ANNA-MARIA DE CESARE, GIULIA MANTOVANI[1]
(Technische Universität Dresden)

*Valutare la naturalezza dell'*output *prodotto dall'IA generativa: uno studio empirico sull'uso dei suffissi alterativi in un corpus di favole sintetiche e scritte da esseri umani*

Abstract

This study empirically assesses the naturalness of AI-generated texts in comparison with human-written ones, using a corpus of fables authored by humans and generated by four LLMs (GPT-3.5 Turbo, GPT-4o mini, Teuken-7B, and Minerva-7B) trained on different datasets. To evaluate the naturalness of the fables, the study focuses on alterative suffixes, a distinctive feature of Italian that commonly occurs in child-oriented text genres. The results reveal that synthetic fables exhibit a lower usage of these suffixes, suggesting an indirect Anglicisation of these texts.

Keywords: GenAI, LLM, fables, Italian, alterative suffixes

1 *Introduzione*

Negli ultimi anni, l'IA cosiddetta *generativa* ha fatto un notevole balzo di qualità nel campo della scrittura. Grazie a tecniche computazionali molto raffinate, che poggiano su *transformer* e meccanismi di attenzione (per approfondimenti, cfr. Vaswani et al. 2017), l'*output* dei modelli linguistici di grandi dimensioni (dall'ingl. *Large Language Model*, d'ora in poi anche LLM) è diventato più naturale, vale a dire simile a quello prodotto da esseri umani (sull'uso del termine *naturale* in questo contesto, cfr. De Cesare 2026 e la spiegazione nella nota 2). In una serie di studi che valutano la qualità dell'italiano generato, si osserva però che l'*output* include anche tratti e strutture poco naturali, modellati sulla lingua inglese. I fenomeni in questione riguardano tutti i piani della lingua: la grafia e l'ortografia (*Dicembre* invece di *dicembre*; *clarezza* invece di *chiarezza*; gli esempi sono citati, rispettivamente, in De Cesare 2024: 73 e Tavosanis 2024: 17), il lessico (sono palesi soprattutto i calchi, come in *uno stile unico* invece di *uno stile personale*: Tavosanis 2024: 18) e la sintassi (ci limitiamo qui a menzionare l'uso del gerundio con funzione relativa, come in *Le due ragazze crescono in un quartiere povero di Napoli, con Lila dimostrando un'intelligenza straordinaria*; per dettagli, cfr. Cicero 2023: 739).

Le interferenze con l'inglese rilevate nell'italiano generato non sono sorprendenti: nei corpora di addestramento dei modelli linguistici, come per esempio quelli della famiglia GPT, i testi scritti in lingua inglese sono molto più numerosi di quelli scritti in italiano (su questo punto, cfr. per esempio Navigli, Conia & Ross 2023). L'impatto dei dati di addestramento sulla naturalezza dell'*output* generato in italiano è però ancora poco chiaro, anche perché mancano studi empirici qualitativi approfonditi e sistematici sulla questione

[1] Il contributo è frutto del lavoro di entrambe le autrici. La redazione dei singoli paragrafi è tuttavia la seguente: Anna-Maria De Cesare ha scritto i §§ 1 e 2, Giulia Mantovani, il § 3, mentre i risultati e le conclusioni (§§ 4, 5) sono stati redatti congiuntamente. Ringraziamo le tre persone che hanno valutato il contributo per i commenti costruttivi. Il limite di spazio a disposizione ci consente di tenere conto solo delle osservazioni principali.

(cfr. tuttavia De Cesare 2024, che propone una prima rassegna di impronte algoritmiche dell'inglese a partire dall'analisi di un corpus di biografie brevi). Gli studi a disposizione (cfr. i già citati Tavosanis 2024 e Cicero 2023, ai quali possiamo aggiungere almeno Antonelli 2024 e Cicero 2025) propongono osservazioni molto acute ma generalmente basate sull'analisi di pochissimi dati. Questi studi, inoltre, si concentrano sull'*output* di un solo modello linguistico o di più modelli ma sempre appartenenti a una stessa famiglia, in generale GPT di OpenAI. A nostra conoscenza, non disponiamo ancora di uno studio basato su corpora che valuti la naturalezza dell'*output* di modelli sviluppati *ad hoc* per generare testi di qualità in lingua italiana.

Il presente contributo persegue due obiettivi che mirano a colmare una parte delle lacune descritte sopra: a) capire meglio la qualità dei testi sintetici[2] confrontandoli con testi naturali, scritti da esseri umani; e b) capire la qualità di testi sintetici generati da modelli linguistici addestrati su corpora in cui l'italiano non ha lo stesso peso; si tratta dunque anche di capire meglio il ruolo che giocano i dati di addestramento sull'*output* generato in italiano da modelli linguistici diversi. Il campo che abbiamo scelto di analizzare in questa sede è molto circoscritto, ma specifico della lingua italiana, anche nel panorama romanzo (Simone 2010: 829): la morfologia derivazionale tramite l'uso di suffissi alterativi. In questo campo è noto che i suffissi non sono solo più numerosi in italiano che in inglese ma anche notevolmente più produttivi (sul concetto di *produttività*, cfr. per esempio Gaeta & Ricca 2006).

Per fare luce sull'uso dei suffissi alterativi nei testi naturali e nei testi sintetici generati in italiano abbiamo scelto di analizzare produzioni appartenenti al genere della favola (cfr. anche Mantovani in stampa, che analizza la presenza di suffissi diminutivi del nome in confronto a forme analitiche alternative formate tramite l'aggettivo *piccolo*). Questo genere testuale, tipico della comunicazione incentrata sul bambino (secondo la proposta di Dressler & Merlini Barbaresi 1994), costituisce in effetti uno dei terreni più fertili per l'uso di questi suffissi in italiano e in inglese. Più in particolare, confronteremo l'uso di sette suffissi alterativi in un corpus di favole naturali, scritte da esseri umani, con l'uso degli stessi suffissi in un corpus di favole sintetiche, generate da quattro modelli linguistici: due modelli americani (GPT-3.5 Turbo e GPT-4o mini) e due modelli europei (Teuken-7B e Minerva-7B).

Il lavoro è organizzato nel modo seguente: la prima parte (§ 2) mette in luce le principali differenze tra i suffissi alterativi dell'italiano e dell'inglese, confronta i quattro modelli linguistici utilizzati per generare il corpus di favole, prestando particolare attenzione alla quantità di dati di addestramento in lingua italiana, e formula quattro ipotesi di lavoro che permetteranno di valutare la naturalezza delle favole sintetiche; la seconda parte (§ 3) presenta il corpus di lavoro e i metodi di ricerca, mentre la terza parte (§ 4) descrive i principali risultati quantitativi ottenuti sull'uso dei suffissi alterativi nel corpus di favole naturali e sintetiche. Nelle conclusioni (§ 5) formuleremo un giudizio sulla naturalezza delle favole sintetiche alla luce dei risultati ottenuti in merito alle quattro ipotesi di lavoro e forniremo alcuni elementi per spiegare questi risultati tornando sulla questione di un possibile influsso dell'inglese.

