

Una lettura “a distanza” e in diacronia dei luoghi artificiali in un corpus di prosa letteraria italiana e tradotta (XIX- XXI secolo)

Floriana Carlotta Sciumbata

Dipartimento di Scienze Giuridiche, del Linguaggio, dell’Interpretazione e della Traduzione, Università di Trieste, Italia
fsciumbata@units.it

Abstract

I luoghi in cui si svolgono le opere letterarie riflettono i cambiamenti economici, politici e sociali e rappresentano un elemento rilevante nella definizione di correnti e generi. Il presente studio propone un’analisi quantitativa delle ambientazioni nella narrativa italiana e nella narrativa tradotta in italiano tra XIX e XXI secolo, un periodo caratterizzato da profonde trasformazioni storiche, sociali e letterarie. L’indagine si basa su un corpus di testi e utilizza un approccio di *distant reading* per individuare le ambientazioni attraverso liste lessicali costruite *ad hoc*. Dopo aver descritto il corpus e le scelte metodologiche, il contributo illustra gli strumenti impiegati e presenta i risultati dell’analisi. L’obiettivo è duplice: da un lato, analizzare la rappresentazione dello spazio narrativo; dall’altro, proporre una procedura di estrazione delle ambientazioni, applicabile anche ad altre categorie lessicali e a diverse tipologie testuali. I risultati consentono di delineare un profilo delle ambientazioni, intese come luoghi artificiali, nella narrativa italiana degli ultimi due secoli e di metterle a confronto con quelle presenti nelle opere straniere tradotte in italiano.

Parole chiave: distant reading; luoghi artificiali; studio diacronico; narrativa italiana; corpus letterario

The places in which literary works are set reflect economic, political, and social changes and represent a relevant element in the definition of movements and genres. This study presents a quantitative analysis of settings in Italian narrative and in works translated into Italian between the 19th and 21st centuries, a period marked by profound historical, social, and literary transformations. The analysis is based on a corpus of texts and adopts a distant reading approach to identify settings through ad hoc lexical lists. After describing the corpus and the methodological choices, the article outlines the tools employed and presents the results of the analysis. The aim is twofold: on the one hand, to examine the representation of narrative space; on the other, to propose a procedure for the extraction of settings that can be applied to other lexical categories and different types of texts. The results make it possible to outline a picture of settings, understood as artificial places, in Italian narrative over the last two centuries and to compare them with those found in works translated into Italian.

Keywords: distant reading; artificial places; diachronic study; Italian fiction; literary corpus

1. Introduzione

Il presente lavoro dà conto dei risultati di una ricerca esplorativa di tipo prevalentemente quantitativo, condotta con i metodi della linguistica dei corpora e dell'analisi automatica dei dati testuali, che mira a estrarre il lessico delle ambientazioni da un corpus di letteratura italiana e tradotta pubblicata tra il XIX e il XXI secolo. Le ambientazioni costituiscono un punto cardine della letteratura: non sono solo lo spazio in cui si muovono i personaggi e si svolgono le trame, ma, insieme a chi li occupa,¹ sono anche lo specchio di cambiamenti economici, politici e sociali e diventano un elemento cruciale per la definizione di correnti e generi. Si pensi alla contrapposizione di luoghi rurali e urbani,² all'emergere di nuovi contesti, come la fabbrica, o all'evoluzione e al declino di altri, come la corte o il castello. Con "luoghi artificiali" si intendono in questo lavoro quelli costruiti o organizzati dall'essere umano: edifici, strutture, spazi urbani e rurali di diversa natura, in contrapposizione ai luoghi naturali come boschi, fiumi o montagne. Questi luoghi non riflettono i cambiamenti storici in modo diretto o meccanico: la loro rappresentazione è mediata da convenzioni letterarie, scelte narrative o editoriali, e può registrare trasformazioni sociali con ritardi o anticipazioni rispetto alla realtà. Come afferma efficacemente Moretti nel suo *Atlante del romanzo europeo* (1997, 74), "quel che succede dipende dal dove succede", che riassume l'aspetto spaziale del concetto di "cronotopo", ripreso dalla fisica da Bachtin (1979, 231) per descrivere "l'interconnessione sostanziale dei rapporti temporali e spaziali dei quali la letteratura si è impadronita artisticamente". Anche Ryan, Foote e Azaryahu (2016) hanno mostrato come lo spazio narrativo non si limiti a fare da sfondo agli eventi, ma svolga funzioni molteplici: può essere oggetto di attenzione, portatore di significato simbolico, principio di organizzazione del testo e persino supporto fisico della narrazione.

La ricostruzione degli ambienti in cui ha luogo la narrativa³ è un argomento di grande interesse (cfr. Tally 2018), a cui però non sono stati dedicati studi linguistici ad ampio spettro e condotti con metodi quantitativi,⁴ soprattutto per quanto riguarda la letteratura italiana: l'analisi si limita spesso a luoghi (naturali o artificiali, reali o inventati) specifici o alle ambientazioni di singoli autori. A differenza di quanto avviene per altre tradizioni letterarie europee, per le quali esistono analisi delle ambientazioni anche su campioni ridotti, per l'italiano mancano studi sistematici di

¹ Cfr. i dati sugli agentivi presentati da Sciumbata, Nadalutti e Tringali (2021) e quelli sulle gerarchie ecclesiastiche, militari e nobiliari (Sciumbata 2023).

