linguistico-di-grandi-dimensioni_\(Neologismi\)](https://www.treccani.it/vocabolario/neo-modello-linguistico-di-grandi-dimensioni_(Neologismi)))

testi, le strategie linguistiche per rendere visibili le donne quando si fa riferimento a gruppi misti di persone, il significato degli aggettivi impiegati per descrivere i referenti femminili (rispetto a quelli maschili) ecc.³ Il presente studio indaga un altro aspetto: la forma e l’uso degli antroponimi riferiti alle donne (per approfondimenti sui nomi di persona in italiano, cfr. Marcato 2010). I tipi di antroponimi di cui terremo conto sono tre:

- I. nome + cognome: “Giorgia Meloni”
- II. solo cognome: “Meloni”
- III. articolo determinativo al sing. (*la*) + cognome: “la Meloni”

In questo studio ci concentriamo sull’uso degli antroponimi del tipo III, che confronteremo con quelli del tipo I e II. L’interesse per gli antroponimi del tipo *la Meloni* rispetto a *Meloni* si giustifica principalmente alla luce delle caratteristiche dell’italiano contemporaneo. Secondo D’Achille 2016, infatti, “l’assenza dell’articolo davanti ai cognomi, diffusa da tempo per gli uomini, viene ora spesso estesa (almeno nello scritto) anche alle donne, in rapporto alla tematica dell’uso non sessista della lingua” (p. 176)⁴. La regressione degli antroponimi femminili del tipo III si giustifica, più in particolare, in base alla presenza di un *bias* di genere: forme come *la Meloni* si usano più spesso quando ci si riferisce a una donna che a un uomo (cfr. Sabatini 1993 e Burr 1997, entrambi basati su dati tratti da testi giornalistici); ma soprattutto, essi si adoperano anche per nominare personalità femminili ben note, il cui sesso biologico non ha alcun bisogno di essere specificato.

Questo lavoro si articola in quattro parti. Nella prima (§ 2) facciamo brevemente il punto su questioni di norma relative agli antroponimi femminili del tipo *Giorgia Meloni / Meloni / la Meloni* e riportiamo alcuni dati quantitativi sull’uso delle due ultime forme nell’ambito di giornali pubblicati tra il 1977 e il 1997. Nella seconda (§ 3) presentiamo il corpus di lavoro allestito per analizzare la forma e la frequenza dei tre tipi di antroponimi femminili nell’ambito di testi biografici generati da ChatGPT e pubblicati su *Wikipedia*. A nostra conoscenza, questa tematica non è ancora stata trattata né in rapporto ai testi generati in modo automatico da LLM né alle voci biografiche di *Wikipedia*.⁵ Nella terza parte (§ 4) riportiamo i risultati della nostra analisi *corpus-based*, in particolare sul rapporto quantitativo tra i tipi II e III; proporremo anche alcune ipotesi per spiegare la frequenza d’uso di forme come *la Meloni* nelle biografie generate da ChatGPT. Nelle conclusioni (§ 5) facciamo il

³ Su questi (e molti altri) aspetti c’è un’ampissima bibliografia: si veda, a titolo puramente esemplificativo, Marcato & Thüne 2002, Gomez Gane (a c. di) 2017, Formato 2019, De Cesare 2022a e gli studi inclusi in De Cesare & Casoni 2021.

⁴ Per questo motivo, sulla scia di quanto espresso da Sabatini (1993 [1987]: 106), importanti istituzioni politiche, scientifiche e accademiche (cfr. anche la decisione comunicata dall’Accademia della Crusca sul proprio sito il 9.3.2023), raccomandano di evitare l’uso dell’art. det. davanti ai cognomi femminili.

⁵ Tra gli studi che si sono occupati della scrittura generata in modo automatico a partire da sistemi intelligenti diversi da LLM, cfr. Tivosanis 2018, 2021 e Fan & Gardent 2022 (quest’ultimo in ottica di studi di genere). Diversi studi si sono poi occupati di questione di genere in testi tradotti automaticamente tramite reti neurali (Monti 2020, Rescigno et al. 2020, Pescia 2021).

punto sui risultati ottenuti, mettendo in luce che la lingua dei testi generati da ChatGPT è molto più politicamente corretta e al passo con i tempi. Questo dato apre importanti domande sul ruolo che i testi generati potrebbero avere in futuro sull’assetto dell’italiano contemporaneo.

2. GIORGLIA MELONI, MELONI, LA MELONI: NORMA, RACCOMANDAZIONI E USO

2.1. Norma d’uso e raccomandazioni, tra grammatiche e manuali di stile

Oggi non vige più una chiara norma grammaticale in merito all’uso degli antroponomi femminili. Sfogliando le tre principali grammatiche italiane di riferimento, pubblicate alla fine degli anni Ottanta, sembra che la norma sia cominciata a cambiare proprio in quegli anni (Renzi 1988: 391-395; Schwarze 1988: 53; Serianni 2000 [1988]: 120-121⁶). Scrive infatti Serianni (2000):

con i cognomi femminili la norma tradizionale, cui è bene continuare ad attenersi, prescrive l’obbligo dell’articolo (si veda per esempio Fogarasi 1983: 169). Tuttavia, andrà notato che la tendenza attuale è verso l’uso del semplice cognome senza articolo, come per il maschile. [...] anche l’uso giornalistico attuale favorisce la soppressione dell’articolo: per esigenza di rapidità (specie nei titoli [...]) o di omogeneità (quando il cognome femminile è abbinato a un cognome maschile senza articolo [...]). Ma anche – si direbbe – per la minore importanza attribuita al sesso di una persona rispetto alla sua attività professionale o politica⁷.

I fattori che spiegano la presenza o assenza dell’art. det. davanti al cognome di una donna sono molteplici: entrano in gioco il luogo testuale in cui si manifesta l’antroponomo (titolo vs. corpo del testo), aspetti sintattici (la comparsa dell’antroponomo in un costrutto coordinativo o meno), nonché questioni storico-sociali (il ruolo che gioca il sesso biologico della donna quando si fa riferimento al mestiere che esercita o alla sua attività politica).

Oltre al contenuto delle grammatiche, è importante osservare anche quello che suggeriscono i manuali di stile, per esempio quelli ad uso dei giornalisti. Nel *Manuale di linguaggio e di stile per l’informazione scritta e parlata* pubblicato da Sergio Lepri nel 2011, l’uso dell’articolo determinativo davanti a cognome femminile è sconsigliato. Nel manuale si legge infatti:

L’articolo determinativo non si usa con i cognomi di personalità contemporanee e, a differenza di come accadeva fino a qualche anno

⁶ Per una sintesi sul punto di vista discordante delle grammatiche pubblicate tra Otto e Novecento, cfr. Burr (1997: 149-153) e Cirillo (1998: 274-276).

⁷ Lepschy & Lepschy 1981: 152. [Serianni 2000: 120.]

fa, neppure con le donne; si scrive “Bindi” non “*la* Bindi”, “Moratti” non “*la* Moratti” (ma è molto meglio con il nome, anche per riconoscere se si tratta di uomo o di donna, “Rosy Bindi”, “Letizia Moratti”)⁸.

