

# Vocabolari d'autore e intelligenza artificiale: è possibile creare il vocabolario della lingua di Boccaccio con ChatGPT?

Claudia Palmieri

(Università per Stranieri di Siena/Ente Nazionale Giovanni Boccaccio)

claudia.palmieri(at)unistrasi.it

## Abstract

L'innovazione tecnologica ha profondamente influenzato la lessicografia, aprendo nuove prospettive per la redazione e la consultazione dei vocabolari. Questo contributo si propone di indagare il ruolo dell'intelligenza artificiale generativa, in particolare di ChatGPT, nella creazione di un vocabolario del lessico impiegato da Giovanni Boccaccio nel *Decameron*. Attraverso un'indagine sperimentale condotta su un *corpus* di vocaboli appartenenti all'area botanica e naturalistica, il lavoro esplora le capacità e i limiti di ChatGPT nella generazione di voci lessicografiche. Vengono esaminati aspetti come la struttura delle definizioni, l'accuratezza nell'individuazione dei diversi significati e l'adeguatezza degli esempi testuali tratti dalla principale opera del Certaldese. Il contributo intende fornire una base di riflessione sulle prospettive future dell'IA generativa nel supporto alla compilazione di un vocabolario in lingua italiana, delineando opportunità e sfide per una sua integrazione efficace nella pratica lessicografica.

Technological innovation has profoundly influenced lexicography, opening new perspectives for both the compilation and consultation of dictionaries. This paper investigates the role of generative artificial intelligence in the creation of a vocabulary of the lexicon used by Giovanni Boccaccio in the *Decameron*. Through an experimental analysis conducted on a corpus of terms belonging to the botanical and naturalistic domain, this work explores the capabilities and limitations of ChatGPT in generating lexicographic entries. Key aspects examined include the structure of definitions, the accuracy in distinguishing meanings, and the appropriateness of textual examples drawn from Boccaccio's masterpiece. This contribution seeks to foster reflection on the future prospects of generative AI in supporting the compilation of Italian dictionaries, highlighting both the opportunities and challenges associated with its integration into lexicographic practice.

## Parole chiave

Lingua italiana, Giovanni Boccaccio, lessicografia, intelligenza artificiale generativa, ChatGPT

## Keywords

Italian language, Giovanni Boccaccio, lexicography, Generative Artificial Intelligence, ChatGPT

## 1 Introduzione

Il *Dizionario di linguistica* diretto da Gian Luigi Beccaria, alla voce *lessicografia*, offre una definizione tanto essenziale quanto efficace: “Tecnica di compilazione di dizionari” (Beccaria 1994: 427). Tra le informazioni che sono contenute nella voce, spiccano i riferimenti al primo *Vocabolario degli Accademici della Crusca* (1612),



Palmieri, Claudia. 2025. Vocabolari d'autore e intelligenza artificiale  
Special Issue: *AI-Driven Linguistic Studies*, Vol. 2. No. 2  
DOI: 10.62408/ai-ling.v2i2.20  
ISSN: 2943-0070

che è stato un modello per tutta la lessicografia monolingue europea, e all'applicazione delle tecnologie digitali al lavoro di redazione di vocabolari:

La linguistica computazionale ha migliorato le tecniche della lessicografia, e ha facilitato la compilazione di concordanze, liste di frequenza, dizionari inversi, rimari. La cosiddetta *corpus linguistics*, la linguistica che si basa su studi condotti a partire da corpora di testi su supporto elettronico, ha facilitato la redazione di liste di parole nuove e diffuso in lessicografia l'abitudine a ricorrere a esempi tratti da corpora e ad organizzare lemmario e glosse tenendo conto della frequenza d'uso (Beccaria 2004: 447).

Questi accenni mettono in luce due fatti centrali nella storia della lessicografia italiana. Innanzitutto, viene posta l'attenzione sull'affermazione della Crusca come punto di riferimento nell'ambito della lessicografia. Fondato nel decennio 1570-1580, il sodalizio fiorentino intraprende la redazione del *Vocabolario* alla fine del XVI secolo.<sup>1</sup> Una delle peculiarità del lavoro della Crusca è la collegialità nelle scelte e nel modo di condurre il lavoro. A più di quattro secoli di distanza della prima edizione del *Vocabolario*, la collaborazione tra gli addetti ai lavori è uno dei principi che animano le esperienze di ricerca nell'ambito dell'umanistica digitale: “i progetti di digital humanities presuppongono un lavoro di équipe e incoraggiano una concezione aperta e plurale dell'autorialità all'interno dell'opera” (Previtali 2022: 26).

Questo ponte tra passato e presente ci permette di passare all'altro aspetto messo in rilievo da Beccaria: il connubio tra lessicografia e tecnologie digitali.<sup>2</sup> L'uso di corpora digitali, anche annotati, e la creazione e l'utilizzo di piattaforme per la redazione, la pubblicazione e la consultazione di vocabolari elettronici appartengono ormai alla moderna pratica lessicografica, sensibile alla valorizzazione delle innovazioni. È, dunque, coerente con questa prospettiva la scelta di Claudio Marazzini di chiudere il suo fondamentale volume sulla storia dei vocabolari italiani, *L'ordine delle parole*, con un paragrafo intitolato *Il vocabolario diventa elettronico*. Lo studioso riconosce che “l'informatica può essere di valido aiuto per la lessicografia attuale, sia nella fase di elaborazione del materiale, sia nella fase di realizzazione del prodotto destinato all'utente” (Marazzini 2009: 414). La riflessione sull'uso del calcolatore in lessicografia ha preso avvio prima degli anni Duemila. Già nel 1996, Carla Marellò evidenzia i vantaggi del ricorso all'informatica, sottolineando la possibilità, da parte del redattore, di conservare e rielaborare grandi quantità di dati. Quando Marellò scrive *Le parole dell'italiano. Lessico e dizionari*, non erano ancora stati messi a disposizione del pubblico lessici ideati e realizzati per essere distribuiti in forma digitale. “Si sono studiati programmi per potenziare la consultazione di dizionari a stampa trattandoli come basi di dati, ma nessuna casa editrice ha finanziato un progetto di dizionario-banca di dati linguistica nato per essere consultato attraverso il computer” (Marellò 1996:

<sup>1</sup> Sulla storia della prima Crusca, sono imprescindibili i lavori di Severina Parodi e di Maurizio Vitale. Cfr. Parodi (1983), Parodi (a c. di) (1993) e Vitale (1986).