² Si rimanda al saggio di Williams (1973) sulla letteratura inglese, emblematicamente intitolato *The Country and the City*.

³ Un approccio che ha riscontrato un certo successo è quello delle rappresentazioni geografiche e cartografiche della letteratura (Moretti 1997; 2007).

⁴ Un'eccezione per l'area germanofona è rappresentata dagli studi di Grisot e Herrmann (2023; 2025), che analizzano la rappresentazione dello spazio nella narrativa svizzera in lingua tedesca contrapponendo ambienti urbani e rurali e includendo sia toponimi sia termini spaziali generici. In questi lavori le entità spaziali sono individuate tramite liste lessicali applicate al corpus e utilizzate principalmente per studiare la relazione tra spazio narrativo ed emozioni attraverso tecniche di *sentiment analysis* e, nel caso dello studio su *Heidi* (2025), anche tramite il confronto con annotazioni umane. Per un esempio di analisi quantitativa diacronica del lessico narrativo su larga scala applicata al romanzo britannico ottocentesco, con risultati che collegano cambiamenti lessicali a trasformazioni sociali, si vedano invece Heuser e Le-Khac (2012).

questo tipo.⁵ L'assenza di sistematicità si può ricondurre anche alla carenza di repertori completi e digitalizzati, come una lista lessicale di luoghi da estrapolare da un corpus sufficientemente grande e bilanciato, a sua volta difficile da comporre per i numerosi fattori da tenere in considerazione.

L'analisi presentata si propone come una prima ricognizione delle ambientazioni nella narrativa in italiano, a partire da un corpus definito nei paragrafi successivi. La trattazione si soffermerà prima di tutto sulle caratteristiche del corpus usato come “banco di prova”, composto da opere di narrativa scritte in italiano o tradotte tra il XIX e il XXI secolo, rappresentative di un periodo caratterizzato da profondi mutamenti, sia dal punto di vista storico sia da quello letterario. Sarà poi illustrato il processo di creazione delle liste lessicali, realizzate appositamente per l'estrazione, e gli strumenti e i metodi impiegati, che includono metodi quantitativi e qualitativi. In ultimo, verranno presentati i risultati e alcune osservazioni conclusive sui dati ottenuti. L'obiettivo è, in primo luogo, proporre un metodo per l'estrazione prevalentemente automatica delle ambientazioni che possa essere applicato anche ad altre categorie lessicali e a tipologie testuali diverse. I risultati attesi riguardano una progressiva riduzione delle ambientazioni rurali e religiose a favore di spazi urbani e privati, in linea con i processi di urbanizzazione e secolarizzazione che caratterizzano il periodo. Il secondo obiettivo della ricerca è tracciare un quadro delle ambientazioni, intese come luoghi artificiali, nella letteratura italiana degli ultimi due secoli, anche in confronto con opere straniere tradotte in italiano. Ci si aspetta inoltre che la letteratura tradotta abbia svolto un ruolo di anticipazione rispetto alla narrativa nativa nell'introduzione di nuovi spazi narrativi, veicolando generi e ambientazioni propri di altre tradizioni letterarie prima che questi si affermassero nella produzione italiana originale.

L'approccio metodologico adottato è quello del *distant reading*⁶ (Moretti 2000; 2013), cioè una lettura “a distanza” che si serve di metodi computazionali, quantitativi e statistici per estrarre e analizzare dati da opere letterarie al fine di individuare tendenze e collegamenti e quindi di comprendere meglio la letteratura, ricostruirne la storia, ricollocare generi ecc. Come rileva lo stesso autore, l'osservazione a distanza “fa vedere meno dettagli, vero: ma fa capire meglio i rapporti, i pattern, le forme” (Moretti 2005, 3): il *distant reading* può infatti far emergere tendenze che sarebbero impossibili da notare attraverso il *close reading*, cioè la lettura ravvicinata e l'interpretazione attenta di testi letterari. Inoltre, permette di considerare opere solitamente escluse dagli studi “da vicino”, come quelle di consumo, particolarmente interessanti anche dal punto di vista linguistico (cfr. Alfieri 1994; Ricci 2013).

⁵ Si vedano, a titolo di esempio, i saggi raccolti da AA. VV. 2011; Anselmi e Ruozzi 2003; Ubbidente e Rybicki 2010; le considerazioni di Iacoli (2018) e l'*Atlante Calvino* <https://atlantecalvino.unige.ch/> (ultima data di consultazione dei link: 12 febbraio 2026).

⁶ Negli ultimi anni, il *distant reading* è stato oggetto del progetto europeo *Distant Reading for European Literary History* (COST Action CA16204) tra novembre 2017 e aprile 2022, volto a sviluppare metodi e a creare risorse per un'analisi della letteratura europea. Uno degli obiettivi del progetto è stato quello di assemblare un corpus di romanzi, pubblicati tra Ottocento e Novecento, in dieci lingue europee, tra cui anche l'italiano, denominato ELTeC (*European Literary Text Collection*). Lo stesso approccio metodologico è oggetto anche del PRIN RIND (*Reading the Italian Novel at a Distance*) (1830-1930), che vede la collaborazione delle università di Trieste, Padova e Tor Vergata e si concentra proprio sull'osservazione di aspetti come l'uso del discorso diretto e l'evoluzione delle ambientazioni in un secolo di narrativa in italiano (traduzioni comprese). Le specifiche del progetto si trovano sul sito <https://rind.units.it>.