[2] L'aggettivo *sintetico* si riferisce qui all'*output* dei modelli linguistici e si contrappone all'aggettivo *naturale*, impiegato per riferirsi alla comunicazione umana (per dettagli, cfr. De Cesare 2026). L'accezione dell'aggettivo è dunque diversa da quella che ha nella dicotomia *sintetico-analitico*, in cui fa riferimento alle caratteristiche morfosintattiche di una struttura linguistica.

2 *Italiano-inglese a confronto*

2.1 *Paradigmi di suffissi alterativi, produttività e contesti d'uso*

Nel campo che ci interessa, ovvero il paradigma dei suffissi alterativi, l'italiano e l'inglese presentano numerose differenze importanti (per dettagli, cfr. Dressler & Merlini Barbaresi 1994, § 3.3; Merlini Barbaresi 2002; Cacchiani 2011). La prima differenza riguarda il numero di suffissi alterativi a disposizione nelle due lingue. Come mostra l'elenco (peraltro non esaustivo[3]) nella Tab. 1, il paradigma dei suffissi alterativi è molto più ampio in italiano che in inglese. Questo vale per tutte e tre le categorie semantiche individuate: i diminutivi, gli accrescitivi e i peggiorativi.

Tab. 1. Paradigma di suffissi alterativi in italiano e inglese (Merlini Barbaresi 2004: 265; Cacchiani 2011: 759)

	italiano	inglese
diminutivi	***-etto***, ***-ello***, *-iccio*, ***-ino***, *-occio*, *-onzolo*, ***-uccio***	***-ette,*** *-ey*, ***-ie/-y, -let,*** *-ling*
accrescitivi	*-asso*, ***-one***, ***-otto***, *-ozzo*	-
peggiorativi	***-accio***, *-astro*, *-ucolo*	-

Come si può osservare a partire dalla Tab. 1, l'inglese possiede pochi suffissi alterativi, tutti appartenenti alla classe dei diminutivi (Cacchiani 2011: 759, 780; Dressler & Merlini Barbaresi 1994: 112). Per esprimere i tratti semantici veicolati dalle due categorie di suffissi mancanti (quelli accrescitivi e peggiorativi), l'inglese ricorre ad altre strategie, tra cui la prefissazione (*riccone* > *mega**rich***, *cervellone* > *mega**mind***; es. tratti da Cacchiani 2011: 777) e soprattutto la modificazione lessicale tramite aggettivi (*filmaccio* > *bad movie*, *giallaccio* > *awful yellow*; cfr. Cacchiani 2011: 777 e 779) o avverbi (*benone* > *very well*).

Una seconda differenza tra italiano e inglese riguarda la produttività dei suffissi alterativi disponibili nel sistema delle due lingue, misurata in base alla capacità di creare forme nuove e alla flessibilità di combinarsi con basi appartenenti a parti del discorso diverse. I suffissi più produttivi dell'italiano sono quelli in grassetto nella Tab. 1 (Merlini Barbaresi 2004: 266); la stessa cosa vale per l'inglese (Dressler & Merlini Barbaresi 1994: 112–116). Rispetto all'italiano, i diminutivi inglesi sono comunque molto meno produttivi (Dressler & Merlini Barbaresi 1994: 112, 402). Questo spiega perché, anche in questa categoria semantica, l'inglese ricorre a un ventaglio di altre strategie linguistiche per esprimere quello che è veicolato in italiano dai suffissi diminutivi (Dressler & Merlini Barbaresi 1994: 178, 403). La lista include, tra altre possibilità, la reduplicazione (Merlini Barbaresi & Dressler 2020), l'uso di aggettivi e avverbi che esprimono il tratto semantico 'piccolo' (*(a) little, small, tiny, a bit*, come in *pezzettino* > *little piece*) o ancora, nel dominio dei peggiorativi, il suffissoide *x-head*, come in *crackhead* 'fumatore di cocaina; idiota' (Mattiello & Sánchez Fajardo 2025: 2). Le strategie sono *in primis* di natura analitica.

[3] L'elenco non è esaustivo soprattutto in merito all'italiano. La Tab. 1 non include, per esempio, il suffisso elativo *-issimo* e molte altre forme elencate in Merlini Barbaresi (2004: 265). In un ulteriore studio, più approfondito dal punto di vista qualitativo, terremo anche conto di altri valori valutativi dei suffissi, in particolare di quelli proposti nella classificazione di Grandi (2002).

Comune a entrambe le lingue è invece la manifestazione privilegiata dei suffissi alterativi nella comunicazione incentrata sul bambino, in cui il bambino funge da emittente e/o destinatario del messaggio (Dressler & Merlini Barbaresi 1994: 112 parlano di "child-centered speech") o in qualsiasi situazione che faccia riferimento, anche metaforicamente, a un mondo infantile. Rispetto ai suffissi italiani, che compaiono anche in altre situazioni comunicative, quelli inglesi sono praticamente circoscritti alla comunicazione con i bambini. Basta considerare il caso di *-ie/-y*, il suffisso inglese in assoluto più produttivo (Dressler & Merlini Barbaresi 1994: 112). Oltre agli antroponimi, in particolare ai nomi di battesimo (*Sus-ie, Bill-y*), che si usano però anche nella comunicazione tra adulti, il diminutivo si combina con nomi caratteristici della comunicazione che coinvolge bambini piccoli: quelli denotanti rapporti di parentela (*mumm-y; dadd-y*) e animali (*bunn-y, mous-ie, dogg-ie*).

Nell'ambito della comunicazione con i bambini, rientra naturalmente *in primis* il cosiddetto *baby talk*, la varietà del parlato che "designa il modo di rivolgersi a bambini in tenera età da parte degli adulti che si prendono cura di loro" (Bernini 2010: 139); a livello morfosintattico questa varietà è caratterizzata dalla "diffusa presenza di suffissi diminutivi coi nomi (*acquina*, *cagnolino*, *lattuccio*, *orsettino*)" (*ibid*.), come in *Facciamo il bagnetto al nostro tesoruccio?* (es. tratto da Gaeta 2010: 372). Per quanto riguarda lo scritto, la manifestazione più prototipica della comunicazione incentrata sui bambini è la letteratura per l'infanzia (Dressler & Merlini Barbaresi 1994: 385, 386). In questo studio, ci concentreremo sul genere della favola (da distinguere da quello della fiaba), caratterizzato da tre elementi: è di regola scritta da un autore o da un'autrice, e nasce dunque come testo scritto; ha generalmente come protagonisti animali antropomorfici; e si chiude con un apologo morale, con il quale "si vuole insegnare un comportamento o condannare un vizio umano" (Detti 2005).