<sup>2</sup> Sulla lessicografia digitale, cfr. Granger e Paquot (2012) e Tarp e Gouws (2023).

156). Nel quinto capitolo del libro, la studiosa dedica una particolare attenzione ad alcuni vocabolari in formato elettronico, in particolare su CD-ROM, e alle questioni che all'epoca si presentavano ai lessicografi.

Negli anni Novanta era già chiara la differenza tra i lessici cartacei e quelli elettronici. Nell'analizzare i due dizionari elettronici della lingua italiana allora in commercio, lo *Zingarelli minore* (XII ed.) e il *Devoto-Oli*, Riccardo Degl'Innocenti e Maria Ferraris esprimono chiaramente l'idea che è necessario considerare “un dizionario elettronico come un'opera concettualmente diversa da quella cartacea, rivolta a lettori che vanno messi in grado di scoprire e assumere nuovi punti di vista e capaci di porre nuove domande al dizionario” (Degl'Innocenti e Ferraris 1994: 165).

Trent'anni dopo, il pubblico ha a disposizione diverse risorse digitali consultabili attraverso il Web, tra le quali ci sono anche i capisaldi della lessicografia italiana: le cinque edizioni del *Vocabolario degli Accademici della Crusca*,<sup>3</sup> il *Dizionario della lingua italiana* di Niccolò Tommaseo<sup>4</sup> e il *Grande Dizionario della Lingua Italiana*<sup>5</sup> (GDLI). Tra le grandi imprese di lessicografia digitale in Italia si annovera il TLIO, un vocabolario storico di tutte le varietà dell'italiano antico con un limite cronologico che è posto convenzionalmente alla fine del XIV secolo.<sup>6</sup> Di recente, nel 2022, anche la grande impresa del *Lessico Etimologico Italiano* (LEI), avviata da Max Pfister, è stata pubblicata in Rete, grazie a un lavoro di digitalizzazione delle parti già pubblicate e dei materiali documentari, che si è affiancato alla progettazione di un'infrastruttura digitale per la redazione delle voci mancanti.<sup>7</sup>

Il caso del LEI permette di ribadire che la svolta digitale nella lessicografia riguarda non solo coloro che consultano i vocabolari, ma anche coloro che si occupano della redazione. Considerato l'interesse che negli ultimi due anni l'IA generativa ha suscitato, risulta opportuno chiedersi se un chatbot basato su un modello linguistico di grandi dimensioni come GPT4o possa supportare efficacemente la redazione di voci lessicografiche in lingua italiana,<sup>8</sup> principalmente per quanto riguarda la costruzione della definizione semantica e l'individuazione delle diverse accezioni di significato di un vocabolo.

Il presente contributo offre alcune osservazioni su prove condotte con un gruppo di vocaboli dell'area semantica della botanica e del mondo naturale. L'esperimento ha riguardato sia l'italiano contemporaneo sia l'italiano antico, con particolare

---

<sup>3</sup> Le edizioni sono consultabili al sito del progetto *Lessicografia della Crusca in Rete* (<http://new.lessicografia.it>).

<sup>4</sup> La versione digitale del *Dizionario* è consultabile all'indirizzo <https://www.tommaseobellini.it>.

<sup>5</sup> È possibile consultare il GDLI all'indirizzo <https://www.gdli.it>.

<sup>6</sup> Le voci del TLIO sono pubblicate on-line nel sito <http://tlio.ovl.cnr.it/TLIO>.

<sup>7</sup> Sui risultati del lavoro di digitalizzazione del LEI, cfr. Virgilio (2023).

<sup>8</sup> La comunità accademica internazionale ha dedicato al rapporto tra IA generativa e lessicografia alcuni contributi. Cfr. de Schryver 2023 e la bibliografia ivi citata. Per la lingua italiana, risultano fondamentali i contributi di Anna-Maria De Cesare (2023 e 2024), Mirko Tavosanis (2024) e Francesco Cicero (2023).

attenzione al lessico del *Decameron* di Giovanni Boccaccio, che comprende oltre 6550 lemmi.<sup>9</sup>

L'indagine che qui si presenta prende avvio dal progetto del VocaBO, il *Vocabolario di Boccaccio Online: il Decameron*. Si tratta di un vocabolario d'autore sviluppato interamente in ambiente digitale, ancora in fase di realizzazione, promosso dall'Università per Stranieri di Siena e dall'Ente Nazionale Giovanni Boccaccio in collaborazione con il Cnr-Istituto di Linguistica Computazionale "Antonio Zampolli" di Pisa (CNR-ILC) e con il sostegno dell'Accademia della Crusca.<sup>10</sup> Il progetto è stato avviato alla fine del 2022 sotto la direzione di Giovanna Frosini, presidente dell'Ente Nazionale Giovanni Boccaccio. La redazione avviene attraverso MAIA-VocaBO, uno strumento Web collaborativo per la lessicografia digitale sviluppato dal gruppo di ricerca KLAB ('Knowledge Lexica & Applications Bridging') dell'ILC. All'interno di un unico ambiente, MAIA unisce funzionalità tradizionalmente gestite da software distinti e separati, tra cui la creazione e l'annotazione multilivello di un corpus testuale, la costruzione di una risorsa lessicale e il *Lexical Linking*, ossia il collegamento diretto tra le annotazioni linguistiche del corpus e gli elementi di suddetta risorsa.<sup>11</sup>