A partire da queste premesse, la ricerca si articola attorno a due domande principali. In primo luogo, intende verificare in che modo la rappresentazione delle ambientazioni artificiali cambi nella narrativa italiana tra XIX e XXI secolo e se tali variazioni possano essere messe in relazione con trasformazioni storiche e sociali più ampie. In secondo luogo, il lavoro si propone di osservare se esistano differenze tra opere scritte originariamente in italiano e opere tradotte, ipotizzando che le due componenti possano riflettere modelli diversi di costruzione dello spazio narrativo. Includere le traduzioni serve a rendere possibile un confronto tra testi nativi e testi tradotti che condividono la stessa lingua di pubblicazione: le traduzioni sono infatti considerate come parte del sistema letterario in lingua italiana nel momento in cui entrano in circolazione. Seguendo l'impostazione del modello polisistemico (Even-Zohar 1990), la letteratura tradotta non è intesa come un insieme eterogeneo di testi importati, ma come un sistema integrato e attivo all'interno della letteratura di arrivo, che interagisce con le altre componenti del sistema e contribuisce alla definizione del repertorio disponibile in un dato momento storico.

2. Il corpus

La creazione del corpus rappresenta una fase cruciale dell'analisi (cfr. Ondelli 2018), in cui è necessario bilanciare criteri teorici e vincoli pratici, come la disponibilità dei testi in formato digitale e le questioni legate al diritto d'autore. In questo studio si è scelto di utilizzare il corpus già proposto da Ondelli (2019),⁷ di 8.352.265 token.⁸ Nonostante alcuni limiti, tra cui la dimensione relativamente contenuta della raccolta, la distribuzione non uniforme delle opere nei diversi periodi e la rappresentazione non sempre bilanciata di autori e generi, il corpus può essere usato come punto di partenza per questa indagine, soprattutto considerando le difficoltà della compilazione di una raccolta completa per un intervallo di tempo così ampio e la mancanza di alternative.

Questo studio riprende anche la suddivisione in quattro periodi storici già proposta da Ondelli (*ibid.*) in base alla data di prima pubblicazione dell'opera nativa o tradotta (in questo caso senza considerare l'anno di pubblicazione dell'opera originale). La scelta di datare le opere tradotte in base alla prima pubblicazione in italiano, e non a quella dell'originale, non è priva di conseguenze interpretative: un'opera di Dickens tradotta nel 1930 riflette un immaginario spaziale vittoriano, non quello del primo Novecento italiano. Tuttavia, la decisione è deliberata e risponde a una logica precisa: ciò che interessa in questa sede non è la cultura d'origine del testo, ma il momento in cui esso entra nella circolazione letteraria italiana e può esercitare un'influenza sul sistema ricevente. La data della traduzione coincide quindi con la data della ricezione, e il corpus tradotto rappresenta la letteratura straniera così come è stata mediata e proposta al lettore italiano in un dato momento storico.⁹

La periodizzazione ripresa da Ondelli è stata costruita dall'autore in base ad aspetti sociali, culturali, ma anche editoriali. Il primo periodo è costituito dal XIX secolo (1800-1899; circa 860.000 parole per le opere di italiani e altrettante per le traduzioni), d'ora in avanti 1800,

⁷ Si veda anche Sciumbata (2024) per un'analisi più approfondita dello stesso corpus.

⁸ I dettagli sulla composizione del corpus sono riportati in Appendice. Per le questioni di copyright già menzionate, il corpus nel suo insieme non può essere reso pubblicamente disponibile né redistribuito.

⁹ Va inoltre considerato che le traduzioni storiche sono spesso introvabili anche nelle biblioteche, che tendono a conservare le edizioni più recenti.

considerato come unitario per raggiungere un numero sufficiente di parole, vista la difficoltà di reperire testi pubblicati nel XIX secolo già digitalizzati o in stati di conservazione che non consentono la digitalizzazione. L'Ottocento costituisce un unico blocco anche in considerazione del predominio linguistico e culturale francese e della situazione politica, sociale e culturale italiana, ancora frammentata e poco alfabetizzata. Il Novecento è invece suddiviso in tre fasi: la prima (1900-1945; circa 1.200.000 parole sia per il subcorpus italiano sia per quello tradotto), che comprende le due guerre mondiali e, a partire dagli anni Venti, il consolidarsi del fascismo e delle sue politiche linguistiche nazionaliste. La fase successiva è quella del *boom* economico del dopoguerra e delle nuove questioni linguistiche (1946-1975; circa 915.000 parole per subcorpus); un periodo conclusivo (1976-2005; circa 1.200.000 parole per subcorpus), segnato dall'oralizzazione dell'italiano e dall'influsso sempre più consistente della lingua inglese (cfr. Antonelli 2011).

La composizione dei sottoperiodi del corpus è rappresentata in Figura 1. Come si può notare, le dimensioni dei subcorpora originali e tradotti dei singoli periodi sono comparabili, ma periodi diversi hanno dimensioni differenti: ciò rende meno immediato il confronto dei risultati, che sono infatti stati normalizzati calcolando la frequenza relativa delle occorrenze rispetto al numero totale di parole di ciascun sottocorpus, così da rendere confrontabili sottoperiodi con dimensioni diverse. Va tenuto presente che la normalizzazione per numero di token non elimina completamente le differenze strutturali tra i subcorpora, che dipendono anche dalla composizione tematica e generica dei testi: i valori normalizzati vanno pertanto interpretati come indicatori di tendenza e non come misure assolute.