2.2 *Modelli linguistici*

I modelli linguistici di grandi dimensioni (o LLM) sono strumenti basati su algoritmi di intelligenza artificiale generativa, in grado di produrre *output* testuale *ex novo*, a partire da un'istruzione chiamata *prompt*. I modelli linguistici disponibili sono oramai centinaia. Tra quelli più noti, anche per le loro abilità linguistiche avanzate, si possono citare i modelli multilingue delle famiglie GPT (sviluppati da OpenAI), Gemini (Google), DeepSeek (DeepSeek), Claude (Anthropic) e Llama (Meta). Si tratta di modelli chiusi (o *closed-source*), sui quali mancano informazioni cruciali, a cominciare dai loro dati di addestramento.

Molto importanti sono poi i modelli multilingue sviluppati in Europa per generare testi di qualità nelle principali lingue europee (inglese, tedesco, francese, spagnolo, italiano, portoghese, neerlandese ecc.), come Teuken-7B (OpenGPT-X). Esistono anche modelli sviluppati per generare testi di qualità in lingua italiana: è il caso della famiglia dei modelli Minerva creati alla Sapienza di Roma. A differenza della maggior parte dei modelli americani, molti modelli europei sono aperti (o *open-source*): sono dunque noti il numero dei loro parametri di addestramento, i dati sui quali sono stati addestrati, la forma della loro architettura interna ecc. (cfr. per esempio Navigli, Conia & Ross 2023).

Nel presente studio vogliamo valutare la naturalezza dell'*output* di quattro modelli linguistici di grandi dimensioni: GPT-3.5 Turbo e GPT-4o mini – due modelli americani chiusi della famiglia GPT sviluppati da OpenAI – così come Teuken-7B di OpenGPT-X e Minerva-7B della Sapienza di Roma – due modelli aperti sviluppati rispettivamente in Germania e in Italia. La Tab. 2 riporta alcuni dati relativi a quattro fattori che incidono in modo importante sulla qualità dell'*output* generato (per approfondimenti, cfr. Navigli, Conia, Ross 2023), ovvero (i) il numero dei parametri di addestramento; (ii) la dimensione del corpus di addestramento iniziale, ovvero dei dati testuali usati nella fase di creazione del modello di base (ingl. *foundation model*); (iii) la dimensione del corpus di addestramento iniziale in lingua italiana rispetto a quello in lingua inglese; infine, (iv) le varietà di lingua e la gamma di generi testuali rappresentati nei corpora di addestramento. I fattori considerati non sono noti per i modelli americani. La Tab. 2 riporta dunque i dati relativi a GPT-3 descritti in Brown et al. 2020 e Johnson et al. 2022, partendo dal presupposto che i dati relativi ai modelli successivi (GPT-3.5 Turbo, ma anche GPT-4o mini) sono stati ampliati.

*Tab. 2. Fattori per valutare la qualità dell'*output *in italiano*

Modelli linguistici	Numero di parametri di addestramento	Dimensione del corpus di addestramento (in *token*)	Dimensione del corpus di addestramento in italiano vs inglese (in *token* o *parole*[4])	Varietà di lingua; generi testuali
GPT-3.5 Turbo	> 175 miliardi (GPT-3)	> 500 miliardi (GPT-3: Brown et al. 2020: 9)	It. > 1 miliardo Ingl. > 181 miliardi (GPT-3: OpenAI 2020)	Testi scritti: Web (Common Crawl), libri, Wikipedia
GPT-4o mini	> 175 miliardi (GPT-3)	> 500 miliardi (GPT-3: Brown et al. 2020: 9)	It. > 1 miliardo Ingl. > 181 miliardi (GPT-3: OpenAI 2020)	Testi scritti: Web (Common Crawl), libri, Wikipedia
Teuken-7B[5]	7 miliardi	4.000 miliardi	It. 189 miliardi Ingl. 1.668 miliardi	Testi scritti: Web (Common Crawl), testi educativi (Fine-Web EDU), testi accademici, finanziari e tecnici
Minerva-7B	7 miliardi	ca. 2.500 miliardi[6]	It. ca. 1.100 miliardi Ingl. ca. 1.300 miliardi	Testi scritti: Web (Common Crawl), libri (Gutenberg), Wikipedia, testi giuridici (EurLex, Gazzetta ufficiale), articoli di giornale

I dati della Tab. 2 mostrano che ci sono differenze molto significative tra i quattro modelli linguistici. I due modelli americani sono chiaramente molto più grandi dei due modelli europei per quanto riguarda il numero

[4] Per quanto riguarda GPT-3, abbiamo solo informazioni relative ai dati di addestramento sotto forma di *parole*, ovvero prima della loro tokenizzazione. Secondo OpenAI, 75 parole corrispondono a circa 100 *token* per l'inglese (https://platform.openai.com/tokenizer) (Ultimo accesso 08/01/2025). La tokenizzazione, tuttavia, varia a seconda della lingua e della rispettiva ricchezza morfologica (cfr. Navigli, Conia, Ross 2023).
[5] Si veda la tabella 42 in Ali et al. 2024.
[6] Per la precisione, i dati di addestramento ammontano a 2.480 miliardi di *token* (cfr. Orlando et al. 2024: 708).

di parametri di addestramento (> 175 *vs* 7 miliardi). I due modelli europei sono invece uguali da questo punto di vista. Per quanto riguarda il secondo fattore, non è difficile ipotizzare che il corpus di addestramento dei due modelli americani sia molto più grande di quello utilizzato per addestrare la versione precedente, GPT-3 (composto da 500 miliardi di *token*), superando forse anche la dimensione del corpus di addestramento dei due modelli europei, tra i quali in questo caso c'è una differenza importante: Teuken-7B è stato addestrato su un corpus di 4.000 miliardi di *token*, Minerva-7B su uno di "soli" 2.500 miliardi di *token*.

Il terzo fattore che vogliamo testare in questa sede riguarda la mole dei dati di addestramento in lingua italiana rispetto a quelli in lingua inglese. Nei due modelli americani, l'inglese prevale in modo massiccio (it. 1 *vs* ingl. 181 miliardi di parole), e può dunque essere considerato come sovrarappresentato; le due lingue sono un po' più bilanciate nel caso di Teuken-7B (it. 189 *vs* ingl. 1.668 miliardi di *token*), mentre sono rappresentate in modo praticamente equo nei dati di addestramento di Minerva-7B (1.100 miliardi di *token* per l'it. e 1.300 per l'inglese), sviluppato per produrre testi di più elevata qualità in lingua italiana.