Prima di illustrare alcuni risultati significativi, è necessario fare una precisazione. La generazione automatica di voci di dizionario da parte di un modello linguistico come GPT4o si distingue in maniera sostanziale dalla redazione tradizionale, non soltanto per il mezzo tecnologico impiegato, ma anche per la natura profondamente diversa dei processi sottesi. Nella pratica lessicografica, per la redazione di un dizionario basato su un *corpus*, il lessicografo esegue una serie di attività intellettuali con l'obiettivo di rappresentare il significato di un lemma in modo coerente e funzionale alla consultazione: seleziona il *corpus* di riferimento, analizza i contesti d'uso, compie un lavoro di interpretazione e costruisce la struttura semantica. Al contrario, un modello come GPT4o genera testi attraverso un meccanismo di previsione statistica,<sup>12</sup> basato su correlazioni apprese durante l'addestramento, senza disporre di una capacità di riflessione metalinguistica; per questo motivo non è possibile ricostruire con certezza i processi attraverso cui vengono selezionati i significati, strutturate le definizioni o organizzati gli esempi all'interno di una voce. L'osservazione delle voci generate permette di rilevare alcune caratteristiche del comportamento del modello, inerenti sia al piano formale,

---

<sup>9</sup> Il lessico impiegato da Boccaccio solo di recente ha attirato una maggiore attenzione negli studi linguistici. Cfr. Frosini (2014), Manni (2016), Ricotta (2017), Ricotta (2019) e De Vecchis (2022). Nel 2013 è stata pubblicato per la prima volta, in appendice all'edizione commentata del *Decameron* della BUR, il repertorio *Le cose (e le parole) del mondo*, a cura di Amedeo Quondam, che offre una ricognizione dei vari ambiti lessicali presenti nel *Centonovelle*. Cfr. Boccaccio (2017: 1669-1815).

<sup>10</sup> La redazione del VocaBO è attualmente composta da Veronica Ricotta (caporedattrice), Ester Baldi, Leila Bencharaki, Cecilia Cartoceti, Sara Di Giovannantonio, Claudia Palmieri e Emanuela Gioia Pisco.

<sup>11</sup> Sul funzionamento di MAIA, cfr. Giovannetti, Albanesi, Bellandi et al. (2024).

<sup>12</sup> Per un approfondimento sull'architettura alla base del funzionamento dei modelli linguistici di grandi dimensioni, si rimanda all'articolo fondativo di Vaswani, Shazeer, Parmar et al. (2017).

ovvero alla struttura della voce, sia a quello contenutistico, in particolare alla selezione e all'organizzazione dei significati.

## 2 Alcune prove preliminari

Prima di dedicarsi al lessico boccacciano, è stato realizzato un piccolo vocabolario contemporaneo con il modello ChatGPT 4o. Le voci sono state generate per trenta lemmi italiani appartenenti all'ambito botanico e a quello naturalistico.<sup>13</sup> Il campione comprende diverse parti del discorso. Per ogni lemma è stata creata una conversazione distinta. Al modello è stata proposta la seguente richiesta: “Genera una voce di dizionario per il vocabolo [vocabolo]”.

Alcune caratteristiche dei risultati ottenuti da questa prima indagine meritano di essere evidenziate. Innanzitutto, considerata la genericità del *prompt*, non sorprende il fatto che il modello abbia risposto con voci dalla struttura differente. In alcuni casi, vengono fornite solo le definizioni e l'etimologia, oltre alla categoria grammaticale; in altri casi, la voce contiene anche altre informazioni. Per esempio, la voce *lago* include anche i campi relativi a sinonimi, contrari, derivati ed espressioni comuni:

- (1) *lago sostantivo maschile*
1. Grande massa d'acqua raccolta in una conca naturale o artificiale  
Un lago è una vasta distesa d'acqua, generalmente dolce, circondata da terre emerse e situata in una depressione del terreno. Può avere origini glaciali, vulcaniche, tettoniche o essere artificiale.  
Esempio: Il Lago di Garda è il più grande d'Italia.
  2. Bacino d'acqua di minori dimensioni rispetto a un mare  
Può essere alimentato da fiumi, sorgenti o precipitazioni e avere un emissario che lo collega a un altro corpo idrico.
  3. (Figurato) Grande quantità di liquido raccolto in uno spazio limitato  
Esempio: Dopo la pioggia, la strada era un lago d'acqua.
  4. (Tecnico) In alcuni settori, deposito o ristagno di liquidi  
Esempio: Lago di metallo fuso in una fonderia.  
Etimologia: Dal latino *lacus*, con lo stesso significato.  
Sinonimi: specchio d'acqua, bacino, laghetto (se di piccole dimensioni).  
Contrari: terraferma, deserto.  
Derivati: lacustre, laghetto, lagunare.  
Espressioni comuni: “Fare il lago” (in senso figurato, creare una pozza di liquido).

---

<sup>13</sup> I vocaboli scelti appartengono al settore del lessico di cui mi occupo all'interno del progetto del VocaBO. Le voci generate con ChatGPT sono: *abete, acero, albero, alloro, betulla, castagno, cedro, ciliegio, collina, faggio, fiore, fiorire, fiume, fragola, frassino, germogliare, ginepro, legnoso, mela, montagna, monte, noce, pioppo, quercia, ramificare, rampicante, rosa, salice, tiglio, zucca*.

L'esempio appena citato permette di notare alcune particolarità che riguardano le definizioni. In particolare, i primi due significati sono accompagnati da una sorta di precisazione. Per esempio, la prima definizione proposta per *lago*, “Grande massa d’acqua raccolta in una conca naturale o artificiale”, è seguita da un’ulteriore definizione in forma discorsiva: “Un lago è una vasta distesa d’acqua, generalmente dolce, circondata da terre emerse e situata in una depressione del terreno. Può avere origini glaciali, vulcaniche, tettoniche o essere artificiale”. Inoltre, si può notare che le accezioni 3 e 4 sono precedute da due parole – *Figurato* e *Tecnico*<sup>14</sup> – che sono assimilabili alle marche semantiche e alle marche d’uso che si trovano nelle opere lessicografiche. È interessante anche l’inserimento dell’espressione *fare il lago*<sup>15</sup> ‘creare una pozza di liquido’, che non risulta registrata nei vocabolari (cfr. GDLI/GRADIT/Zingarelli, s.v. *lago*).