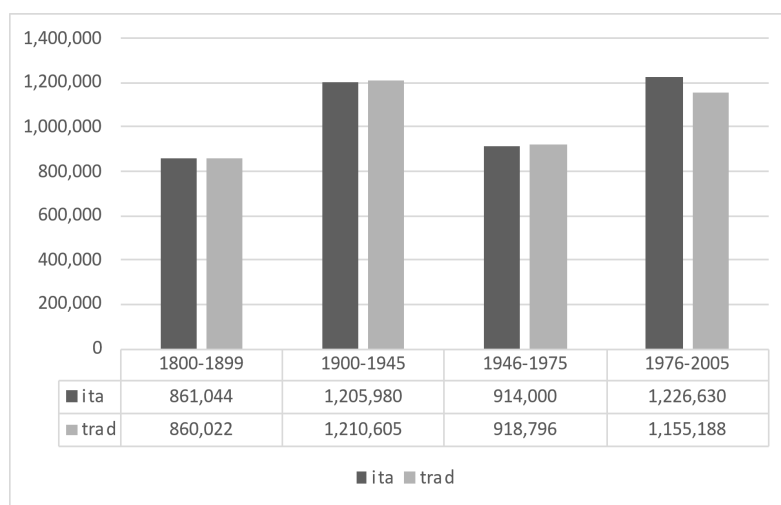


Figura 1: distribuzione di token per sottoperiodo del corpus

Il grafico in Figura 2 riassume invece il peso delle lingue di partenza dei testi tradotti: si nota la predominanza del francese nell'Ottocento e quella dell'inglese in tutti i periodi successivi, con il sottoperiodo postbellico totalmente costituito da testi in questa lingua.

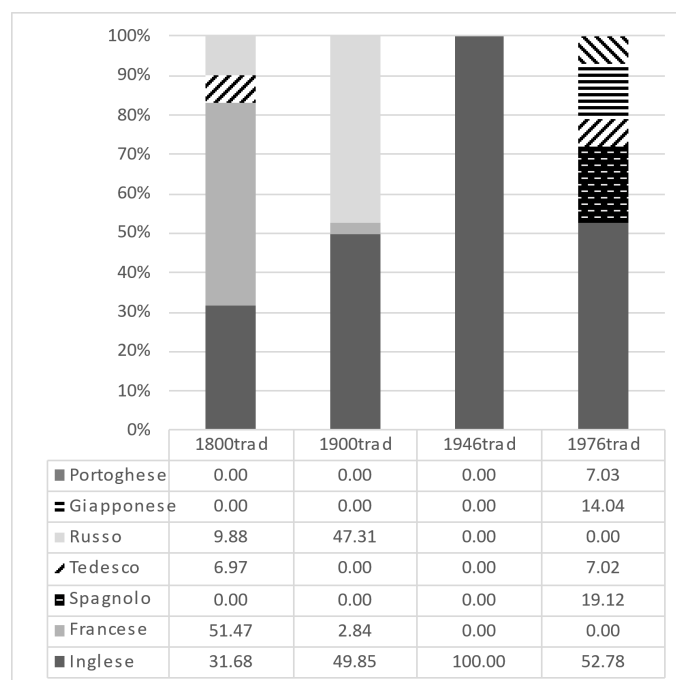


Figura 2: percentuale delle lingue che compongono i testi tradotti per sottoperiodo

Un'altra premessa importante riguarda la rappresentazione degli autori: alcuni di loro, come Pirandello e Salgari, sono presenti con più opere, un aspetto che può far prevalere risultati influenzati dai temi ricorrenti nelle loro produzioni. Per distinguere l'influenza dei singoli autori da quella delle caratteristiche di genere, un'analisi sistematica della distribuzione delle ambientazioni richiederebbe una classificazione esplicita dei testi per genere e sottogenere. Tuttavia, tale classificazione risulta spesso problematica, soprattutto per la narrativa di lungo periodo, in cui i confini tra generi possono essere sfumati o oggetto di interpretazioni diverse, poiché i generi letterari non costituiscono categorie fisse ma "famiglie" soggette a trasformazioni storiche e a continui processi di ridefinizione (cfr. Fowler 1982). Per questo motivo il presente studio non prende in considerazione la dimensione del genere letterario e si concentra invece sull'analisi diacronica delle categorie lessicali delle ambientazioni.