Ci sono probabilmente differenze importanti tra i quattro modelli linguistici anche per quanto riguarda l'ultimo dato. Non è però chiaro quali varietà di lingua e generi testuali siano rappresentati nei dati di addestramento dei modelli da analizzare. Partiamo dunque anche qui da una serie di presupposti: i testi impiegati per addestrare i quattro modelli sono perlopiù rappresentativi dello scritto (più che del parlato) e includono un'ampia gamma di generi testuali, appartenenti a diverse tipologie (testi giuridici, testi giornalisti, testi enciclopedici, comunicazione sui social media). Non sappiamo invece se i dati inclusi nella categoria "libri" comprendano anche testi appartenenti al genere della favola. I quattro modelli sui quali si concentra l'analisi sono però perfettamente in grado di produrre *output* appartenenti a questo genere testuale (§ 3).

2.3 *Ipotesi di lavoro*

Alla luce di quanto esposto sopra (nei §§ 2.1 e 2.2), ci aspettiamo di osservare alcune differenze significative nell'uso dei suffissi alterativi nelle favole naturali e sintetiche scritte in italiano, così come tra le favole sintetiche generate in italiano dai quattro modelli linguistici. In particolare, tenendo conto del fatto che i suffissi alterativi sono poco numerosi e meno produttivi in inglese, ci aspettiamo che il loro impiego sia globalmente diverso nelle favole generate da modelli linguistici i cui dati di addestramento includono molti testi scritti in inglese. Ci aspettiamo anche di osservare una minore ricchezza di forme proprie della lingua italiana (nella fattispecie di suffissi alterativi e di forme con suffissi alterativi) nelle favole generate da modelli allenati su un corpus di testi in cui l'inglese è sovrarappresentato, come è il caso dei due modelli linguistici americani. Le ipotesi di lavoro che vogliamo verificare sono più precisamente le seguenti:

I. Rispetto alle favole naturali, quelle sintetiche includono un paradigma più ridotto di suffissi alterativi;
II. Rispetto alle favole naturali, quelle sintetiche includono un numero più ridotto di lemmi con suffissi alterativi;
III. Rispetto alle favole naturali, quelle sintetiche includono un numero più ridotto di occorrenze (*token*) di parole con suffissi alterativi;

IV. Rispetto alle favole sintetiche generate dai LLM addestrati con una maggiore quantità di *token/parole* in lingua italiana (Teuken-7B e soprattutto Minerva-7B), quelle generate dai LLM addestrati su corpora in cui l'italiano è numericamente meno/poco rappresentato (GPT-3.5 Turbo e GPT-4o mini) includono complessivamente un paradigma meno ampio di suffissi alterativi, meno lemmi e meno occorrenze (*token*) di parole con un suffisso alterativo. Di conseguenza, le favole generate dai modelli americani sono anche meno simili a quelle scritte da esseri umani.

3 *Corpus e metodi*

3.1 *Corpus design*

Il corpus usato in questa sede è stato costruito appositamente per testare le ipotesi formulate nel § 2.3. Esso si compone di due sottocorpora comparabili di favole (HUM e LLM).

Il primo sottocorpus (HUM) è costituito da 200 favole naturali redatte in lingua italiana (cfr. Tab. 3). Le favole sono state pubblicate sul web prima di novembre 2022, ovvero prima della comparsa di LLM ad uso commerciale.[7] I testi si possono a loro volta suddividere in due sottogruppi: da un lato, sono state accolte 69 *Favole al telefono* di Rodari, rappresentative della scrittura d'autore; dall'altro, abbiamo incluso favole redatte da scrittori non professionisti, ovvero redatte da utenti del web.[8]

Tab. 3. Sottocorpus favole naturali (HUM)

Sottocorpus HUM	**N. di favole**	**N. di parole grafiche**	**Data di pubblicazione**
Le Favole di Morfeus	19	8.461	2018
Letture per bambini	18	10.775	2016
Storie brevi della buonanotte	37	19.458	2015-2022
Favole per bambini	6	1.056	2020
15 brevi racconti per bambini sulla natura	10	1.299	2021
Falcone, Favole della buonanotte	10	5.628	2016
Rodari, Favole al telefono	69	26.052	1995 [1962]
Falchi, Fiabe illustrate per bambini	10	7.110	2011
Ianniciello, Racconti e favole per bambini	4	787	2017
Coletta (a cura di), C'erano una volta… le Favole	14	14.161	2008
Romani, Favole per bambini	3	2.701	2020-2021
Tot.	**200**	**97.488**	

Il secondo sottocorpus (LLM) comprende 336 favole generate con quattro modelli linguistici: GPT-3.5 Turbo, GPT-4o mini, Minerva-7B e Teuken-7B (cfr. Tab. 4). La varietà dei modelli serve a valutare l'incidenza dei dati di addestramento sulla qualità dell'*output*. Sebbene al momento in cui si scrive ChatGPT sia giunto al

[7] Le favole sono indicate con il titolo della raccolta, preceduto dall'eventuale cognome di chi ha curato la raccolta o scritto la storia. Sono state raccolte, laddove possibile, solo le favole firmate per garantirne l'originalità.
[8] Un'analisi delle occorrenze di parole che terminano con i suffissi selezionati nei due sottogruppi di testi non rivela differenze sostanziali: la frequenza relativa delle parole che terminano con uno dei suffissi alterativi è per entrambi di circa 6 ogni 1.000 parole. Di primo acchito non si rilevano quindi differenze dovute né alla scrittura d'autore, né a questioni diacroniche.

modello 5.5, sono stati selezionati modelli precedenti poiché, in linea con i dati riassunti nella Tab. 2, ipotizziamo che questi siano più piccoli a livello di parametri e di dati di addestramento e che siano dunque maggiormente paragonabili con i due modelli europei. I testi sono stati generati attraverso l'API di GPT e Teuken-7B, mentre per Minerva-7B è stato usato LM Studio, che permette di impostare i parametri *temperatura* e *system prompt*. I *prompt* utilizzati sono volutamente semplici e impliciti per non distorcere i dati relativi alla produzione di suffissi alterativi; l'indicazione del tipo di testo e del genere testuale fornisce infatti già l'orientamento necessario sui tratti tipici che dovrebbero comparire nell'*output*:

- System prompt: "Il tuo compito è generare testi narrativi."
- User prompt: "Scrivi una favola per bambini (3-5 anni)." oppure "Scrivi una favola per bambini (6-10 anni)."[9]

Le favole sono state generate con una temperatura differente (la temperatura è il parametro che permette di modulare la probabilità di co-occorrenza degli elementi tokenizzati): in un primo ciclo di generazione (09/2025) è stata impostata una temperatura di 0.7, mentre i testi del secondo ciclo (11/2025) sono stati generati con la temperatura 0.6. Quest'ultima temperatura diminuisce la probabilità di co-occorrenza e consente di ottenere testi leggermente diversi.