Del passo citato è importante notare un’altra raccomandazione: l’uso dell’antroponimo del tipo ‘Nome + Cognome’ (nostro tipo I) al posto di ‘Art. det. + Cognome’ (tipo III). Le due forme avrebbero la stessa funzione semantico-pragmatica: chiarire che il referente di cui si parla è una donna. L’idea che gli antroponimi del tipo I e III siano funzionalmente equivalenti si ritrova anche in un intervento sull’uso dell’articolo prima del prenome pubblicato sul sito dell’Accademia della Crusca:

sta scomparendo l’uso di far precedere dall’articolo i cognomi di donne famose o con funzioni pubbliche: “la Deledda”, “la Moratti” tendono a essere sostituite dalle forme complete del nome proprio “Grazia Deledda”, “Letizia Moratti” che permettono di individuare il genere della persona di cui si sta parlando senza bisogno dell’articolo che, in questi casi, poteva apparire come discriminante in quanto non presente con i cognomi di uomini famosi o pubblici (Calvino, Andreotti, ecc.)⁹.

I due ultimi passi mettono in luce altri fattori che condizionano l’uso dell’art. det. davanti a cognome di donna: la vicinanza cronologica della donna e la sua notorietà.¹⁰

2.2. *Uso nella prosa giornalistica*

Anche negli studi empirici che hanno descritto l’uso degli antroponimi femminili in diversi generi testuali, in particolare nella prosa giornalistica,¹¹ c’è accordo sul fatto che il tipo *la*

⁸ Lepri 2011: 40.

⁹ Setti, 2003.

¹⁰ Secondo Korzen (1996: 605) “La funzione principale dell’articolo determinativo [con gli antroponimi] è sempre quella dei [sintagmi nominali determinati] indicanti oggetti unici, cioè il rimando dimostrativo a un’entità nota a un gruppo di persone più o meno vasto”. Si può differenziare tra persona nota (a una cerchia ristretta di persone) e famosa (nota a un’ampia fetta della popolazione). Le personalità al centro delle biografie del nostro corpus (descritto nel § 3) dovrebbero essere soprattutto del secondo tipo.

¹¹ I principali studi che descrivono l’uso degli antroponimi – anche in ottica di pari trattamento linguistico tra uomo e donna – sono Sabatini (1993: 26, 53, 106), Lepschy, Lepschy & Sanson (2001: 12), Marcato & Thüne (2002: 202-203), Setti 2003, Pescia (2010: 68-69), Gomez Gane (2017: 45-46) e Formato (2019: 66-67). Finora, a mia conoscenza, gli unici lavori *corpus-based* che offrono dati empirici sull’uso e la forma degli antroponimi sono tratti da giornali: cfr. Hall 1964, sugli antroponimi maschili del tipo II e III, e Cirillo 1998, che si occupa delle stesse forme ma riferite a donne. Osservazioni interessanti da un punto di vista diacronico sono anche in Herczeg (1972: 119-132), che si concentra *in primis* sugli antroponimi maschili, e soprattutto in Burr (1997: 146-159), che paragona varie forme di antroponimi riferiti a donne e uomini.

Meloni (III) è in regresso e che, a seguito del suo calo, sono in aumento quelli del tipo *Meloni* (II). Secondo lo studio di Cirillo 1998, condotto su un campione di testi giornalistici pubblicati in tre decenni successivi (1977-1987-1997), il cambiamento sarebbe iniziato negli anni Settanta.¹² La Tab. 1 riporta i principali risultati quantitativi ottenuti da Cirillo 1998, così come quelli di Burr 1997 in un campione di testi comparabili pubblicati nel 1989.

Data pubbl. N. occ.	Cirillo 1998		Burr 1997	Cirillo 1998
	1977 50 occ.	1987 50 occ.	1989 183 occ.	1997 100 occ.
II. cognome	10% (5 occ.)	14% (7 occ.)	66% (121 occ.)	25% (25 occ.)
III. art. det. + cognome	90% (45 occ.)	86% (43 occ.)	34% (62 occ.)	75% (75 occ.)
Fonte	Quotidiani: <i>Corriere della Sera, la Repubblica, La Stampa, L'Unità</i>		Quotidiani (nove edizioni): <i>Corriere della Sera, Il Mattino, la Repubblica, La Stampa</i>	Quotidiani (cfr. prima colonna a sinistra) e riviste (settimanali e mensili): <i>L'Espresso, Io Donna, la Repubblica delle Donne, L'indice dei Libri</i>

Tab. 1. Frequenza degli antroponimi femminili: tipi II e III

I dati di Cirillo 1998 indicano che l'uso degli antroponimi femminili del tipo III è in progressivo calo: da 90% di casi nel 1977 si passa a 75% nel 1997, mentre aumentano – proporzionalmente, perché i dati sono calcolati l'uno in rapporto all'altro – quelli del tipo II (10% nel 1977 > 25% nel 1997). Come si sarà notato, i risultati di Cirillo 1998 poggiano su dati non del tutto omogenei (il primo dato riguarda i quotidiani, il secondo quotidiani e riviste) e alquanto esigui: il campione analizzato comprende in tutto solo 250 antroponimi.¹³ I risultati ottenuti in Burr 1997 (cfr. p. 147), che riguardano solo l'annata 1989 e poggiano su un campione più ampio di antroponimi per quella data (183 occ.),

¹² Lo conferma anche l'imponente studio di Hall 1964, basato su un ampio campione di testi (appartenenti a varie tipologie) pubblicati nel 1962. Nei suoi dati, gli antroponimi femminili del tipo II sono solo sei (mentre, nei soli giornali, sono 4188 gli antroponimi maschili composti dal solo cognome: lo spoglio riguarda 11 testate giornalistiche diverse). Il commento di Hall è lampante: “When used referring to a woman, the family name is still universally accompanied by the definite article: e.g. *la Scardigno* “[Rosaria] Scardigno”, *la Deledda* “[Grazia] Deledda”, etc.” (p. 166).

¹³ L'autrice giustifica l'esiguità del campione in base alla difficoltà di riscontrare antroponimi riferiti alle donne “in order to find 250 occurrences of women's surnames we had to look at more than thirty issues of newspapers, magazines and periodicals. Besides, some occurrences were taken from the same article, and very often the same character was the protagonist of many articles” (Cirillo 1998: 287).

sono diversi: ammontano a 34% i cognomi femminili del tipo III e a 66% quelli del tipo II.

Le osservazioni da fare sui risultati degli studi di Burr 1997 e Cirillo 1998 sono sostanzialmente queste: a livello numerico, importa innanzitutto osservare che, a distanza di due decenni, il calo nell’uso degli antroponimi del tipo III è di 15 punti percentuali. Nella prosa giornalistica della fine degli anni Ottanta, c’è però un divario importante tra il dato al quale arriva Cirillo e quello ottenuto da Burr 1997: gli antroponimi del tipo III sono due volte più numerosi in Cirillo (86% vs. 34%), anche se entrambi gli studi si basano su un campione molto simile di quotidiani, pubblicati a soli due anni di distanza (1989 e 1987). Da notare ancora che i due studi non danno indicazioni sulla frequenza degli antroponimi nelle singole testate giornalistiche (tanto meno nelle loro rubriche) e non divulgano la lista dei cognomi riscontrati. Non sappiamo dunque quanti cognomi si riferiscono a personalità femminili note o famose (secondo le osservazioni proposte nella nota 7) e se i dati relativi a diverse annate e/o quotidiani riguardano gli stessi referenti (cf. nota 10). Infine, per spiegare il dato relativo agli antroponimi del tipo III i due studi non tengono conto di altre forme, in particolare di ‘Nome + Cognome’ (tipo I).