In altre voci, come *cedro*, sono proposte come definizioni delle informazioni di carattere enciclopedico:

- (2) *cedro* s.m. [dal lat. *cedrus*, dal gr. *kédros*]
1. (Botanica) Albero sempreverde della famiglia delle Rutacee (*Citrus medica*), originario dell’Asia sudorientale, coltivato per i suoi frutti profumati e rugosi, utilizzati nell’industria alimentare e profumiera. Il frutto del cedro è di grandi dimensioni, con una buccia spessa e ricca di oli essenziali, spesso candita o impiegata nella preparazione di liquori e dolci.
  2. (Legno) Nome comune di diverse specie arboree della famiglia delle Pinacee, tra cui il cedro del Libano (*Cedrus libani*), il cedro dell’Atlante (*Cedrus atlantica*) e il cedro dell’Himalaya (*Cedrus deodara*), caratterizzate da un legno aromatico e resistente, impiegato in ebanisteria e costruzioni.
  3. (Simbolismo e cultura) Il cedro, in particolare il cedro del Libano, è spesso associato alla maestosità, alla longevità e alla sacralità, ed è presente in stemmi e bandiere, come nel caso del Libano.
  4. (Uso figurato, raro) Persona o cosa di grande robustezza e longevità.

Per il lemma vengono fornite quattro accezioni distinte: la prima è quella di ‘albero sempreverde’; la seconda fa riferimento alle ‘diverse specie di piante della famiglia delle Pinacee’; la terza, accompagnata dall’etichetta *Simbolismo e cultura*, contiene un’informazione di tipo culturale; la quarta chiarisce un uso figurato del vocabolo che non risulta documentato nei vocabolari (cfr. GDLI/GRADIT/Zingarelli, s.vv. *cedro 1* e *cedro 2*). Il confronto di questa struttura semantica con quella proposta da alcuni lessici dell’italiano risulta utile per individuare il significato di *cedro*.

<sup>14</sup> Il significato effettivamente registrato dai lessici che maggiormente si avvicina a quello dell’accezione 4, che fa però riferimento a un preciso uso tecnico del vocabolo, è quello di ‘*bacino d’acqua stagnante*’, attestato in autori antichi, a partire da Dante (cfr. TLIO, s.v. *lago*), e di epoca moderna (cfr. GDLI, s.v. *lago*).

<sup>15</sup> Il GDLI (s.v. *lago*) riporta solamente una locuzione con il verbo *fare, fare lago (di qualcosa)*, con il significato di ‘fare soverchio e inutile sfoggio (di qualcosa)’. *Fare il lago* potrebbe essere piuttosto assimilabile a *fare un lago* ‘versare una grande quantità di liquido’.

Innanzitutto, è importante sottolineare che i vocabolari registrano due lemmi *cedro* differenti, che derivano da due etimi distinti: dal lat. *citru(m)* deriva l'italiano *cedro* 'albero delle Rutacee che produce il caratteristico frutto giallo'; dal lat. *cedru(m)* si ha in italiano *cedro* 'pianta delle Pinacee'.

Il GDLI propone due voci. In *cedro 1* vengono presentate le seguenti definizioni: “Grande e maestosa conifera della famiglia Pinacee (*Cedrus libanotica* o *Cedrus Libani*), nota come *cedro del Libano* [...]; un'altra specie, il *Cedro dell'Atlante* (*Cedrus atlantica*) [...]; il *Cedrus brevifolia* è proprio di Cipro; il *Cedrus deodara* (dell'Himalaya occidentale)”; “Legno di cedro (usato per costruzioni, per mobili, per lavori di ebanisteria, per sculture, ecc.)”; “Nome di varie conifere del genere *Juniperus*”. Per *cedro 2* le definizioni sono: “Albero della famiglia Rutacee (*Citrus medica*), alto fino a 3-4 m, con rami spinosi, foglie ampie, lunghe, fiori in grappoli, frutti grossi, di forma ovoidale con prominenze basali e apicali, colore giallo canarino”; “Il frutto del cedro; l'essenza che se ne ricava”; “Oggetto a forma di cedro”; “Gelato aromatizzato con sciroppo di cedro”. Lo Zingarelli registra sotto una medesima voce (*cedro 1*) i due significati di “alberetto sempreverde delle Rutacee con foglie persistenti semplici, fiori bianchi e frutto tondeggiate di colore giallo (*Citrus medica*)” e “il frutto della pianta omonima, grosso, verrucoso, con buccia bitorzoluto, è usato in pasticceria, liquoreria e farmacia”, mentre in un'altra voce (*cedro 2*) registra il significato di “genere delle Pinacee comprendente piante arboree sempreverdi, con scorza bruna, foglie aghiformi, strobili eretti, cono con squame persistenti (*Cedrus*)”.

Rispetto alle voci appena presentate, quella generata da ChatGPT raccoglie in un'unica voce le definizioni che i vocabolari inseriscono sotto due entrate lessicali differenti, sebbene omonime, e offre come etimo solo il lat. *cedrus*. Inoltre, la voce generata non registra tutti i significati di *cedro*, tra i quali quello di 'frutto'. Concentrando l'attenzione sull'etimologia, è interessante l'esempio di un altro lemma, *fragola*. Secondo il DELIN (s.v. *fragola*), il vocabolo deriva dal lat. parlato \**fragula(m)*, dim. di *fragum* 'fragola', mentre il DEI offre un'ulteriore specificazione: “lat. \**fragula*, dimin. di *fraga* 'fraga', class. *fragum*, relitto mediterraneo come il sinonimo 'magiostra'” (DEI, s.v. *fragola*). Nella voce di ChatGPT, viene indicato come etimo *fragum* e non viene citata la forma *fragula*.

Dopo aver osservato alcuni esempi di voci, che hanno già permesso di mettere in evidenza alcuni comportamenti del modello – capace di generare testi che riproducono la struttura delle voci lessicografiche, ma che talvolta contengono informazioni non corrette –, risulta opportuno presentare i risultati ottenuti dalla generazione di voci boccacciane.