Tab. 4. Sottocorpus favole sintetiche (LLM)

Sottocorpus LLM	**N. di favole**	**Temperatura**	**Data di generazione**
GPT-3.5 Turbo	42 42	0.7 0.6	09/2025 11/2025
GPT-4o mini	42 42	0.7 0.6	09/2025 11/2025
Teuken-7B	42 42	0.7 0.6	09/2025 11/2025
Minerva-7B	42 42	0.7 0.6	09/2025 11/2025
Tot.	**336**		

La Tab. 5 fornisce indicazioni sulla dimensione dei due sottocorpora (HUM e LLM) e delle loro sottoparti (i dati sono calcolati con Python dopo aver tokenizzato i testi con la libreria *Stanza*[10]). La maggiore quantità di testi sintetici (336) rispetto a quelli naturali (200) intende bilanciare i due sottocorpora poiché le favole generate sono tendenzialmente più brevi.

Tab. 5. Dimensione corpus di favole

Sottocorpus	**HUM**	**LLM**	**Totale**
N. di favole	200	336	536
N. di parole grafiche	97.488	97.721	195.209

[9] Le favole sono state generate per due fasce di età per analizzare, in uno studio futuro, la distribuzione dei suffissi in base all'età.
[10] Gli script Python sono stati scritti con l'aiuto di Gemini 3 pro.

I dati dei due sottocorpora (HUM e LLM) sono assolutamente comparabili (ca. 100.000 parole grafiche diverse per sottocorpus). Tuttavia, i testi generati presentano una lunghezza media costante (tra le 250 e le 450 parole), mentre le favole umane mostrano una forte eterogeneità, con oscillazioni tra le 75 e le 2.100 parole grafiche. Tale disparità può influenzare la distribuzione dei lemmi e dei *token*, poiché la maggiore lunghezza di un testo favorisce la ripetizione (*token*) a scapito della varietà lessicale (lemmi). Come evidenziano Gaeta & Ricca (2006: 58–65) la produttività di un affisso, intesa come la velocità con cui il lessico viene arricchito con nuovi *type*, decresce all'aumentare del corpus. In questa sede, tuttavia, si è scelto di fornire i dati basati sul totale delle parole dei due sottocorpora.

Per testare l'ipotesi (iv), forniamo le stesse indicazioni anche sui singoli LLM (cfr. Tab. 6).

Tab. 6. Dimensione sottocorpus LLM

Sottocorpus LLM	**GPT-3.5 Turbo**	**GPT-4o mini**	**Teuken-7B**	**Minerva-7B**
N. di favole	84	84	84	84
N. di parole grafiche	21.188	34.082	23.330	19.121

In questo caso, i modelli sono solo in parte comparabili perché le favole di GPT-4o mini sono mediamente più lunghe, risultando in un numero totale di parole più elevato rispetto a quello degli altri modelli (quasi il doppio di Minerva-7B e circa una volta e mezzo di GPT-3.5 Turbo e Teuken-7B).

3.2. *Suffissi alterativi analizzati e metodi per individuare le occorrenze (lemmi e token) nel corpus*

La nostra analisi riguarda l'uso di sette suffissi: *-ino, -etto, -ello* (diminutivi); *-one, -asso* (accrescitivi); *-otto, -ozzo* (polifunzionali: diminutivi o accrescitivi). L'analisi esclude dunque per il momento i suffissi peggiorativi, il suffisso elativo *-issimo* e tutti i suffissi che si uniscono ai verbi.

Per individuare le parole che terminano con questi suffissi ci siamo avvalse di un procedimento a tre tappe. In un primo momento, abbiamo utilizzato uno script di Python per cercare ed estrarre dal corpus tutte le parole terminanti con una delle forme elencate sopra (tenendo conto delle relative forme flesse) in modo automatico, ordinate in liste di lemmi e occorrenze (*token*).

In un secondo momento abbiamo individuato a mano le forme con suffissi alterativi nelle liste, eliminando i falsi positivi (come *vicino*, *emozione*, *notte*). Abbiamo inoltre escluso 554 nomi di animali usati nel senso di 'cucciolo/giovane di X' (come *cagnolino*). In questi casi, infatti, si tratta di forme lessicalizzate (cfr. a proposito Merlini Barbaresi 2004: 266; sulla questione dei nomi per indicare i giovani di animali si veda anche Grandi 2002: 57–68). In questa fase dell'analisi abbiamo consultato due risorse lessicografiche di riferimento: Il Nuovo De Mauro e il Vocabolario on line di Treccani. Dato l'ingente lavoro di revisione manuale, abbiamo adottato alcune scelte di comodo: 1) non abbiamo considerato l'alterazione dei nomi propri; ci siamo concentrate invece sui dati relativi ai nomi comuni, agli aggettivi, agli avverbi e ai numerali; 2) quando i nomi di animali erano più di 40 per lemma, ovvero *coniglietto* (126), *uccellino* (253), *pesciolino* (43) e *topolino* (62), li abbiamo esclusi in blocco. È interessante però notare che i *type* provengono soprattutto dai quattro LLM, ad esclusione di *pesciolino* che si trova solo in Teuken-7B e GPT-4o mini.

Dopo l'analisi individuale da parte delle autrici, i risultati sono stati confrontati per raggiungere un accordo sui casi dubbi. Nell'ultima tappa, abbiamo creato le liste di lemmi e occorrenze (*token*), su cui si basa la riflessione proposta nel § 4.

4 *Risultati*

In ciò che segue presentiamo i risultati quantitativi ottenuti confrontando in un primo momento i dati dei due sottocorpora (HUM vs LLM), nei quali si rileva una distribuzione di suffissi alterativi in tutti i testi di cui si compongono, sebbene con una frequenza diversa. Il confronto verte dapprima sul paradigma di suffissi e sul numero di lemmi con uno dei suffissi alterativi oggetto dell'analisi (§ 4.1); riportiamo poi il numero di occorrenze (*token*) di parole con uno dei suffissi ricercati (§ 4.2); in un secondo momento, riproduciamo la stessa analisi distinguendo però i risultati per modello linguistico (§ 4.3). Nelle tabelle proposte, i suffissi sono sempre presentati in ordine alfabetico.[11]

4.1 *HUM* vs *LLM a confronto: paradigma di suffissi e lemmi con suffissi alterativi*

Vediamo dapprima i dati relativi ai lemmi che terminano con uno dei sette suffissi alterativi nei due sottocorpora di favole (HUM *vs* LLM). I dati riportati nella Tab. 7 raggruppano dunque sotto un'unica forma parole con lo stesso suffisso ma con proprietà grammaticali diverse (per esempio, nel sottocorpus HUM, le quattro occorrenze di *sorellina* e le sei occorrenze di *sorelline* sono ricondotte al lemma *sorellina*).