3. CORPUS DI LAVORO

3.1. *Campione di testi raccolti per il corpus BioCGPT/BioWiki*

Per capire quali e quanti antroponimi femminili sono usati nei testi generati da LLM e in quelli scritti da essere umani, abbiamo creato un corpus di testi comparabili appartenente al genere delle biografie. Il corpus si compone di due parti: BioCGPT e BioWiki.¹⁴

Il sottocorpus BioCGPT è stato creato con la versione gratuita di ChatGPT, disponibile online dal 30 novembre 2022 sul sito di OpenAI. ChatGPT è “un chatbot basato su un grande modello linguistico [LLM]” (ChatGPT, *Wikipedia*¹⁵). Il LLM su cui si basa ChatGPT (ovvero GPT-3.5) è stato addestrato su un campione molto ampio di testi raccolti su Internet tramite webcrawling. Questi testi, pubblicati in rete fino a settembre 2021, includono voci di *Wikipedia* (3%), libri, giornali, riviste, blog ecc. Come indica il nome dell’interfaccia, ChatGPT è stato progettato per essere utilizzato principalmente in modalità dialogica, in sequenze di coppie “richiesta dell’utente-risposta di ChatGPT”. Il corpus BioCGPT comprende però testi generati con un singolo “prompt”: i testi sono quindi da considerarsi monologici piuttosto che dialogici.

Il sottocorpus BioCGPT comprende 34 biografie femminili generate da ChatGPT-3.5 in italiano. Queste biografie sono state generate individualmente tra il 7.1.2023 e il 14.6.2023 e soddisfano due condizioni: (i) riguardano personalità femminili note o famose nate (per lo più) nel Novecento; si tratta di personalità attive in diversi campi (scienza, arte, spettacolo ecc.) e (ii) sono anche disponibili su *Wikipedia*. Tutti i testi sono stati generati

¹⁴ Un sentito grazie a Claudia Rausch e Tom Weidensdorfer per il loro aiuto nella raccolta dei testi che fondano il corpus, creato e interrogato da chi scrive sulla piattaforma Sketch Engine.

¹⁵ https://en.Wikipedia.org/w/index.php?title=ChatGPT&oldid=1175956865#Use_and_implications

manualmente utilizzando lo stesso prompt: “Scrivi un testo di 1000 parole per spiegare chi è x”, dove x corrisponde a una variabile che coincide con il nome e il cognome di una donna, ad esempio Dacia Maraini. Un esempio di prompt e dell’output generato è riportato nella Fig. 1 dell’Appendice. Ogni biografia femminile è stata generata almeno quattro volte (in date diverse), sia per verificare una possibile evoluzione dell’output testuale del LLM sia per confrontare i risultati ottenuti all’interno della stessa biografia. Il sottocorpus BioWiki, a sua volta, è stato creato con l’aiuto di Trafilatura, un software in grado di estrarre automaticamente testi dalla rete (Barbaresi 2021). Abbiamo raccolto il contenuto delle 34 voci femminili di *Wikipedia*¹⁶ menzionati al punto (ii) sopra e lo abbiamo salvato in un unico file txt. Per garantire la massima comparabilità con i testi generati, abbiamo scartato dal file diversi contenuti tipici delle entrate di *Wikipedia*, come le note finali, la filmografia o l’elenco completo delle opere della personalità descritta nella voce, i riferimenti e i link esterni.

3.2. BioCGPT/BioWiki: Corpus design

Il corpus BioCGPT/BioWiki creato per questo studio comprende biografie generate in italiano da ChatGPT-3.5 su 34 donne diverse (BioCGPT) e le corrispondenti 34 biografie disponibili in italiano su *Wikipedia* (BioWiki). La *Tab.2* mostra la composizione generale del corpus.

BioCGPT	BioWiki
34 biografie di donne italiane	34 biografie di donne italiane

Tab.2. Corpus BioCGPT e BioWiki

Il sottocorpus BioCGPT comprende un totale di 168 testi (come indicato nel § 3.1, ogni biografia femminile è stata generata almeno quattro volte), mentre il sottocorpus BioWiki include esclusivamente il testo delle 34 biografie menzionate nella *Tab.2*. La *Tab.3* indica le principali proprietà quantitative dei due sottocorpora utilizzati in questo studio.

	BioCGPT	BioWiki
Testi	168	34
Token	75,932	84,351
Parole	65,624	68,192

Tab.3. Corpus BioCGPT/BioWiki (dati calcolati da Sketch Engine)

Come si può osservare dalla *Tab.3*, i due sottocorpora includono un numero relativamente simile di token (vale a dire unità grafiche autonome: parole, segni di punteggiatura, simboli ecc.) e un numero relativamente vicino di parole, anche se BioCGPT include 168 testi e

¹⁶ Per una descrizione relativa alla produzione delle voci di *Wikipedia*, cfr. Tavosanis 2021, con cenni alla generazione automatica (mediante un *template*) della prima frase delle voci di tipo biografico (pp. 426-427).

BioWiki solo 34. Questo risultato è dovuto al fatto che le biografie generate sono generalmente più brevi di quelle disponibili su *Wikipedia*. La biografia di Dacia Maraini riprodotta nelle *Fig.1* e *Fig.2* (cfr. Appendice) permette di cogliere la differenza di lunghezza: il testo generato comprende 339 parole, mentre la voce di *Wikipedia* ne comprende 984 (non teniamo conto dell’elenco delle opere e dei riferimenti finali). Si noti che la *Fig.1* riproduce l’intero testo generato, mentre la *Fig.2* ne mostra solo una parte. La *Fig.1* permette anche di evidenziare che l’output di ChatGPT risponde solo in parte alla nostra richiesta. Un conteggio del numero di parole presente in ogni testo di BioCGPT rivela che l’output generato non supera mai le 500 parole (non di rado, come si può osservare alla fine della *Fig.1*, il testo si interrompe anche durante la generazione, lasciando l’ultima frase incompleta): i testi più lunghi comprendono 489 parole. Ma il prompt non è stato seguito in merito a un altro punto importante: un testo (su 168) è stato generato in inglese nonostante il prompt fosse scritto in italiano (si tratta di un testo su Anna Magnani, generato l’11.1.2023).

4. CODIFICA DEGLI ANTROPONOMI NEL CORPUS BIOCGPT/BIOWIKI

4.1. *Giorgia Meloni / Meloni / la Meloni. Risultati quantitativi*

Nella *Tab. 4* riportiamo i dati quantitativi relativi alle forme di antroponomo del tipo I, II e III nei due sottocorpora del corpus BioCGPT/BioWiki.

	BioCGPT	BioWiki
Tipo I. Nome + Cognome	537 occ. (46%)	308 occ. (57%)
Tipo II. Cognome	553 occ. (48%)	126 occ. (24%)
Tipo III. Art. det. + Cognome	67 occ. (6%)	103 occ. (19%)
Tot.	1157 occ.	537 occ.