### 3 Un vocabolario d'autore della lingua di Boccaccio con ChatGPT

Il lemmario di riferimento comprende cinquanta voci, anche in questo caso appartenenti all'ambito naturalistico e a quello botanico.<sup>16</sup> Lo scopo del vocabolario è quello di indagare l'uso che Boccaccio fa di questa sezione del lessico e di individuare i significati che i vocaboli hanno nel contesto del *Decameron*. Le voci sono state generate con il modello GPT-4o, con un *prompt one-shot* che ha fornito al modello un esempio da seguire per svolgere il compito richiesto. Per ogni voce è stata avviata una conversazione e al *prompt* è stato allegato il testo del *Centonovelle*. Il *prompt* fornito al modello è:

- (3) Genera una voce di dizionario per il vocabolo [vocabolo] presente nel Decameron di Giovanni Boccaccio, utilizzando il testo allegato come riferimento. La voce deve seguire il modello dell'esempio fornito e includere le seguenti sezioni:
1. lemma
  2. categoria grammaticale
  3. definizione: fornisci una spiegazione chiara e accurata del significato della parola nel contesto dell'opera. Se trovi più significati, elencali tutti.
  4. esempi d'uso tratti dal *Decameron*: per ogni significato riporta almeno una frase in cui la parola compare nel testo, con riferimento alla giornata e alla novella (es. *Decameron* II, 4).

Importante: gli esempi devono essere citazioni dirette dal *Decameron*.

Esempio di voce:

albero s.m.

1. Pianta perenne con fusto eretto legnoso che presenta ramificazioni con foglie.

Esempio: Ma senza arme trovandosi, ricorse a prendere un ramo d'albero in luogo di bastone e cominciò a farsi incontro a' cani e contro al cavaliere (*Decameron* V, 8).

2. Nelle navi, palo che serve a sostenere la vele.

Esempio: La giovane, la quale senza misura della partita di Martuccio era stata dolente, udendo lui con gli altri esser morto, lungamente pianse e seco dispose di non voler più vivere; e non sofferendole il cuore di se medesima con alcuna violenza uccidere, pensò nuova necessità dare alla sua morte: e uscita segretamente una notte di casa il padre e al porto venutasene, trovò per ventura alquanto separata dall'altre navi una navicella di pescatori, la

---

<sup>16</sup> Le voci generate riguardano i seguenti lessemi: *abete, aglio, arancio, baccello, basilico, biada, bosco, campo, castagna, caverna, cespuglio, cipolla, cipresso, colle, erba, fico, fiore, fiorire, fiume, frassino, fronda, fronzuto, gelsomino, giardino, ginestra, giunco, grano, grotta, isola, mandorlo, mare, marina, mela, melarancio, pelago, persa, spiaggia, pianta, piantare, pianura, pineta, pruno, ramo, rena, scalogno, scoglio, selva, selvatico, terra, vigna.*

quale, per ciò che pure allora smontati n'erano i signori di quella, d'albero e di vela e di remi la trovò fornita (*Decameron* V, 2).

A differenza di quelle generate con il *prompt zero-shot*, le voci create a partire dal *prompt one-shot* sono organizzate secondo la struttura descritta e l'esempio fornito. Nel formulare la richiesta, si è rivelato necessario fornire nel *prompt* la forma esatta in cui il vocabolo compare nel *Decameron*. Il termine *abete*, per esempio, ricorre solo una volta nel capolavoro boccacciano, al plurale. Al *prompt* contenente la forma al singolare, il modello ha risposto generando una voce lessicografica senza l'esempio:

- (4) abete s.m.  
 1. Pianta conifera sempreverde appartenente al genere *Abies*, caratterizzata da aghi sottili e da un portamento slanciato, spesso utilizzata per il legname e per decorazioni.  
 Esempio: (mancano citazioni dirette nel testo analizzato).

Alla richiesta di generare una voce per *abeti*, la risposta riporta un esempio, che costituisce una sorta di parafrasi del passo originale. Come indicato nella voce, il vocabolo ricorre effettivamente nella *Conclusione* della VI giornata:<sup>17</sup> “Il piano appresso, senza aver più entrate che quella donde le donne venute v'erano, era pieno d'abeti, di cipressi, d'allori e d'alcun pini sì ben composti e sì bene ordinati, come se qualunque è di ciò il migliore artefice gli avesse piantati” (*Decameron* VI, *Conclusione*, 24).

- (5) abeti s.m. pl.  
 1. Piante conifere sempreverdi appartenenti al genere *Abies*, caratterizzate da fusti slanciati e aghi sottili, frequentemente utilizzate per il legname e in contesti ornamentali.  
 Esempio: “... tutti quegli alberi li avesse piantati il migliore artefice ...” (*Decameron* VI giornata).

In altri casi, la citazione proposta corrisponde al testo decameroniano. Per esempio, la prima definizione di *bosco* – “Ampia area coperta da alberi e vegetazione spontanea, spesso scenario di eventi narrativi caratterizzati da incontri, fughe o avventure” – è accompagnata dalla citazione di *Decameron* IX, 7, 5: “Ora avvenne una notte, essendo Talano con questa sua Margherita in contado a una lor possessione, dormendo egli, gli parve in sogno vedere la donna sua andar per un bosco assai bello, il quale essi non guari lontano alla lor casa avevano”. Il caso di *bosco* permette di mettere in evidenza anche un altro fatto. Se la prima definizione è accettabile, la seconda non introduce un diverso significato del vocabolo, ma

---

<sup>17</sup> Tutte le citazioni del *Decameron* sono tratte dall'edizione curata da Giancarlo Alfano, Maurizio Fiorilla e Amedeo Quondam (Boccaccio 2017).

piuttosto una specificazione di un eventuale valore simbolico del bosco: “Luogo simbolico nella narrazione, spesso associato al mistero, alla paura o alla trasgressione”. Inoltre, la citazione inserita come esempio per questa seconda accezione non contiene il vocabolo *bosco*: “La donna crollando il capo disse: «Chi mal ti vuol, mal ti sogna: tu ti fai molto di me pietoso ma tu sogni di me quello che tu vorresti vedere»” (*Decameron* IX, 7, 8).