Tab. 7. Frequenza assoluta (lemmi)

Suffisso	**HUM**	**LLM**	**Totale lemmi unici nel corpus**
-asso	0	0	0
-ello	15	4	17
-etto	67	21	77
-ino	106	19	109
-one	20	2	21
-otto	2	0	2
-ozzo	0	0	0
Tot.	**210 (93%)**	**46 (20%)**	**226[12] (100%)**

I dati della Tab. 7 permettono innanzitutto di fare un'osservazione importante sul paradigma di suffissi effettivamente utilizzati nei due sottocorpora. Nelle favole sintetiche compaiono solo quattro dei sette suffissi alterativi (nell'ordine di frequenza: *-etto, -ino, -ello, -one*), vale a dire uno in meno rispetto a quanto si trova nelle favole naturali (*-ino, -etto, -one, -ello* e *-otto*).

Un secondo dato importante riguarda la frequenza delle categorie di lemmi suffissati nei due sottocorpora. Nelle favole sintetiche predominano due categorie di lemmi suffissati (quelle uscenti in *-etto* e

[11] Le liste di lemmi e occorrenze sono disponibili in Mantovani & De Cesare 2026.
[12] I lemmi che si ripetono in entrambi i sottocorpora (HUM e LLM) non sono inclusi in questo dato.

in *-ino*), mentre nelle favole naturali predominano largamente le forme uscenti con un singolo suffisso (*-ino*). I dati riportati nell'ultimo rigo della Tab. 7 mostrano poi che le favole sintetiche sono associate a una varietà lessicale molto minore: esse contengono solo il 20% dei lemmi con suffissi alterativi trovati nell'intero corpus. Concretamente, le favole sintetiche includono solo 46 lemmi suffissati diversi (sul totale di 226), mentre se ne riscontrano ben 210 nelle favole naturali.

4.2 *HUM* vs *LLM a confronto: occorrenze di parole (*token*) con suffissi alterativi*

Passiamo ai dati relativi alla frequenza delle parole con suffissi alterativi misurata in termini di singole occorrenze alterate, ovvero di *token* (in questo caso, le quattro occorrenze di *sorellina* e le sei occorrenze di *sorelline* sono contate come 10 *token*). La Tab. 8 mostra la frequenza assoluta di ogni forma alterata nei due sottocorpora di favole (HUM *vs* LLM).

*Tab. 8. Frequenza assoluta (*token*)*

Suffisso	**HUM**	**LLM**	**Totale *token* nel corpus**
-asso	0	0	0
-ello	39	12	51
-etto	184	93	277
-ino	349	113	462
-one	26	2	28
-otto	4	0	4
-ozzo	0	0	0
Tot.	**602 (73%)**	**220 (27%)**	**822 (100%)**

I dati riportati nell'ultimo rigo della Tab. 8 permettono di osservare che, rispetto alle favole naturali, le favole sintetiche includono nettamente meno forme alterate (220 su 822, ovvero circa un quarto dei dati complessivi: 27%).[13] Nelle favole sintetiche spicca ancora una volta l'uso predominante di due suffissi: *-ino* (113 occ.) e *-etto* (93 occ.). Sono invece poco frequenti le forme in *-ello* (12 occ.) e decisamente molto rare quelle in *-one* (2 occ.). Nelle favole naturali, invece, predominano largamente le forme uscenti con il suffisso *-ino* (349 occ.), seguite, ma appunto da abbastanza lontano, da quelle terminanti in *-etto* (184 occ.). Hanno poi una discreta presenza sia le forme in *-ello* (39) che in *-one* (26 occ.).

Il dato sulle forme uscenti in *-one* – poco presenti sia nelle favole sintetiche (2 occ.) sia in quelle naturali (26 occ.) – è alquanto sorprendente in quanto *-one* è "l'unico suffisso dell'italiano standard che formi alterati accrescitivi ed è ampiamente produttivo" (Merlini Barbaresi 2004: 287). Il suffisso permette infatti di alterare un'ampia tipologia di nomi, lavorando su tratti diversi, in particolare quelli relativi alla dimensione e

[13] È difficile prevedere il ruolo che potrebbe giocare la (minore) lunghezza dei testi generati sulla generazione di parole con un suffisso alterativo. Come già accennato, l'attivazione di un determinato suffisso alterativo in un testo potrebbe forse favoreggiare il suo riuso nel resto del testo. Si tratta di una questione interessante, che andrebbe approfondita in altra sede.

all'esagerazione (*librone*); esso può anche essere usato con aggettivi (*grandone*) e, più raramente, con avverbi (*benone*; per dettagli, cfr. Merlini Barbaresi 2004: 287–288).

4.3 *LLM a confronto: paradigma di suffissi, lemmi e occorrenze di parole con suffissi alterativi*

Vediamo infine la distribuzione dei suffissi nel sottocorpus LLM, vale a dire nei dati generati da ciascun modello linguistico considerato, soffermandoci dapprima sui lemmi.

Tab. 9. Risultati quantitativi (lemmi)

Suffisso	**GPT-3.5 Turbo**	**GPT-4o mini**	**Teuken-7B**	**Minerva-7B**	**Totale lemmi unici nel corpus**
-asso	0	0	0	0	0
-ello	3	1	1	0	4
-etto	11	6	8	6	21
-ino	5	8	9	4	19
-one	1	0	1	0	2
-otto	0	0	0	0	0
-ozzo	0	0	0	0	0
Tot.	**20 (44%)**	**15 (33%)**	**19 (41%)**	**10 (22%)**	**46[14] (100%)**

La Tab. 9 mostra che le favole generate da Minerva-7B includono la minore varietà di suffissi: i suffissi presenti sono infatti solo due (*-etto* e *-ino*); le favole generate con GPT-4o mini includono invece tre suffissi (si aggiunge *-ello*), mentre quelle di GPT-3.5 Turbo e Teuken-7B includono quattro suffissi (compaiono anche forme uscenti in *-one*). In tutti i modelli testati si osserva comunque una predominanza dei suffissi uscenti in *-etto* e *-ino*.

L'ultimo rigo della tabella mostra inoltre che in Minerva-7B si registra la minore varietà lessicale nei lemmi con suffissi alterativi, che ammontano a solo 10, seguito da GPT-4o mini (15 lemmi), Teuken-7B (19 lemmi) e GPT-3.5 Turbo (20 lemmi).

Riportiamo ora i dati relativi ai *token* per ciascun LLM.

*Tab. 10. Risultati quantitativi (*token*)*

Suffisso	**GPT-3.5 Turbo**	**GPT-4o mini**	**Teuken-7B**	**Minerva-7B**	**Totale *token* nel corpus**
-asso	0	0	0	0	0
-ello	4	2	6	0	12
-etto	44	6	27	16	93
-ino	11	25	68	9	113
-one	1	0	1	0	2
-otto	0	0	0	0	0
-ozzo	0	0	0	0	0
Tot.	**60 (27%)**	**33 (15%)**	**102 (46%)**	**25 (11%)**	**220 (100%)**

[14] I lemmi che si ripetono nei quattro sottocorpora non sono inclusi in questo dato.