Tab.4. Forme di antroponomi femminili: frequenza assoluta e relativa (in percentuale)

Le tre forme di antroponomo sono presenti in entrambi i sottocorpora, ma con frequenze d’uso (assolute e relative) molto diverse. In BioCGPT gli antroponomi più frequenti sono quelli del tipo I e II (il dato medio corrisponde 47%), mentre quelli del tipo III sono marginali (6%). In BioWiki, invece, gli antroponomi di gran lunga più rappresentati sono quelli del primo tipo (57%), seguiti da quelli del secondo (24%) e terzo tipo (19%). Gli antroponomi del tipo III sono tre volte meno frequenti nel sottocorpus BioCGPT (6%) che in quello BioWiki (19%). Di seguito forniamo alcuni esempi delle tre forme di antroponomo presenti nei due sottocorpora:

- (1) **Eleonora Duse** (1858-1924) è stata un’attrice teatrale italiana, considerata una delle più grandi attrici della storia del teatro. Nata a Vigevano, in Lombardia, **Duse** iniziò la sua carriera come attrice a soli 16 anni, recitando in piccole compagnie teatrali in giro per l’Italia. Nel 1881, entrò a far parte della compagnia del famoso attore Ettore Scola, con il quale recitò in molte produzioni di successo. (Eleonora Duse, BioCGPT)

- (2) **La Duse** fu molto influenzata dal lavoro di alcuni importanti registi e attori del suo tempo, tra cui André Antoine e Konstantin Stanislavskij. Fu proprio quest'ultimo a rivelarsi particolarmente importante nella sua vita artistica: quando **la Duse** incontrò per la prima volta Stanislavskij, nell'estate del 1898, l'incontro cambiò radicalmente il suo modo di recitare. **La Duse**, infatti, fu una delle prime attrici a comprendere l'importanza del metodo Stanislavskij per la creazione di personaggi credibili e realistici sul palcoscenico. (Eleonora Duse, BioCGPT)
- (3) È proprio in questo periodo, gli anni ottanta, che **Eleonora Duse** compirà le scelte di repertorio che segneranno il suo percorso artistico e la sua carriera. [...] Spesso, sia sul palcoscenico sia fuori dalla scena, **Duse** congiungeva le mani e intrecciava le dita, poi sollevava gli indici e li conduceva alla bocca, sfiorandosi le labbra in un'attitudine pensosa e molto delicata. [...] **La Duse** ispirò una parte molto importante dell'opera dannunziana, tale da essere la musa ispiratrice della raccolta poetica *Alcyone*, la più celebre delle raccolte poetiche dannunziane. (Eleonora Duse, *Wikipedia*¹⁷)

Gli antroponimi del tipo I sono usati in modo sistematico la prima volta che ci si riferisce alla personalità al centro della biografia. Questo si osserva sia nelle biografie generate (cf. l'es. 4, oltre a quello proposto al punto 1) sia in quelle pubblicate su *Wikipedia* (5):

- (4) **Liliana Segre** è una figura storica e un simbolo di coraggio, resilienza e impegno nella difesa dei diritti umani. (Liliana Segre, BioCGPT)
- (5) **Liliana Segre** (Milano, 10 settembre 1930) è un'antifascista e politica italiana, superstite dell'Olocausto e testimone attiva della Shoah (Liliana Segre, BioWiki¹⁸)

4.2. *Meloni / la Meloni: frequenza d'uso*

In quanto segue ci concentriamo sul rapporto tra le forme di antroponimo del tipo *Meloni* (II) e *la Meloni* (III). La loro frequenza reciproca è indicata nella *Tab.5*.

	BioCGPT	BioWiki
Tipo II. Cognome	89% (553 occ.)	55% (126 occ.)
Tipo III. Art. det. + Cognome	11% (67 occ.)	45% (103 occ.)
Tot.	100% (620 occ.)	100% (229 occ.)

Tab.5. Forme di antroponimi femminili: frequenza reciproca sia assoluta sia relativa

¹⁷ https://it.wikipedia.org/w/index.php?title=Eleonora_Duse&oldid=134989604

¹⁸ https://it.wikipedia.org/w/index.php?title=Liliana_Segre&oldid=135632041

Gli antroponimi del tipo *Meloni* (II) sono più frequenti in BioCGPT che in BioWiki (89% vs. 55%), mentre vale il contrario per quelli del tipo III, *la Meloni* (11% vs. 45%). Nelle biografie generate gli antroponimi del tipo III si riscontrano per 10 donne su 34 (ca. 29%), mentre in BioWiki si trovano per 24 donne su 34 (71%).

Per capire meglio l’uso degli antroponimi del tipo II e III nei dati raccolti è utile osservare i risultati relativi alle singole 34 personalità femminili. Nelle *Tabb.6-8* forniamo dunque i risultati relativi a tre casi diversi: la *Tab.6* propone i dati quantitativi relativi ai 10 antroponimi femminili il cui cognome è accompagnato dall’art. det. nei sottocorpus BioCGPT e i dati relativi alle stesse 10 personalità in BioWiki; la *Tab.7* propone i dati relativi ai casi in cui l’art. det. davanti al cognome si riscontra solo in BioWiki e la *Tab.8* quelli in cui l’art. det. davanti al cognome non si riscontra né in BioCGPT né in BioWiki.

	BioCGPT		BioWiki	
	Freq. ass. III / II	% III	Freq. ass. III / II	% III
1. la Merini / Merini (1931-2009)	8 / 0	100	1 / 0	100
2. la Loren / Loren (1934-)	2 / 0	100	12 / 1	92
3. la Cortellesi / Cortellesi (1973-)	17 / 0	100	1 / 6	14
4. la Hack / Hack (1922-2013)	13 / 7	65	1 / 6	14
5. la Ferrante / Ferrante (1943-)	7 / 18	28	0 / 5	0
6. la Duse / Duse (1858-1924)	8 / 22	27	11 / 2	85
7. la Fallaci / Fallaci (1929-2006)	4 / 16	20	13 / 2	87
8. la Maraini / Maraini (1936-)	6 / 26	19	0 / 3	0
9. la Brandeis / Brandeis (1848-1926)	1 / 13	7	3 / 8	27
10. la Deledda / Deledda (1871-1936)	1 / 15	6	2 / 30	6
Totale freq. assoluta e %	67 / 117	36	44 / 63	41

Tab.6. Forme di antroponimi: tipo III / tipo II e % del tipo III

I dati nell’ultimo rigo della *Tab.6* permettono di osservare che, presi nella loro globalità, i dieci antroponimi femminili del tipo III sono leggermente più numerosi in BioWiki (41%) che in BioCGPT (36%). Scendendo nel dettaglio, si può però constatare che in 5 biografie su 10 (Loren, Cortellesi, Hack, Ferrante, Maraini), la percentuale degli antroponimi del tipo III è più alta in BioCGPT, con scarti a volte molto importanti, come nel caso di Cortellesi (100% vs. 14%); a volte trascurabili, come in quello di Loren (100% vs. 92%); in una biografia, la percentuale è identica in entrambi i sottocorpora (si tratta di Merini, alla quale ci si riferisce sempre ed esclusivamente con ‘art. det. + cognome’), mentre nelle altre quattro è più alta in BioWiki: anche qui con scarti a volte molto ampi, come nel caso di Duse (27% vs. 85%).

In BioCGPT vi sono poi 15 cognomi femminili che non sono mai preceduti dall’art. det., a differenza di ciò che si osserva in BioWiki. I dati relativi a BioWiki proposti nella *Tab.7* mostrano anche che in ben 8 casi su 15 il cognome è sempre preceduto dall’art. *la*.