Non sempre sono individuate tutte le accezioni di un vocabolo. Per esempio, *mare* compare nel *Decameron* con il significato di ‘distesa di acqua salata che circonda le terre emerse’ e con il significato estensivo di ‘onde (del mare)’. Il lessema ricorre anche in interessanti costruzioni sintattiche, come *da mare* ‘rispetto al mare (come punto di riferimento spaziale)’, *in mezzo mare* ‘in mezzo al mare’ e *rompere in mare* ‘fare naufragio’, e nell’espressione *non rimproverare al mare d’averlo fatto crescere il piccolo ruscelletto* ‘non rimproverare il mare se il piccolo ruscelletto lo ha fatto crescere di livello; non confondere causa ed effetto’ (cfr. Boccaccio 2017: 1286). Nella voce generata da ChatGPT sono proposte due definizioni. Anche in questo caso, quella riportata nel secondo paragrafo della voce non è una vera e propria definizione, ma piuttosto una breve descrizione del valore dello spazio marittimo nella narrazione:

- (6) mare s.m.
1. Distesa d’acqua salata che circonda le terre emerse, spesso scenario di avventure, naufragi e viaggi nel *Decameron*.  
Esempio d’uso: “Ma come che il fatto s’andasse, adivenne che solutosi subitamente nell’aere un groppo di vento e percosso nel mare sì grande in questa cassa diede e la cassa nella tavola sopra la quale Landolfo era, che, riversata, per forza Landolfo lasciatala andò sotto l’onde” (*Decameron* II, 4).
  2. Simbolo dell’incertezza e del destino avverso, il mare appare spesso come un luogo di pericoli e mutamenti improvvisi.  
Esempio d’uso: “Levandosi la sera uno scilocco, il quale non solamente era contrario al suo cammino ma ancora faceva grossissimo il mare, il quale il suo picciolo legno non avrebbe bene potuto comportare...” (*Decameron* II, 6).

Anche la voce *giardino* contiene una prima definizione che corrisponde all’effettivo significato del vocabolo all’interno del *Decameron* – “Spazio verde recintato, coltivato con piante, fiori e alberi” –, ma anche altre definizioni che non contengono la descrizione di altre accezioni della parola: “Luogo simbolico di piacere e perfezione”; “Scenario di eventi narrativi e luogo di incontri, segreti e amori”. A questi esempi si può aggiungere quello della voce *erba*, che nel *Centonovelle* ricorre con i significati di ‘pianta dal fusto basso e non legnoso’, ‘insieme di piante erbacee’ e ‘insieme delle piante erbacee che ricopre un area di terreno e il terreno stesso che ne è rivestito’. Anche in questo caso, la prima definizione è adeguata al contesto allegato come esempio. Il secondo e il terzo paragrafo rappresentano perlopiù un

breve commento ai passi del *Decameron*. Inoltre, l'esempio citato per la terza accezione è tratto dalla decima novella della terza giornata, tranne che per il soggetto, Masetto, che è il protagonista della prima novella della terza giornata:

- (7) erba s.f.
1. Vegetazione erbacea, prato: Indica l'insieme delle piante erbacee che coprono il terreno, spesso usato per descrivere ambientazioni naturali e piacevoli nel *Decameron*.  
Esempio: "E quivi, sentendo un soave venticello venire, sì come volle la lor reina, tutti sopra la verde erba si puosero in cerchio a sedere" (*Decameron* I, introduzione).
  2. Elemento ornamentale e simbolico: L'erba è descritta come elemento decorativo e simbolico, spesso associata alla bellezza e al benessere.  
Esempio: "Nel mezzo del quale, quello che è non meno commendabile che altra cosa che vi fosse ma molto più, era un prato di minutissima erba e verde tanto, che quasi nera pareva" (*Decameron* III, 1).
  3. Nutrimento e ambiente per la vita: In alcuni passaggi, l'erba è citata come parte integrante dell'ambiente naturale che accoglie i personaggi.  
Esempio: "Masetto, che di radici d'erba e d'acqua vivea, poteva male rispondere alle poste" (*Decameron* III, 1).

A partire dalla voce generata da ChatGPT e dagli esempi testuali in essa utilizzati, l'indagine è proseguita con la redazione di una voce lessicografica che ha assunto l'output del modello come punto di partenza, rielaborandolo attraverso un controllo sistematico di tutte le occorrenze del lessema nel *Decameron*:

- (8) erba s.f.
1. (Bot.) Pianta dal fusto basso e non legnoso.  
Esempio: Molti altri servavano, tra questi due di sopra detti, una mezzana via, non strignendosi nelle vivande quanto i primi né nel bere e nell'altre dissoluzioni allargandosi quanto i secondi, ma a sufficienza secondo gli appetiti le cose usavano e senza rinchiudersi andavano a torno, portando nelle mani chi fiori, chi erbe odorifere e chi diverse maniere di spezierie, quelle al naso ponendosi spesso, estimando essere ottima cosa il cerebro con cotali odori confortare, con ciò fosse cosa che l'aere tutto paresse dal puzzo de' morti corpi e delle infermità e delle medicine compreso e puzzolente (*Decameron*, I, *Introduzione*, 24)
  2. (Estens.) Insieme di piante erbacee.  
Esempio: Rustico, che di radici d'erba e d'acqua vivea, poteva male rispondere alle poste; e dissele che troppi diavoli vorrebbero essere a potere il ninferno attutare ma che egli ne farebbe ciò che per lui si potesse (*Decameron* III, 10, 30).  
2.1 L'insieme delle piante erbacee che ricopre un'area di terreno e, in alcuni casi, anche il terreno stesso che ne è rivestito.

Esempio: E quivi, sentendo un soave venticello venire, sì come volle la lor reina, tutti sopra la verde erba si puosero in cerchio a sedere (*Decameron*, I, *Introduzione*, 109).