La Tab. 10 mostra che, fra i quattro LLM testati, Minerva-7B si colloca di nuovo in ultima posizione per numero di parole terminanti con uno dei suffissi alterativi considerati (25 occ.). Segue, da relativamente vicino, GPT-4o mini (33 occ.). In questo caso è importante notare che il sottocorpus di favole generato con questo modello contiene un numero maggiore di parole rispetto ai sottocorpora generati con gli altri LLM: il dato relativamente basso di *token* con uno dei suffissi in questione (33 occ., appunto) è dunque degno di nota (cfr. anche la Tab. 11). Al terzo posto si posiziona GPT-3.5 Turbo (con 60 occ.). Teuken-7B offre il maggior numero di parole con uno dei suffissi in questione (102 su 220, ovvero quasi la metà dei dati relativi all'intero corpus).

Confrontando la forma dei suffissi nelle favole generate dai quattro modelli si osserva poi un dato in parte inatteso: mentre in GPT-3.5 Turbo e Teuken-7B sono più frequenti le forme che terminano con il suffisso *-ino* (25 e 68 occ.), in Minerva-7B e GPT-4o mini predominano invece le forme uscenti con il suffisso *-etto* (rispettivamente 16 e 44 occ.). Nelle favole generate dai quattro modelli sono poi poco presenti o del tutto assenti le forme in *-ello*, mentre le forme in *-one* sono molto rare e si trovano solo nelle favole generate da GPT-3.5 Turbo e Teuken-7B.

5 *Conclusioni*

Alla luce dei risultati ottenuti, possiamo affermare che le prime tre ipotesi formulate nel § 2.3 sono verificate: rispetto alle favole naturali, che includono parole uscenti con 5 dei 7 suffissi analizzati (*-ino, -etto, -ello, -one, -otto*; sono assenti le parole uscenti in *-asso* e *-ozzo*), quelle sintetiche includono un paradigma più ridotto di suffissi alterativi (sono presenti solo 4 suffissi: *-ino, -etto, -ello, -one*); le favole sintetiche presentano poi un numero molto più ridotto di lemmi con i suffissi alterativi analizzati (circa un quinto) e un numero più ridotto di occorrenze (*token*) con uno dei suffissi in questione (circa un quarto).

Per valutare la quarta ipotesi è necessario considerare i dati proposti nella Tab. 11, che riassume i risultati ottenuti per tutti e tre i parametri analizzati. Nel secondo rigo abbiamo anche aggiunto il dato relativo al suffisso più frequente nei cinque sottocorpora.

*Tab. 11. Riepilogo dei risultati ottenuti: favole naturali a confronto con le favole generate da quattro LLM (*frequenza normalizzata su 10.000 parole)*

Fattori	**HUM**	**GPT-3.5 Turbo**	**GPT-4o mini**	**Teuken-7B**	**Minerva-7B**
N. suffissi	5	4	3	4	2
Suffisso più frequente	*-ino*	*-etto*	*-ino*	*-ino*	*-etto*
N. lemmi*	22	9	4	8	5
N. *token**	62	28	10	44	13

I dati della Tab. 11 mostrano che la quarta e ultima ipotesi è verificata solo in parte. In effetti, da una parte si osserva che i risultati migliori – da intendere qui come più vicini a quelli del sottocorpus di favole naturali – sono quelli di Teuken-TB, allenato su un corpus di addestramento in cui l'italiano ha un peso maggiore rispetto ai LLM americani (cfr. Tab. 2). I risultati di Teuken-7B possono essere dovuti ad almeno due fattori: la presenza

di testi redatti in lingue in cui i suffissi alterativi sono produttivi (spagnolo, portoghese, rumeno, tedesco) e l'uso di un tokenizzatore non anglocentrico, ma adatto per corpora multilingue (Ali et al. 2024).

D'altra parte, però, i risultati relativi al modello americano GPT-3.5 sono più vicini a quelli ottenuti nel corpus di favole naturali che non a quelli prodotti da Minerva-7B (addestrato su un corpus in cui l'italiano rappresenta il 50% dei dati). I risultati del modello americano sono migliori da tre punti di vista: (i) rispetto alle forme di suffisso effettivamente prodotte nei testi generati (4 *vs* 2); (ii) rispetto ai lemmi generati con un suffisso alterativo (9 *vs* 5 per 10.000 parole) e (iii) rispetto al numero di occorrenze di parole con un suffisso alterativo (28 *vs* 13 occ., sempre per 10.000 parole). I risultati, inattesi, relativi a Minerva-7B potrebbero spiegarsi con il fatto che il modello che abbiamo testato in questa sede è quantizzato; si tratta infatti di una versione compressa e dunque meno performante.[15]

Sorprendente è anche la preferenza per il suffisso *-etto*, nelle favole naturali meno produttivo di *-ino*, di due LLM (GPT-3.5 e Minerva-7B). A ben guardare il suffisso è impiegato con una certa frequenza soprattutto nelle favole generate da GPT-3.5 (11 lemmi dei 21 inclusi nel sottocorpus di favole sintetiche; e 44 occorrenze sulle 93 riscontrate nelle favole generate). Questo dato, tuttavia, non è immediato da spiegare senza un'analisi qualitativa dei lemmi generati, che per motivi di spazio non possiamo però offrire in questa sede. In un futuro lavoro sarebbe anche interessante soffermarsi sulla creatività dei LLM nel generare parole con suffissi alterativi, osservando per esempio se, e quanto frequentemente, i LLM producono casi con doppio suffisso.

I risultati ottenuti in questa sede confermano quanto già osservato in altri studi: i testi sintetici sono diversi dai testi naturali. Nel caso indagato, possiamo costatare in particolare che le favole sintetiche presentano una quantità molto minore di suffissi e di parole con suffissi alterativi, caratteristiche del genere testuale in questione. Questo risultato potrebbe essere dovuto a un'anglicizzazione indiretta e non chiaramente visibile (cfr. Johnson et al. 2022) dei testi generati, addestrati su una mole importante di testi in inglese. Esso potrebbe però anche spiegarsi con il funzionamento dei modelli linguistici, basati su algoritmi statistico-probabilistici che premiano le forme più frequenti nei dati di addestramento. Per capire meglio i fattori che entrano in gioco nella generazione dei suffissi alterativi abbiamo bisogno di condurre altri studi, in particolare che tengano conto dell'uso dei suffissi alterativi in un corpus di favole naturali e sintetiche scritte in inglese.