	BioCGPT		BioWiki	
	Freq. ass. III / II	% III	Freq. ass. III / II	% III
11. la Montessori / Montessori (1870-1952)	0 / 18	0	22 / 0	100
12. la Magnani / Magnani (1908-1973)	0 / 14	0	14 / 0	100
13. la Ginzburg / Ginzburg (1916-1991)	0 / 29	0	2 / 0	100
14. la Ercoli-Finzi / Ercoli-Finzi (1937-)	0 / 22	0	2 / 0	100
15. la Cavaglieri Tesoro / Cavaglieri Tesoro (1921-2002)	0 / 0	0	2 / 0	100
16. la Fini / Fini (1907-1996)	0 / 41	0	1 / 0	100
17. la Spaziani / Spaziani (1922-2014)	0 / 15	0	1 / 0	100
18. la Cattaneo / Cattaneo (1962-)	0 / 18	0	1 / 0	100
19. la Ortese / Ortese (1914-1998)	0 / 22	0	2 / 1	67
20. la Cristoforetti / Cristoforetti (1977-)	0 / 27	0	1 / 1	50
21. la Aulenti / Aulenti (1927-2012)	0 / 13	0	1 / 1	50
22. la Bindi / Bindi (1951-)	0 / 13	0	6 / 10	38
23. la Morante / Morante (1912-1985)	0 / 35	0	1 / 3	25
24. la Iotti / Iotti (1920-1999)	0 / 0	0	1 / 3	25
25. la Boldrini / Boldrini (1961-)	0 / 26	0	2 ¹⁹ / 11	15
Totale freq. assoluta e %	0 / 300	0	59 / 30	66

Tab.7. Forme di antroponimi: tipo III / tipo II e % del tipo III

Questa volta la differenza tra i dati delle 15 biografie presenti in BioCGPT e BioWiki è molto importante: globalmente (cf. ultimo rigo), gli antroponimi del tipo III coprono il 66% dei casi riscontrati in BioWiki, mentre non sono mai documentati in BioCGPT. Uno sguardo più dettagliato ai dati della Tab. 6 permette poi di cogliere scarti quantitativi molto

¹⁹ Abbiamo escluso un caso reperito nell’ambito del discorso riportato diretto: “Altrettanto vaste sono le reazioni politiche quando il leader della Lega, Matteo Salvini, durante un comizio a Soncino il 25 luglio 2016, espone sul palco una bambola gonfiabile accompagnando questo gesto con la frase: "C'è una sosia della Boldrini qui sul palco. Non so se sia già stata esibita..." [...]” (https://it.Wikipedia.org/w/index.php?title=Laura_Boldrini&oldid=134941309).

marcati in alcune biografie: è il caso di Maria Montessori, una delle figure cronologicamente più lontane presenti nel nostro corpus, e di Anna Magnani. Nei testi di BioCGPT l’assenza del determinante davanti ad alcuni cognomi femminili (*Montessori / Magnani*) è un dato davvero notevole.

Nella *Tab.8* riportiamo infine i dati dei 9 cognomi femminili che non compaiono mai con l’art. det., né in BioCGPT né in BioWiki.

	BioCGPT		BioWiki	
	Freq. ass. III / II	% III	Freq. ass. III / II	% III
26. la Meloni / Meloni (1977-)	0 / 39	0	0 / 23	0
27. la Levi-Montalcini / Levi-Montalcini (1909-2012)	0 / 13	0	0 / 8	0
28. la Marabini / Marabini (1930-)	0 / 14	0	0 / 1	0
29. la Rossanda / Rossanda (1924-2020)	0 / 35	0	0 / 1	0
30. la Segre / Segre (1930-)	0 / 8	0	0 / 0	0
31. la Strada / Strada (1884-1970)	0 / 0	0	0 / 0	0
32. la Lollini / Lollini (1888-1987)	0 / 8	0	0 / 0	0
33. la Gianotti / Gianotti (1960)	0 / 21	0	0 / 0	0
34. la Bo Bardi / Bo Bardi (1914-1992)	0 / 10	0	0 / 0	0
Totale freq. assoluta e %	0 / 148	0	0 / 33	0

Tab.8. Forme di antroponimi: tipo III / tipo II e % del tipo III

I dati della *Tab.8* mostrano che in entrambi i sottocorpora la maggior parte di questi antroponimi sono codificati come tipo II: si veda per esempio il caso di *Meloni*. In un numero non irrilevante di casi, però, non si riscontrano antroponimi né del tipo II né del tipo III. In BioCGPT è solo il caso di Strada, ma in BioWiki si tratta di cinque donne diverse (Segre, Strada, Lollini, Gianotti, Bo Bardi). Nei due sottocorpora ci si riferisce a queste donne con la forma di antroponimo del tipo I (‘Nome + Cognome’).

4.3. *Alcune ipotesi sui risultati ottenuti*

Nelle biografie generate, più che la presenza dell’art. det. davanti ai cognomi femminili spicca il dato quantitativo relativo alla sua assenza. In BioCGPT predominano infatti largamente i casi in cui l’art. det. non c’è: in 20 profili femminili su 34 non c’è traccia di antroponimo del tipo *la Meloni*. Un altro dato che mostra in modo chiaro la bassa frequenza degli antroponimi del tipo *la Meloni* rispetto a *Meloni* sono le percentuali: gli antroponimi del tipo III coprono 11% dei casi, quelli del tipo II, l’89%. Molto diversi sono invece i risultati in BioWiki, dove gli antroponimi del tipo *la Meloni* si riscontrano in ben 24 profili femminili su 34 e dove il dato relativo al tipo III corrisponde al 45%, mentre

quello del tipo II equivale al 55%. La differenza tra le biografie generate da ChatGPT e quelle pubblicate su *Wikipedia* è dunque macroscopica.

Le ipotesi che spiegano, a nostro avviso, perché le biografie generate includono una percentuale così elevata di casi in cui l’art. det. non è presente davanti a un cognome femminile sono molteplici e non si escludono a vicenda. I risultati si spiegano innanzitutto in base ai testi italiani sui quali GPT-3.5 è stato addestrato. L’ipotesi sarebbe che, in questi testi, i cognomi femminili del tipo III sono meno frequenti di quelli del tipo II; questa ipotesi, però, non può essere verificata perché non abbiamo accesso ai dati con i quali il LLM è stato addestrato. Questo sarà vero soprattutto nei testi redatti negli ultimi decenni (dagli anni Ottanta in poi; su questo punto, cfr. § 2).

Un altro fattore che può spiegare la bassa frequenza dell’art. det. davanti ai cognomi femminili è la sua assenza davanti ai cognomi maschili. Nei testi usati per allenare GPT-3.5, i casi in cui manca l’art. det. davanti al cognome maschile sono sicuramente molto numerosi. A questo si aggiunge il fatto che in molte tipologie testuali i referenti maschili sono più numerosi di quelli femminili. Nei media scritti, per esempio, le donne sono chiaramente sottorappresentate: il dato riportato da Burr è del 9%²⁰ mentre è del 30% nell’ultimo rapporto “Who makes the News. Europe. Regional Report 2020” (pubblicato ogni cinque anni).²¹ Importa qui anche notare che nei dati di Burr 1997 i cognomi maschili non sono praticamente mai preceduti dall’art. det.