Un ulteriore caso che si intende illustrare riguarda il verbo *fiorire*. In una finestra di conversazione, è stata prima prodotta la voce a partire dalla richiesta “Genera una voce di dizionario per il vocabolo fiorire”. Il risultato è una voce che propone diverse accezioni accompagnate da esempi:

- (9) fiorire v.intr. (aus. essere o avere)  
 (Botanica) Svilupparsi e aprirsi in fiore; detto di piante, produrre fiori.  
 Es.: In primavera i ciliegi iniziano a fiorire.  
 (Figurato) Raggiungere un periodo di massimo splendore, prosperare.  
 Es.: Durante il Rinascimento, l’arte e la cultura hanno fiorito in Italia.  
 (Letterario) Manifestarsi in modo evidente e abbondante.  
 Es.: Sul suo viso fioriva un sorriso radioso.  
 (Di fenomeni o movimenti) Diffondersi, svilupparsi ampiamente.  
 Es.: Negli ultimi anni, le piccole imprese tecnologiche hanno fiorito nel mercato digitale.

Successivamente, nella medesima finestra, è stata generata la voce con il *prompt* usato per tutti gli altri lemmi del vocabolario di Boccaccio. Coerentemente con l’uso del verbo *fiorire* che fa il Certaldese, il modello registra solo il significato figurato di ‘essere in una fase di prosperità e di sviluppo’. La definizione è seguita da un passo – erroneamente ricondotto al *Proemio* – dell’*Introduzione* alla quarta giornata (IV, *Introduzione*, 38).

- (10) fiorire v.intr.  
 1. Crescere e svilupparsi in maniera rigogliosa.  
 Esempi d’uso tratti dal *Decameron*: “E già più ne trovarono tralle loro favole i poeti, che molti ricchi tra’ loro tesori, e assai già, dietro alle loro favole andando, fecero la loro età fiorire, dove in contrario molti nel cercar d’aver più pane, che bisogno non era loro, perirono acerbi” (*Decameron*, Proemio).

#### 4 Conclusioni

L’osservazione delle prove di redazione di voci lessicografiche con ChatGPT ha permesso di mettere in luce punti di forza e limiti dell’impiego dell’intelligenza artificiale generativa in questo specifico ambito. L’esperimento, seppur circoscritto a un campione limitato di vocaboli, fornisce interessanti spunti di riflessione sul comportamento del modello nell’elaborazione di definizioni, nell’organizzazione delle informazioni e nella gestione di elementi tipici della pratica lessicografica, come la selezione delle marche d’uso e l’identificazione di più accezioni per uno stesso lemma. Tuttavia, è necessario sottolineare che i risultati ottenuti non

permettono di formulare un giudizio definitivo sull'efficacia dell'impiego di ChatGPT per la redazione di vocabolari in italiano.

Da un lato, si riconosce la capacità di questi strumenti di generare testi articolati<sup>18</sup> che – nel caso delle voci generate con il *prompt one-shot* – seguono la struttura descritta e l'esempio fornito e riportano il significato corretto di un lessema, come nel caso di *fiorire*. Dall'altro, emergono limiti che ne riducono l'applicabilità in un contesto lessicografico rigoroso: il modello, infatti, non è sempre in grado di garantire uniformità nella struttura delle voci, nelle quali vengono talvolta incluse informazioni che riguardano il valore simbolico-narrativo del lessema nel contesto del *Decameron*, come nei casi di *mare*, *giardino* ed *erba*; non sempre sono individuate tutte le accezioni; a volte vengono incluse nella voce anche informazioni non corrette. Un altro elemento di criticità risiede nella natura stessa del modello, che opera in modo non deterministico. Per ottenere voci soddisfacenti, sarebbero necessari un costante intervento e la revisione del testo generato da parte del redattore, come nell'esempio di *erba*. Tale necessità riduce in misura significativa il vantaggio, in termini di tempo di lavoro, offerto dall'impiego dell'IA generativa.

L'esperimento presentato ha avuto, quindi, il principale obiettivo di esplorare il comportamento di ChatGPT rispetto alla lessicografia in italiano e alla redazione di un vocabolario d'autore. I risultati ottenuti rappresentano un punto di partenza per ulteriori approfondimenti, che potranno concentrarsi su altri aspetti del rapporto tra intelligenza artificiale e redazione lessicografica. La complessità del tema richiede indagini più estese e sistematiche, che potranno contribuire a delineare con maggiore chiarezza il ruolo che gli strumenti basati su modelli linguistici di grandi dimensioni possono avere nella pratica lessicografica contemporanea.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this article.

---

<sup>18</sup> La tensione tra riproduzione formale e comprensione del significato da parte dei modelli linguistici è al centro di un ampio dibattito. Sogaard (2025) ha individuato cinque differenti posizioni sulla capacità di comprensione da parte dei modelli linguistici di grandi dimensioni. Secondo alcuni, questi modelli, addestrati esclusivamente a prevedere l'elemento successivo in una sequenza testuale o a ricostruire un elemento mancante, non hanno comprensione semantica. Altri ritengono che essi possiedano soltanto una semantica inferenziale, ossia la capacità di stabilire relazioni logico-associative tra espressioni linguistiche, senza però collegarle a oggetti o eventi reali. Una terza posizione attribuisce ai modelli, oltre alla semantica inferenziale, una forma di semantica referenziale, poiché le loro rappresentazioni interne, di tipo numerico-vettoriale, mostrano somiglianze strutturali che emergono tra lingue, architetture e modalità differenti. Vi è poi chi ritiene che i LLM possiedano una competenza semantica anche referenziale in virtù della partecipazione a catene comunicative e chi sostiene, invece, che una vera semantica referenziale sia acquisibile soltanto dai chatbot, grazie all'interazione e al feedback umano. Sogaard aggiunge anche la propria posizione, secondo cui la capacità semantica di modelli linguistici e di chatbot è legata sia a fattori interni (le rappresentazioni numerico-vettoriali nel modello) sia a fattori esterni (i dati di addestramento, gli sviluppatori e gli annotatori, la scelta del modello e gli agenti umani).