3. La lista di luoghi artificiali

Un punto critico per estrarre luoghi artificiali da una raccolta di testi è rappresentato dalla creazione della lista lessicale da cui partire per estrapolare i dati dal corpus: non sono infatti disponibili liste elettroniche e sistematiche di ambientazioni. Negli ultimi anni, l'informatica e la linguistica computazionale hanno proposto alcuni metodi per individuare automaticamente in grandi moli di testi categorie come persone, luoghi, oggetti, associazioni ed enti, numeri ed espressioni temporali, con il *named-entity recognition* (NER), cioè il riconoscimento e la classificazione di entità nominate attraverso l'applicazione di regole o di metodi statistici. I sistemi NER disponibili per l'italiano, come quello incluso nel software *Tint* (Aprosio e Moretti

2018) oppure *NerIta*,¹⁰ riconoscono i toponimi, come *Italia* o *Trieste*,¹¹ ma non prevedono la classificazione di nomi generici come edifici, costruzioni, stabilimenti e altri luoghi artificiali come *scuola*, *fabbrica*, *chiesa* ecc.¹²

Per l'individuazione delle ambientazioni si è quindi scelto un approccio *lexicon-based*. Infatti, sebbene metodi distribuzionali come i *word embeddings* permettano di individuare relazioni semantiche tra parole sulla base delle loro co-occorrenze nei corpora, la loro applicazione a corpora letterari di dimensioni relativamente limitate può produrre risultati poco stabili. Come osservano Sciandra, Trevisani e Tuzzi (2021), i testi narrativi risultano spesso più “dispersivi” dal punto di vista tematico rispetto ad altri generi testuali, e rendono perciò meno robuste le associazioni tra parole su cui si basa questo tipo di modelli; inoltre, l'addestramento efficace degli *embeddings* richiede generalmente corpora molto più estesi di quelli disponibili per studi diacronici sulla narrativa. Un ulteriore limite, almeno rispetto agli obiettivi di questo studio, riguarda il tipo di similarità individuata dai modelli distribuzionali. Un modello di *word embeddings* addestrato su un corpus narrativo potrebbe avvicinare un lemma come *chiesa* a parole semanticamente o narrativamente associate, come *prete*, *preghiera* o *mesa*, ma non necessariamente ad altri luoghi artificiali, come *castello*, *stazione* o *fabbrica*, che appartengono alla stessa categoria funzionale senza condividere contesti distribuzionali ricorrenti.

Tra le risorse esistenti per l'annotazione semantica, sono stati presi in considerazione *USAS*¹³ e *ItalWordNet*¹⁴ (Roventini *et al.* 2016), ma nessuna delle due soluzioni è stata adottata. La prima include nella categoria dei luoghi sia nomi comuni sia nomi propri e presenta una classificazione ampia, non sempre adeguata a distinguere in modo sistematico i sottotipi presi in considerazione per l'analisi. *ItalWordNet*, pur offrendo una struttura semantica più formalizzata, non prevede una classe unitaria di luoghi: le entità risultano distribuite in diversi rami della gerarchia e spesso associate a più sensi: ciò renderebbe necessarie operazioni di selezione e disambiguazione poco immediate e non banali.

Approcci più recenti basati su modelli transformer contestualizzati, come *BERT*, consentono in linea di principio una classificazione più fine delle entità spaziali, ma richiedono annotazioni manuali specifiche per il dominio letterario e la lingua target. Kababgi, Grisot, Pennino e Herrmann (2024), che hanno sviluppato un classificatore di luoghi per la narrativa svizzera in

¹⁰ https://huggingface.co/bullmount/it_nerIta_trf

¹¹ Un elemento evidentemente usato per distinguere le entità nominate è la presenza di maiuscole non dovute al contesto sintattico (quindi non a inizio frase o dopo un segno di punteggiatura forte). Entrambi i sistemi NER provati identificano come luogo anche un nome inventato nella frase “Sono andata ad *Astranapu”, che viene etichettato come luogo anche grazie ad altri elementi nel contesto, come preposizioni, verbi di movimento ecc.

¹² Negli ultimi anni sono stati sviluppati diversi progetti dedicati al riconoscimento automatico di entità nei testi letterari, spesso con l'obiettivo di costruire corpora annotati o analizzare la rappresentazione di personaggi e luoghi nella narrativa. Tra i contributi più rilevanti si possono citare il dataset di entità letterarie proposto da Bamman, Popat e Shen (2019), il progetto PPORTAL_ner per la letteratura portoghese (Silva e Moro 2024) e il sistema di regole per l'estrazione di personaggi e luoghi nei romanzi francesi sviluppato da Bornet e Kaplan (2017). A questi si affiancano studi basati su annotazioni manuali o semi-automatiche di entità narrative, come il progetto *The Emotions of London* (Heuser, Moretti e Steiner 2016).

¹³ <https://github.com/UCREL/Multilingual-USAS>

¹⁴ <https://dSPACE-clarin-it.ilc.cnr.it/items/3bcdffb9-ea4f-45f3-a321-6c32aa2f9f06>

tedesco, riconoscono esplicitamente che si tratta di "a work in progress", addestrato su soli sei romanzi e non trasferibile ad altre lingue senza nuove risorse annotate.

Per questa analisi si è quindi preferito costruire una lista lessicale specifica per la categoria dei luoghi artificiali, combinando diverse fonti e integrando procedure automatiche e selezione manuale. La scelta di un approccio basato su liste lessicali risponde anche a esigenze di trasparenza e controllabilità del metodo: a differenza dei modelli supervisionati o distribuzionali, che richiedono grandi quantità di dati annotati o corpora molto estesi, l'approccio adottato consente di mantenere un controllo diretto sulle categorie analizzate e sui risultati.