I risultati ottenuti si spiegano però anche in base ai testi inglesi sui quali è stato addestrato GPT-3.5: gli LLM sono allenati simultaneamente con campioni di testi redatti in diverse lingue (inglese, francese, tedesco, italiano ecc.).²² L’inglese è tuttavia largamente predominante nei dati di allenamento perché è la lingua più usata in rete, da dove i testi sono prelevati (Ferrara 2023: 5). La maggiore presenza della lingua inglese nei dati di addestramento di GPT-3.5 potrebbe avere delle ricadute sulla generazione di testi redatti in lingue diverse da quella inglese. Nel caso di cui ci occupiamo si potrebbe ipotizzare che l’alta percentuale di antroponimi femminili del tipo II e la minore percentuale di quelli del tipo III nelle biografie generate in italiano sia (anche) dovuta alla riproduzione dei tipi di antroponimi più frequenti in inglese. In questa lingua, gli antroponimi femminili e maschili sono infatti del tipo I o II. Quelli del tipo III sono possibili ma sono circoscritti ad alcuni contesti specifici. Il loro uso dipende da fattori sintattici e pragmatici (non gioca invece nessun ruolo il sesso del referente): l’art. det. (naturalmente neutro per quanto riguarda il genere) compare quando il cognome è accompagnato da un modificatore, come in *the Ferrante pseudonym* ‘lo pseudonimo di Ferrante’ (**Ferrante pseudonym*) e *the Ferrante letters* ‘le lettere di Ferrante’ (**Ferrante letters*), o quando si vuole sottolineare l’unicità ed

²⁰ Nei testi giornalistici indagati da Burr 1997 (pubblicati, come abbiamo visto, nel 1989), 91% degli antroponimi (12383 occ.) si riferiscono a uomini e solo 9% (1229 occ.) a donne (cfr. pp. 142-143).

²¹ <https://whomakesthenews.org/wp-content/uploads/2021/07/Europe-Region.-GMMP-report.pdf>

²² Dr. Sunna Torge (c.p.). Una chiara spia della presenza dell’inglese sono le biografie generate in lingua inglese a partire da un prompt redatto in italiano (cfr. il caso riferito alla fine del § 3.2).

eccezionalità di una persona, come in *I met THE Ferrante* ‘Ho incontrato ART. DET. Ferrante’ (con enfasi sull’articolo)²³.

5. CONCLUSIONI

I risultati ottenuti in questo studio indicano che ci sono differenze sostanziali nel modo in cui i testi generati da ChatGPT (modello GPT-3.5) e testi simili redatti da esseri umani si riferiscono alle donne. Abbiamo visto che le forme linguistiche impiegate per codificare gli antroponimi riferiti alle donne sono le stesse nelle biografie generate da ChatGPT e in quelle pubblicate su *Wikipedia*: in entrambi i casi troviamo forme come *Giorgia Meloni*, *Meloni* e *la Meloni*. Quello che cambia è la loro frequenza d’uso. Le biografie femminili generate in italiano usano soprattutto antroponimi del tipo *Giorgia Meloni* e *Meloni*, mentre quelle disponibili su *Wikipedia* presentano un netto sbilanciamento verso il primo tipo. In entrambi i casi, gli antroponimi del tipo *la Meloni* sono quelli meno rappresentati nei dati. Uno dei dati più interessanti emersi in questo studio riguarda la diversa frequenza d’uso degli antroponimi del tipo *la Meloni* nei due sottocorpora analizzati. La bassa frequenza di questa forma di antroponimo nelle biografie generate da ChatGPT (11%) rispetto a quelle che si leggono su *Wikipedia* (45%) è molto significativa e apre importanti domande, in particolare in merito ai fattori che giocano un ruolo nella spiegazione dei dati riscontrati. I risultati ottenuti per i testi generati dovrebbero spiegarsi in prima istanza in base ai dati sui quali è stato addestrato il sistema operativo di ChatGPT. Come indicato nel § 1, l’idea è infatti che i *pattern* più frequenti dei testi “sottostanti” siano riprodotti nei testi generati. La bassa occorrenza di forme come *la Meloni* nei testi generati dovrebbe dunque essere il riflesso della sua bassa frequenza nei testi in italiano sui quali GPT-3.5 è stato addestrato. A incidere sulla frequenza degli antroponimi del tipo ‘art. det. + cognome femminile’ potrebbero però esserci altri fattori, come la forma degli antroponimi italiani riferiti agli uomini e la forma degli antroponimi usati più di frequente in inglese, indipendentemente dal sesso del referente. Insomma, il dato che si ricava sulla frequenza degli antroponimi femminili come *la Meloni* nei testi generati potrebbe essere in buona parte da attribuire a fattori indiretti: l’influsso degli antroponimi maschili e la codifica degli antroponimi in lingua inglese. Indipendentemente dalle ragioni che spiegano la bassa frequenza degli antroponimi come *la Meloni* nelle biografie generate, i testi di ChatGPT-3.5 analizzati in questa sede si configurano come molto più politicamente corretti e al passo con i tempi di quelli presenti su *Wikipedia*.

La differenza riscontrata in questa sede tra i testi generati e quelli scritti da esseri umani è di primaria importanza per l’assetto dell’italiano dei prossimi decenni. Al momento (sul

²³ È interessante a questo proposito notare che nelle traduzioni di testi giornalistici dall’inglese (ma non solo) all’italiano, l’articolo davanti a cognome femminile “è la prassi consolidata [...], con rarissime eccezioni (“con Brambilla o con Dell’Utri?” si chiede Furio Colombo sull’Unità)” (Ondelli & Viale 2010: 31). Nel passaggio dall’inglese all’italiano, l’art. det. è dunque aggiunto in modo pressoché sistematico, il che indica quanto sia percepito come necessario. Si tratta dunque di una norma (implicita) fortemente sentita.

Anna-Maria De Cesare, ‘Giorgia Meloni, Meloni, la Meloni?’ *La codifica degli antroponimi in biografie generate da ChatGPT e pubblicate su Wikipedia*.

finire del 2023) la produzione e circolazione dei testi generati è ancora circoscritta²⁴, ma essa è sicuramente destinata ad aumentare rapidamente nei prossimi anni. In futuro non è dunque esclusa la possibilità che una maggiore esposizione ai testi generati possa provocare cambiamenti nell’uso linguistico, corretti o scorretti che ci appaiano ora, anche sotto la spinta dell’inglese. Nel caso indagato in questa sede, si potrebbe immaginare una nuova spinta verso il regresso dell’articolo determinativo davanti ai cognomi femminili. Indagare le caratteristiche linguistiche (ma anche testuali, pragmatiche ecc.) dei testi generati da LLM, mettendo in luce i loro tratti salienti rispetto ai testi redatti da esseri umani, può dunque aiutarci a capire meglio quali potrebbero essere alcuni possibili futuri cambiamenti nell’uso linguistico dell’italiano standard e neo-standard.

RIFERIMENTI BIBLIOGRAFICI

- Accademia della Crusca (2023), “L’Accademia risponde a un quesito sulla parità di genere negli atti giudiziari posto dal Comitato Pari opportunità del Consiglio direttivo della Corte di Cassazione, Accademia della Crusca”, <https://accademiadellacrusca.it/Media?c=54099131-d9f9-4f94-b999-e7aa4646b317> [consultato il 13.8.2023]
- Barbarese A. (2021), “Trafilatura: A Web Scraping Library and Command-Line Tool for Text Discovery and Extraction”, in *Proceedings of ACL/IJCNLP 2021: System Demonstrations*, pp. 122-131.
- Burr E. (1997), “Neutral oder Stereotyp. Referenz auf Frauen und Männer in der italienischen Tagespresse”, in Dahmen W. et al. (a c. di), *Sprache und Geschlecht in der Romania. Romanistisches Kolloquium X.*, Tübingen, Narr, pp. 133-179.
- Cicero F. (2023), “L’italiano delle intelligenze artificiali generative”, in *Italiano LingueDue* 15/2: 733-761.
- Cirillo C. (1998), “Corti or la Corti? Definite article + surnames for women”, in *The Italianist* 18: 272-288.
- D’Achille P. (2016), “Architettura dell’italiano di oggi e linee di tendenza”, in Lubello S. (a c. di), *Manuale di linguistica italiana*, De Gruyter, Berlin, pp. 165-189.
- De Cesare A.-M. (2022a), “La codifica linguistica dei referenti umani nella Costituzione svizzera: tra disparità e uguaglianza di genere”, in Ferrari A., Lala L. & Pecorari F. (a c. di), *L’italiano dei testi costituzionali. Indagini linguistiche e testuali tra Svizzera e Italia*, Edizioni dell’Orso, Alessandria, pp. 245-270.