Anna-Maria De Cesare
Giulia Mantovani
Technische Universität Dresden
Wiener Straße 48, 01219 Dresden
anna-maria.decesare@tu-dresden.de
giulia.mantovani@tu-dresden.de

[15] Ringraziamento Luca Moroni (Università Sapienza di Roma) per questa osservazione.

Bibliografia

Ali, Mehdi & Fromm, Michael & Thellmann, Klaudia et al. 2024. Teuken-7B-Base & Teuken-7B-Instruct: Towards European LLMs. *arXiv*. (https://doi.org/10.48550/arXiv.2410.03730)

Antonelli, Giuseppe. 2024. L'italiano 'autentico' dell'intelligenza artificiale. *AggiornaMenti* 5. 6–12.

Bernini, Giuliano. 2010. Baby talk. In *Enciclopedia dell'italiano* (dir. R. Simone), 139–140. Roma: Istituto della Enciclopedia Treccani.

Brown, Tom B. & Mann, Benjamin & Ryder, Nick et al. 2020. Language Models are few-shot learners. *arXiv*. (https://doi.org/10.48550/arXiv.2005.14165)

Cacchiani, Silvia. 2011. Intensifying affixes across Italian and English. *Poznań Studies in Contemporary Linguistics* 47(4). 758–794. (https://doi.org/10.2478/psicl-2011-0038)

Cicero, Francesco. 2023. L'italiano delle intelligenze artificiali generative. *Italiano LinguaDue* 15(2). 733–761. (https://doi.org/10.54103/2037-3597/21990).

Cicero, Francesco. 2025. Esercizi di stile: La narrativa per l'infanzia delle intelligenze artificiali. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 2(2). 1–21. (https://doi.org/10.62408/ai-ling.v2i2.19)

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.

De Cesare, Anna-Maria. 2026. *L'italiano sintetico dell'intelligenza artificiale generativa.* Firenze: Cesati.

Detti, Ermanno. 2005. Fiaba. (https://www.treccani.it/enciclopedia/fiaba_(Enciclopedia-dei-ragazzi)/) (Ultimo accesso 20/11/2025).

Dressler, Wolfgang U. & Merlini Barbaresi, Lavinia. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin/New York: De Gruyter.

Falcone, Angela. 2016. *Favole della buonanotte*. Du Lac et Du Parc Grand Resort.

Gaeta, Livio. 2010. Diminutivo. In *Enciclopedia dell'italiano* (dir. R. Simone), 371–373. Roma: Istituto della Enciclopedia Treccani.

Gaeta, Livio & Ricca, Davide. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44/1. 57–89.

Grandi, Nicola. 2002. *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: FrancoAngeli.

Ianniciello, Riccardo. 2017. *Racconti e favole per bambini*. Canterano: Gioacchino Onorati.

Il Nuovo De Mauro. Dizionario italiano De Mauro. (https://dizionario.internazionale.it/) (Ultimo accesso 12/01/2026).

Johnson, Rebecca L. & Pistilli, Giada & Menédez-González, Natalia et al. 2022. The Ghost in the Machine has an American accent: value conflict in GPT-3. *arXiv*. (https://doi.org/10.48550/arXiv.2203.07785)

Mantovani, Giulia. In stampa. Forme sintetiche e forme analitiche per i diminutivi dei nomi: un confronto tra favole naturali e favole generate dall'intelligenza artificiale. In Lala, Letizia & Castro, Enrico & Garzonio, Jacopo (a cura di), *Il nome in italiano. Morfologia, sintassi, testualità*. Firenze: Cesati.

Mantovani, Giulia & De Cesare, Anna-Maria. 2026. Dati per lo studio *Valutare la naturalezza dell'output prodotto dall'IA generativa: uno studio empirico sull'uso dei suffissi alterativi in un corpus di favole sintetiche e scritte da esseri umani*, in stampa in "Linguistica e Filologia", [Dataset]. *Zenodo*. (https://doi.org/10.5281/zenodo.20054868)

Mattiello, Elisa & Sánchez Fajardo, José A. 2025. The evaluative suffixoid-*head* and its Spanish and Italian equivalents. *Folia Linguistica*. (https://doi.org/10.1515/flin-2025-0063)

Merlini Barbaresi, Lavinia. 2002. Morfopragmatica: Due paradigmi morfologici a confronto. In Porcelli, Gianfranco & Maggioni, Maria Luisa & Tornaghi, Paola (a cura di), *Due codici a confronto*. *Per un'analisi contrastiva dei sistemi linguistici inglese e italiano (Atti del Convegno di Brescia, 28-30 marzo 1996)*, 109–140. Brescia: Editrice La Scuola.

Merlini Barbaresi, Lavinia. 2004. Alterazione. In Grossmann, Maria & Rainer, Franz (a cura di), *La formazione delle parole in italiano*, 264–292. Tübingen: Max Niemeyer.

Merlini Barbaresi, Lavinia & Dressler, Wolfgang U. 2020. Pragmatic explanations in morphology. Word knowledge and word usage. In Pirrelli, Vito & Plag, Ingo & Dressler, Wolfgang U. (eds), *Word Knowledge and Word Usage. A Cross-Disciplinary Guide to the Mental Lexicon*, 405–451. Berlin/Boston: De Gruyter.

Navigli, Roberto & Conia, Simone & Ross, Björn. 2023. Biases in Large Language Models: Origins, Inventory, and Discussion. *ACM J. Data Inform. Quality* 15(2). 101–121. (https://doi.org/10.1145/3597307)

OpenAI. 2020. Languages by word count. (https://github.com/openai/gpt-3/blob/master/dataset_statistics/languages_by_word_count.csv) (Ultimo accesso 08/01/2026).
OpenGPT-X. Teuken 7B Instruct: Multilingual, open source models for Europe – instruction-tuned and trained in all 24 EU languages. (https://opengpt-x.de/en/models/teuken-7b/) (Ultimo accesso 12/01/2026).
Orlando, Riccardo & Moroni, Luca & Huguet Cabot, Pere-Lluís et al. 2024. Minerva LLMs: The first family of Large Language Models trained from scratch on Italian data. *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024)*. 707–719.
Rodari, Gianni. 1995 [1962]. *Favole al telefono*. Trieste: Edizioni EL.
Simone, Raffaele. 2010. Lingue romanze e italiano. In *Enciclopedia dell'italiano* (dir. R. Simone), 826–838. Roma: Istituto della Enciclopedia Treccani.
Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1(1). 1–24. (https://doi.org/10.62408/ai-ling.v1i1.14)
Treccani. Vocabolario on line. (https://www.treccani.it/vocabolario/) (Ultimo accesso 12/01/2026).
Vaswani, Ashish & Shazeer, Noam & Parmar, Niki et al. 2017. Attention is all you need. *arXiv*. (https://doi.org/10.48550/arXiv.1706.03762)