Per costruire la lista di lemmi appartenenti a categorie semantiche relative a luoghi, in un primo tentativo si è cercato di progettare un'estrazione automatica di lessemi collegati a edifici e luoghi artificiali da dizionari ed enciclopedie online tramite una prima consultazione manuale. Queste fonti hanno senza dubbio il vantaggio di contenere numerosi lemmi con le relative definizioni, ma si sono rivelate poco utili per la loro scarsa sistematicità e perché contengono dati non strutturati. Per esempio, nel dizionario De Mauro¹⁵ si trovano le seguenti definizioni:

- villa: "casa signorile, modernamente di dimensioni non eccessive, circondata da un giardino o da un parco";
- bar: "locale in cui si consumano alcolici, caffè, bibite, panini e sim.";
- giardino: "terreno, spec. recintato o delimitato da siepi, muri e sim., coltivato con piante ornamentali e fiori";
- caserma: "edificio destinato all'alloggiamento e all'addestramento di corpi militari o di analoghe organizzazioni";
- fabbrica: "stabilimento in cui si svolge una produzione industriale".

È evidente che le definizioni non sono sufficientemente regolari per poter raccogliere in automatico i lemmi a cui fanno riferimento: se si dovessero scrivere delle regole per avviare procedure di *scraping*, si andrebbe incontro a un'operazione estremamente dispendiosa, e lo stesso vale per lo spoglio manuale delle voci. Un dizionario che è invece risultato utile per costituire la lista è quello analogico (Feroldi e Dal Pra 2011). A partire da alcune parole "seme", come *casa*, *chiesa*, *scuola*, *fabbrica*, *edificio*, è stato stilato un primo nucleo di parole da estrarre. Per ampliarlo, un'altra fonte è stata *Wikidata*,¹⁶ cioè la banca dati alla base dei progetti collegati a *Wikimedia*, di cui fa parte anche l'enciclopedia libera e collaborativa *Wikipedia*. *Wikidata* presenta diversi vantaggi: oltre ad avere una licenza di pubblico dominio che la rende liberamente utilizzabile, contiene dati strutturati, cioè organizzati per campi predefiniti, che facilitano l'estrazione delle informazioni. Infine, *Wikidata* è un database multilingue e ogni voce è collegata ai suoi corrispettivi in tutte le traduzioni disponibili, una caratteristica particolarmente utile perché rende il metodo di creazione delle liste più flessibile e replicabile anche in altre lingue. La banca dati di *Wikidata* contiene diverse sezioni che rimandano a luoghi artificiali che sono etichettati come "factory", "facility", "workplace", "shop", "commercial building", "religious building", "house", "rural building", "fortification" e "venue". Le informazioni di *Wikidata* sono state estratte in automatico e inserite in una tabella: la somma di tutte le fonti, escludendo le voci doppie, contiene un totale di 1.241 entrate (da *abbazia* a *zuccherificio*), tra lemmi singoli, locuzioni

¹⁵ www.dizionario.internazionale.it

¹⁶ Un approccio simile è già stato sperimentato con gli agentivi (Sciumbata, Nadalutti e Tringali 2021) e gradi militari, gerarchie religiose e titoli nobiliari (Sciumbata 2023).

complesse e polirematiche.¹⁷ Pur non essendo esaustiva, la lista copre un’ampia gamma di tipologie di luoghi artificiali, sia urbani sia rurali, e risulta sufficientemente estesa per un’analisi esplorativa, perché permette di intercettare un insieme stabile e ricorrente di categorie spaziali nel corpus.

Per ampliare ulteriormente la lista, è stato fatto un ultimo tentativo con *ChatGPT*, a cui sono state chieste delle liste con *prompt* diversi, come “dammi un elenco di edifici e luoghi artificiali” o “continua questo elenco di edifici”. I risultati non sono però stati soddisfacenti: nel primo caso vengono forniti esempi di luoghi specifici come “Casa Bianca” o “Torre Eiffel”, mentre la seconda richiesta non restituisce particolari elementi di novità rispetto alla lista già ottenuta.¹⁸

In ultimo, i lemmi della lista sono stati suddivisi in sette macrocategorie: religioso (ad es. *chiesa, abbazia, badia*); militare (ad es. *caserma, questura, bunker*); pubblico (ad es. *stazione, aeroporto, clinica*); civile (ad es. *villa, castello, giardino*); rurale (ad es. *casolare, cascina, granaio*); industriale (ad es. *fabbrica, acciaieria, mobilificio*) e commerciale (ad es. *negozio, albergo, hotel*). Tale etichettatura, svolta manualmente, comprende categorie arbitrarie e in parte sovrapposte: per esempio, una stazione di polizia potrebbe essere considerata un edificio civile, militare e pubblico; una chiesa è un edificio religioso e pubblico; un negozio è commerciale, civile e pubblico. La scelta di queste sette categorie è motivata dalla loro rilevanza rispetto ai principali assi di trasformazione storico-sociale del periodo analizzato (urbanizzazione, industrializzazione, secolarizzazione, militarizzazione) e dalla loro capacità di rappresentare distinzioni significative nella rappresentazione dello spazio narrativo. La categorizzazione, che assegna una sola etichetta a ciascun lemma, non ha l’obiettivo di definire classi semantiche rigide, ma di offrire una griglia interpretativa utile per osservare tendenze generali nella distribuzione delle ambientazioni nel corpus.