²⁴ I testi generati in modo automatico usati al momento nei media italiani sono prodotti a partire da un testo predefinito (detto *template*) contenente dei vuoti (*gap*) riempiti da un algoritmo che estrae le informazioni da integrare da una base strutturata di dati. Per approfondimenti, in particolare sulla lingua e testualità di questi testi, cfr. De Cesare, 2022b e De Cesare, Eliasson & Weidensdorfer, 2023.

Anna-Maria De Cesare, ‘Giorgia Meloni, Meloni, la Meloni ?’. *La codifica degli antroponimi in biografie generate da ChatGPT e pubblicate su Wikipedia*.

- De Cesare A.-M. (2022b), “Répétitions et variations des textes générés: Une analyse linguistique basée sur un corpus d’articles financiers rédigés en français”, in *CHIMERA: Romance Corpora and Linguistic Studies* 8, 79–108.
- De Cesare A.-M. (2023), “Assessing the quality of ChatGPT’s generated output in light of human-written texts. A corpus study based on textual parameters”, in *CHIMERA. Romance Corpora and Linguistic studies* 10: 179-210.
- De Cesare A.-M. & Casoni M. (a c. di), (2021). *La rappresentazione delle donne nelle pratiche didattiche, nei discorsi, nelle lingue*, N. speciale di *Babylonia* 3/2021.
<https://babylonia.online/index.php/babylonia/issue/view/16>
- De Cesare A.-M., Eliasson E. & Weidensdorfer T. (2023), “La coesione testuale (basata sulle strutture verbali) nella scrittura generata in ambito finanziario. Italiano e francese a confronto”, in De Cesare A.-M., Ferrari A. & Pecorari F., Lala L. (a c. di), *Forme della scrittura italiana contemporanea in prospettiva contrastiva. La componente testuale*, Cesati, Firenze, pp. 71-86.
- Fan A. & Gardent C. (2022), *Generating Full Length Wikipedia Biographies. The Impact of Gender Bias on the Retrieval-Based Generation of Women Biographies*.
<https://arxiv.org/abs/2204.05879>
- Ferrara E. (2023), “Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models”, in *First Monday* 28/11, pp. 1-39.
<https://doi.org/10.5210/fm.v28i11.13346>
- Fogarasi M. (1983), *Grammatica italiana del Novecento. Sistemazione descrittiva*, Bulzoni, Roma.
- Formato F. (2019), *Gender, Discourse and Ideology in Italian*, Mcmillan, Palgrave.
- Gomez Gane Y. (a c. di) (2017), «*Quasi una rivoluzione*». *I femminili di professioni e cariche in Italia e all'estero*, Accademia della Crusca, Firenze.
- Hall R. A. (1964), “Moot points in Italian Grammar. 3. Definite Article + Family Name (masc.)”, in *Italica* 41/2, pp. 162-167.
- Herczeg G. H. (1972), “Contributi all’uso dell’articolo determinativo in italiano”, in *Acta Linguistica Academiae Scientiarum Hungaricae* 22: 119-39.
- Korzen I. (1996), *L’articolo italiano fra concetto ed entità*. Vol. I-II, Museum Tusulanum Press, Copenhagen.
- Lepri S. (2011), *News. Manuale di linguaggio e di stile per l’informazione scritta e parlata*, Rizzoli Etas, Milano.
- Lepschy A. L. & Lepschy G. (1981), *La lingua italiana*, Bompiani, Milano.
- Lepschy A. L., Lepschy G. & Sanson, H. (2001), “Lingua italiana e femminile”, in *Quaderns d’Italià* 6: 9-18.
- Marcato G. (2010), “Antroponimi”, in Simone R. (dir.), *Enciclopedia dell’italiano*, Treccani, Roma, pp. 84-87.
- Marcato G. & Thüne E.-M. (2002), “Italian. Gender and female visibility in Italian”, in Hellinger M. & Bußmann H. (eds), *Gender Across Languages. The Linguistic Representation of Women and Men*. Vol. 2, Benjamins, Amsterdam-Philadelphia, pp. 187-217.

Anna-Maria De Cesare, ‘Giorgia Meloni, Meloni, la Meloni?’ *La codifica degli antroponimi in biografie generate da ChatGPT e pubblicate su Wikipedia*.

- Monti J. (2020), “Gender issues in Machine Translation: an unsolved problem?” in von Flotow L. & Kamal H. (a c. di), *Routledge Handbook on Translation, Feminism and Gender*, London, Routledge.
- Nissim M. & Pannitto L. (2022), *Che cos'è la linguistica computazionale?* Carocci, Roma.
- Ondelli S. & Viale M. (2010), “L’assetto dell’italiano delle traduzioni in un corpus giornalistico. Aspetti qualitativi e quantitativi”, in *Rivista internazionale di tecnica della traduzione* 12: 1-62.
- Pescia L. (2010), “Il maschile e il femminile nella stampa scritta del Canton Ticino (Svizzera) e dell’Italia”, in Sapegno M. S. (a c. di), *Che genere di lingua? Sessismo e potere discriminatorio delle parole*, Carocci, Roma, pp. 57-74.
- Pescia L. (2021), “La femminilizzazione degli agentivi nell’era digitale: la rappresentazione linguistica delle donne e google translate”, in De Cesare A.-M. & Casoni M. (a c. di), *La rappresentazione delle donne nelle pratiche didattiche, nei discorsi, nelle lingue*, N. speciale di *Babylonia* 3/2021, pp. 102-109.
<https://doi.org/10.55393/babylonia.v3i.133>
- Renzi L. (1988), “L’articolo”, in Renzi L. (a c. di), *Grande grammatica italiana di consultazione*, vol. 1, il Mulino, Bologna, pp. 357-423.
- Rescigno A. A., Monti J. Way A. & Vanmassenhove E. (2020), “A case study of natural gender phenomena in translation. A comparison of Google translate, Bing Microsoft Translator and DeepL for English to Italian, French and Spanish”, in *Association for Machine Translation in the Americas*, pp. 62-90.
- Sabatini A. (1993), *Il sessismo nella lingua italiana. Commissione Nazionale per la parità e le pari opportunità tra uomo e donna*, Presidenza del Consiglio dei Ministri, Roma.
- Schwarze C. (1988), *Grammatik der italienischen Sprache*, Niemeyer, Tübingen.
- Serianni L. 2000 [1988], *Italiano. Grammatica, sintassi, dubbi*, Garzanti, Milano.
- Setti R. (2003), “L’articolo prima di un prenome”, *Accademia della Crusca*, 4.4.2003.
<https://accademiadellacrusca.it/it/consulenza/larticolo-prima-di-un-prenome/98>
[consultato il 14.8.2023]
- Tavosanis M. (2018), *Lingue e intelligenza artificiale*, Carocci, Roma.
- Tavosanis M. (2021), “L’ideologia linguistica e le pratiche di *Wikipedia* in lingua italiana”, in Alamán A. P., Ruggiano F. & Walsh O. (a c. di), *Le ideologie linguistiche: lingue e dialetti nei media vecchi e nuovi*, Berlin, Lang, pp. 413-434.