## Riferimenti bibliografici

- Beccaria = Beccaria, Gian Luigi (dir. da). 2004. *Dizionario di linguistica e di filologia, metrica, retorica*. Torino: Einaudi.
- Boccaccio, Giovanni. 2017. *Decameron*, a c. di Alfano, Giancarlo & Fiorilla, Maurizio & Quondam, Amedeo. Milano: Rizzoli BUR.
- Cicero, Francesco. 2023. L'italiano delle intelligenze artificiali generative. *Italiano LinguaDue* 15 (2). 733–761. <https://doi.org/10.54103/2037-3597/21990>
- De Cesare, Anna-Maria. 2023. Assessing the quality of ChatGPT's generated output in light of human-written texts. A corpus study based on textual parameters. *CHIMERA. Romance Corpora and Linguistic studies* 10. 179–210. <https://revistas.uam.es/chimera/article/view/17979> (ultimo accesso 30/10/2025).
- De Cesare, Anna-Maria. 2024. Nuove dinamiche di contatto linguistico: Le “impronte digitali” dell'inglese nell'italiano generato da LLM. *Lid'O. Lingua Italiana d'Oggi* XXI. 67–91.
- Degl'Innocenti, Riccardo & Ferraris, Maria. 1994. Lexicon Valley. *Italiano & oltre* 9 (5). 262–265.
- DEI = Battisti, Carlo & Alessio, Giovanni. 1975. *Dizionario etimologico italiano*. Firenze: G. Barbera. 5 voll.
- DELIN = Cortelazzo, Manlio & Zolli, Paolo. 1999. *Il nuovo etimologico. DELI. Dizionario Etimologico della Lingua Italiana*, a c. di Cortelazzo, Manlio & Cortelazzo, Michele. Bologna: Zanichelli.
- de Schryver, Gilles-Maurice. 2023. Generative AI and lexicography. The current state of the art using ChatGPT. *International Journal of Lexicography* 36 (4). 355–387. <https://doi.org/10.1093/ijl/ecad021>
- De Vecchis, Kevin. 2022. Su due termini medici in Boccaccio: *emuntorio* e *gavocciolo*. *Lingua e stile* 57 (1). 107–122.
- Frosini, Giovanna. 2014. «Una imagnetta di nostra donna». Parole e cose nel testamento volgare di Giovanni Boccaccio. *Studi sul Boccaccio* 42. 1–24.
- GDLI = Battaglia, Salvatore & Barberi Squarotti, Giorgio (dir. da). 1961-2002. *Grande Dizionario della Lingua Italiana*. Torino: UTET. 21 voll.
- Giovannetti, Emiliano & Albanesi, Davide & Bellandi, Andrea & Marchi, Simone & Papini, Mafalda & Sciolette, Flavia. 2024. MAIA: an open collaborative platform for text annotation, e-lexicography, and lexical linking. *Umanistica Digitale* 18. 27–52. <https://doi.org/10.6092/issn.2532-8816/19705>
- GRADIT = De Mauro, Tullio (dir. da). 1999. *Grande Dizionario Italiano dell'Uso*. Torino: UTET. 6 voll.
- Granger, Sylviane & Paquot, Magali (eds). 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Manni, Paola. 2016. *La lingua di Boccaccio*. Bologna: il Mulino.
- Marazzini, Claudio. 2009. *L'ordine delle parole. Storia di vocabolari italiani*. Bologna: il Mulino.

- Marello, Carla. 1996. *Le parole dell'italiano. Lessico e dizionari*. Bologna: Zanichelli.
- Parodi, Severina. 1983. *Quattro secoli di Crusca. 1583-1983*. Firenze: Accademia della Crusca.
- Parodi, Severina (a c. di). 1993. *Gli atti del primo Vocabolario*. Firenze: Accademia della Crusca.
- Previtali, Giuseppe. 2022. *Che cosa sono le digital humanities*. Roma: Carocci.
- Ricotta, Veronica. 2017. «Istoriare e adornar di lavoro perfetto». Boccaccio e il lessico artistico medievale. In Zamponi, Stefano (a c. di), *Intorno a Boccaccio / Boccaccio e dintorni. Atti del Seminario Internazionale di Studi 2016* (Certaldo Alta, Casa di Giovanni Boccaccio, Settembre, 2016). Firenze: Firenze University Press. 113–123.
- Ricotta, Veronica. 2019. «Con animi e con vocaboli onestissimi si convien dire». Prime attestazioni e hapax in Boccaccio. *Studi di lessicografia italiana* 36. 67–101.
- Søgaard, Anders. 2025. Do language models have semantics? On the five standard positions. In Che, Wanxiang & Nabende, Joyce & Ekaterina Shutova & Pilehvar, Mohammad Taher (eds). *Proceedings of the 63rd annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 25910–25922. <https://aclanthology.org/2025.acl-long.1258.pdf> (ultimo accesso 04/11/2025).
- Tarp, Sven & Gouws, Rufus H. 2023. A necessary redefinition of lexicography in the digital age: glossography, dictionography and implications for the future. *Lexikos* 33. 425–447. <https://lexikos.journals.ac.za/pub/article/view/1826> (ultimo accesso 30/01/2025).
- Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1 (1). <https://doi.org/10.62408/ai-ling.v1i1.14>
- TLIO = *Tesoro della Lingua Italiana delle Origini*, fondato da Pietro Beltrami e continuato da Lino Leonardi, diretto da Paolo Squillacioti. CNR-Opera del Vocabolario Italiano. <http://tlio.ovi.cnr.it/TLIO> (ultimo accesso 07/06/2025).
- Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan N. & Kaiser, Lukasz & Polosukhin, Illia. 2017. Attention is all you need. <https://arxiv.org/pdf/1706.03762> (ultimo accesso 30/09/2025).
- Virgilio, Giulia. 2023. Primi risultati della digitalizzazione del Lessico Etimologico Italiano: il LEI online. *Archivio per il vocabolario storico italiano* 6. 217–224.
- Vitale, Maurizio. 1986. *L'oro nella lingua. Contributi per una storia del tradizionalismo e del purismo italiano*. Milano-Napoli: Ricciardi.
- Zingarelli = Zingarelli, Nicola. 2023. *Lo Zingarelli. Vocabolario della lingua italiana*, a c. di Cannella, Mario & Lazzarini, Beata & Zaninello, Andrea. Bologna: Zanichelli.