Un aspetto che merita una precisazione riguarda il trattamento dei luoghi fittizi, simbolici o ibridi, presenti nella narrativa in misura variabile a seconda del genere e del periodo. L’approccio lessicale adottato non distingue tra luoghi reali e luoghi inventati: parole come *castello* o *abbazia* vengono estratte e conteggiate indipendentemente dal fatto che designino un luogo geograficamente identificabile, un luogo immaginario o un luogo la cui esistenza è deliberatamente ambigua nel testo. La distinzione tra luoghi veri e immaginati richiederebbe tuttavia un oneroso lavoro manuale di etichettatura qualitativa, che dovrebbe tenere in considerazione il contesto narrativo. Inoltre, ai fini del presente studio, ciò che interessa non è la referenzialità geografica del luogo, ma la sua presenza come elemento lessicale e come categoria dello spazio narrativo. Anche i luoghi fittizi contribuiscono a definire l’immaginario spaziale di un’opera, di un genere o di un’epoca, e la loro inclusione nel conteggio non altera la comparabilità dei dati tra i diversi sottoperiodi. Restano invece esclusi dall’analisi i toponimi veri e propri (veri o inventati), come nomi di città, regioni o Paesi.

¹⁷ La lista completa è disponibile su <https://github.com/PRIN-Rind/LinguaPipe>

¹⁸ In prospettiva, l’ampliamento delle liste lessicali potrebbe essere esplorato anche attraverso sistemi di *retrieval-augmented generation* (RAG), che consentono di interrogare modelli linguistici a partire da repertori documentari o lessicali specifici. Strumenti di questo tipo (ad esempio *NotebookLM*) potrebbero essere utilizzati come supporto per individuare nuovi candidati lessicali da aggiungere alle liste.

4. Estrazione dei dati

Dopo aver assemblato la lista, gli elementi non sono stati estratti direttamente dal corpus al fine di ridurre il rischio di ottenere falsi positivi: un esempio è la parola *forte*, che nell'estrazione di luoghi è da considerare solo se è un sostantivo, mentre tutte le sue occorrenze come aggettivo sono da escludere. Un altro problema è rappresentato dal fatto che la lista contiene lemmi solo alla forma base: un'estrazione "diretta" escluderebbe quindi le forme plurali o alterate. Per aggirare entrambi i problemi, il corpus è stato trattato con *Bran* (Sciumbata e Tringali s.d.), un software modulare che permette di gestire facilmente l'estrazione di dati testuali e la loro analisi.¹⁹ *Bran* si serve di *UdPipe 2* (Straka 2018), programma che lemmatizza²⁰ e aggiunge informazioni sulla parte del discorso, le caratteristiche morfologiche e le relazioni di dipendenza sintattica tra i token. In questo modo, è stato possibile individuare i luoghi solo tra le parole ridotte alla forma base e classificate come sostantivi, così da ridurre la necessità di interventi manuali. Questa scelta comporta una limitazione consapevole: i luoghi narrativi sono talvolta evocati anche attraverso aggettivi, verbi o sintagmi nominali complessi che non compaiono nella lista. L'approccio adottato coglie pertanto la presenza esplicita dei luoghi, non la loro rappresentazione complessiva, che richiederebbe un'analisi del contesto più ampia e più onerosa dal punto di vista metodologico.

Il primo tentativo di estrazione ha tuttavia mostrato che alcuni lemmi potevano essere ambigui: parole come *caffè*, *ufficio*, *pensione* e *studio* hanno infatti più significati e non fanno sempre riferimento a luoghi.²¹ Ciò è stato risolto con una lettura corsiva dei contesti in cui comparivano parole polisemiche per eliminare a mano le occorrenze non pertinenti. Un'altra questione è legata alla parola *casa*, il luogo menzionato più di frequente in tutto il corpus: tutte le occorrenze del lemma sono però state eliminate perché non danno informazioni sulla natura dell'edificio (per esempio, se si tratta di un appartamento in città o di una cascina in campagna).

Sebbene non sia stata condotta una valutazione formale della qualità dell'estrazione attraverso un *gold standard* annotato, la fase di disambiguazione manuale descritta sopra ha permesso di individuare le principali fonti di errore sistematico: i falsi positivi più ricorrenti riguardano lemmi polisemici in contesti non spaziali, come *studio* in senso intellettuale o *pensione* in senso economico, oltre ad alcune occorrenze con valore metaforico o traslato. L'assenza di un corpus annotato di riferimento per le ambientazioni nella narrativa italiana rende inoltre difficile stimare

¹⁹ La scelta di *Bran* come interfaccia, piuttosto che di pipeline analoghe come *Stanza*, è motivata dalle funzionalità integrate di gestione del corpus che il software offre: oltre alla lemmatizzazione e al POS-tagging, consente di estrarre e filtrare le occorrenze per categoria grammaticale, di aggregare i risultati per sottocorpus e di esportarli in formato tabulare, riducendo la necessità di passaggi intermedi e di *scripting* aggiuntivo. Questo ha semplificato il lavoro in modo significativo, considerando che il corpus è suddiviso in otto subcorpora distinti.

²⁰ A proposito di lemmatizzazione, si tratta di un processo particolarmente problematico per i testi ottocenteschi, nei quali si riscontrano una maggiore variabilità grafica e morfologica, oltre alla presenza di forme oggi desuete o non standardizzate. I modelli attualmente disponibili per l'italiano sono inoltre generalmente addestrati su testi contemporanei e non letterari (Twitter, Wikipedia, quotidiani online ecc.), quindi possono mostrare prestazioni meno affidabili quando applicati a corpora diacronici e a testi di natura letteraria.