Appendice



Scrivi un testo di 1000 parole per spiegare chi è Dacia Maraini.



Dacia Maraini è una scrittrice, drammaturga e attivista italiana, nata a Fiesole, in provincia di Firenze, il 29 novembre 1936. Figlia del poeta e scrittore Tommaso Landolfi e della pittrice Elvira Maraini, Dacia Maraini ha trascorso l'infanzia e l'adolescenza a Roma, dove ha frequentato il liceo classico.

Nel 1956 si è laureata in Lettere Classiche presso l'Università di Roma, con una tesi sulla figura della madre nella poesia greca antica. Dopo la laurea, ha trascorso un anno in Inghilterra, dove ha studiato teatro e ha iniziato a lavorare come attrice.

Tornata in Italia, Dacia Maraini ha iniziato a scrivere per il teatro, pubblicando il suo primo lavoro, "La lunga vita di Marianna Ucrìa", nel 1960. Da allora, ha scritto numerosi altri testi teatrali, ottenendo un notevole successo sia in Italia che all'estero.

Oltre al teatro, Dacia Maraini si è anche dedicata alla scrittura di romanzi, saggi e poesie. Tra i suoi romanzi più noti si possono citare "Lettere a Marina" (1981), "Buio" (1987) e "La casa del sorriso" (1991), tutti pubblicati con grande successo di critica e di pubblico.

Dacia Maraini è stata anche impegnata in numerose attività di denuncia e di impegno sociale, sia in Italia che all'estero. Ha lottato per i diritti delle donne, per la pace e per la giustizia sociale, diventando una voce importante nella società italiana.

Nel corso della sua carriera, Dacia Maraini ha ricevuto numerosi riconoscimenti, tra cui il Premio Campiello nel 1981 per "Lettere a Marina", il Premio Strega nel 1987 per "Buio" e il Premio Flaiano nel 1991 per "La casa del sorriso".

Dacia Maraini è stata anche insignita della medaglia d'oro al merito della Repubblica Italiana, della medaglia d'oro dell'Accademia dei Lincei e della laurea honoris causa in Scienze della Comunicazione presso l'Università di Siena.

Oggi Dacia Maraini è considerata una delle scrittrici e attiviste più importanti e influenti in Italia, nonché una figura di riferimento per le donne di tutto il mondo. Con la sua voce potente e la sua passione per la giustizia, ha contribuito a far conoscere

Fig.1. BioCGPT-3.5 (generato il 7.1.2023)

Dacia Maraini

Da Wikipedia, l'enciclopedia libera.

Dacia Maraini (Firenze, 13 novembre 1936^{[1][2]}) è una scrittrice, poetessa e saggista italiana.

□

Biografia

Nasce a Firenze nel 1936, primogenita dell'antropologo, orientalista e scrittore fiorentino Fosco Maraini e della pittrice e gallerista palermitana Topazia Alliata, quest'ultima appartenente al ramo siciliano dell'antico casato pisano degli Alliata, ovverosia gli Alliata di Salaparuta. Il nonno paterno della futura scrittrice fu lo scultore e critico d'arte romano d'origini ticinesi e genovesi Antonio Maraini (1886-1963), deputato del Partito Nazionale Fascista dal 1934 al 1939, nonché stretto collaboratore del gerarca Achille Starace e principale fautore delle politiche artistico-culturali del regime fascista, mentre la nonna paterna fu la scrittrice inglese, nata nell'allora Ungheria asburgica e d'origini in parte polacche, Yoï Crosse (1877-1944)^[3]; il nonno materno fu il gastronomo Enrico Maria Alliata di Villafranca (1879-1946), proprietario della rinomata azienda vinicola *Corvo* ed ultimo signore delle antiche cantine di Casteldaccia, mentre la nonna materna fu Oria Maria Amelia "Sonia" Ortúzar Ovalle de Olivares (1892-1981), una cantante lirica, che però non poté esordire, figlia d'un diplomatico cileno^[4].

Maraini trascorse l'infanzia in Giappone^[5], dove i genitori si erano stabiliti nel 1939, e dove nacquero le sue sorelle Yuki e Antonella, detta Toni. A seguito della caduta del fascismo e susseguente dichiarazione dell'armistizio di Cassibile nel 1943, con cui dunque l'Italia spezzava i suoi legami con l'Asse, la famiglia venne internata in un campo di concentramento dalle autorità giapponesi, dove patì la fame.

Soltanto nel 1945 la famiglia riuscì a rientrare in Italia, stabilendosi dapprima in Sicilia, presso la tenuta dei nonni materni, Villa Valguarnera di Bagheria, ed in seguito a Roma. Dopodiché il padre Fosco, da solo, volle tornarsene a Firenze. Questi anni sono raccontati dalla stessa Maraini nel suo romanzo *Bagheria*:



Dacia Maraini nel 2012

Premio Campiello 1990

Premio Strega 1999



La famiglia Maraini, Fosco, Dacia, Yuki, Toni e la madre Topazia Alliata

Fig.2. BioWiki (consultato il 18.9.2023²⁵)

²⁵ https://it.wikipedia.org/w/index.php?title=Dacia_Maraini&oldid=135501906

ABSTRACT

L’obiettivo generale dello studio è indagare la rappresentazione linguistica delle donne all’interno di testi generati da ChatGPT, confrontandola con quella dei testi scritti da esseri umani. A questo fine esaminiamo la forma e frequenza d’uso di tre tipi di antroponimi riferiti alle donne (*Giorgia Meloni*, *Meloni*, *la Meloni*) attraverso un’analisi di testi biografici generati automaticamente da GPT-3.5 e pubblicati su *Wikipedia*. L’analisi è condotta utilizzando un corpus creato appositamente per questa indagine, che comprende biografie relative a 34 personalità femminili di rilievo. I risultati ottenuti mostrano che ci sono differenze sostanziali nel modo in cui le biografie generate da ChatGPT e quelle pubblicate su *Wikipedia* si riferiscono alle donne. Uno dei dati più interessanti riguarda la più bassa frequenza d’uso degli antroponimi del tipo *la Meloni* nelle biografie generate. Questo dato apre importanti domande sia in merito ai fattori che spiegano i dati riscontrati sia sul ruolo che i testi generati potrebbero avere in futuro sull’assetto dell’italiano contemporaneo.

The goal of this study is to investigate the linguistic representation of women within ChatGPT’s generated texts as compared with that of texts written by human beings. To this end, we examine the form and frequency of use of three types of anthroponyms referring to women (*Giorgia Meloni*, *Meloni*, *la Meloni*) through an analysis of biographical texts automatically generated by GPT-3.5 and published on Wikipedia. The analysis is conducted using a corpus created specifically for this investigation, which includes biographies related to 34 prominent female personalities. The results obtained show that there are substantial differences in how the biographies generated by ChatGPT and those published on Wikipedia relate to women. One of the most interesting outcomes concerns the lower frequency of use of anthroponyms of the type *la Meloni* in the generated biographies. This finding opens important questions both about the factors explaining the data found and about the role that the generated texts might play in the future on the structure of contemporary Italian.

KEYWORDS: testi generati, ChatGPT, *Wikipedia*, antroponimi femminili, articolo determinativo

DATA DI PUBBLICAZIONE: 30 gennaio 2024.