²¹ Un caso in parte diverso è rappresentato da lemmi come *chiesa*, *scuola* o *teatro*, che possono indicare sia un luogo fisico sia l'istituzione o l'attività che vi si svolge. Poiché questa ambiguità non è sempre risolvibile automaticamente, tali forme sono state mantenute quando compatibili con un'interpretazione spaziale.

in modo affidabile il *recall*. Le tendenze descritte nei paragrafi seguenti vanno quindi interpretate come indicative e coerenti con un’analisi di tipo esplorativo.

Una volta ripuliti, i risultati sono stati normalizzati e presentati come percentuali rispetto al totale delle occorrenze di ciascun subcorpus, così da rendere confrontabili sottoperiodi con dimensioni diverse (cfr. paragrafo 2).

5. Risultati

L’estrazione ha permesso di ricavare dal corpus 442 luoghi artificiali della lista, per un totale di 48.205 occorrenze. Un primo quadro sull’andamento delle ambientazioni e sulle differenze tra opere native e tradotte emerge dal confronto tra le categorie di luoghi artificiali nel subcorpus italiano e in quello tradotto, rappresentate rispettivamente nella Figura 3 e nella Figura 4.

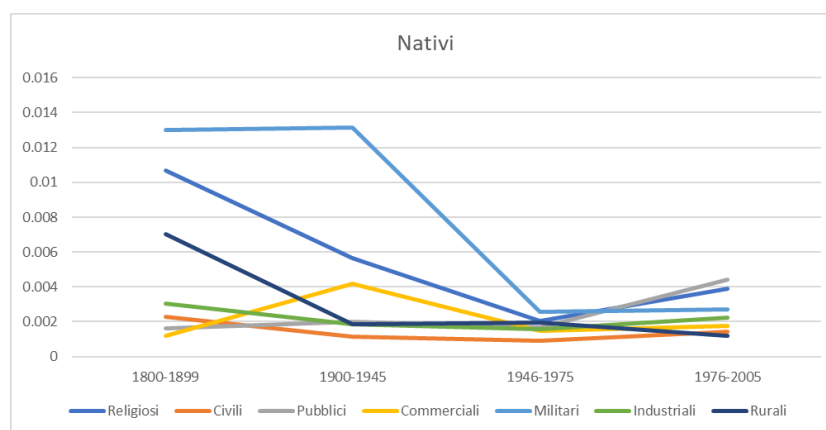


Figura 3: andamento delle categorie di luoghi artificiali nel subcorpus nativo

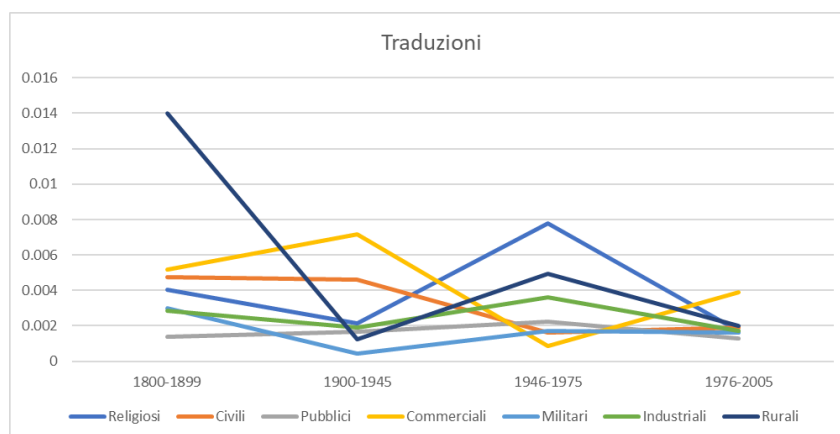


Figura 4: andamento delle categorie di luoghi artificiali nel subcorpus tradotto

Nelle opere di scrittori italiani si registra una prevalenza di luoghi religiosi nel 1800 e nella prima parte del Novecento, seguita da un brusco calo. Anche i luoghi militari compaiono con una certa frequenza nel primo sottoperiodo e restano invariati nel periodo successivo, prima di diminuire (cfr. Figura 3). Nel 1900-1945 ci si sarebbe potuti attendere un aumento e la percentuale invariata potrebbe sorprendere, visto il succedersi delle due guerre mondiali negli stessi anni; tuttavia, il dato si può contestualizzare se si ricostruiscono gli argomenti dei titoli che compongono questa sezione del corpus (v. Appendice), ma anche la loro data di pubblicazione, poiché la metà dei romanzi è stata pubblicata prima del 1915, quando il conflitto non era ancora iniziato.

Le ambientazioni rurali tendono a diminuire dal 1900, come si nota sempre in *Figura 3*: questo calo può essere collegato ai processi di urbanizzazione e industrializzazione²² e al progressivo spostamento dell'attenzione narrativa verso gli spazi urbani. Tuttavia, la rappresentazione letteraria della campagna non segue necessariamente in modo immediato le trasformazioni sociali. Come osserva Williams nella sua “scala mobile” (*escalator*) della nostalgia (1973, 9-12), la letteratura tende a collocare la “vera” vita rurale in un passato sempre appena trascorso, mentre il presente viene percepito come il momento della sua definitiva scomparsa. La dimensione nostalgica della rappresentazione della campagna può prolungarne la presenza narrativa anche quando la sua centralità sociale è già in declino, e ciò può spiegare la stabilità nel terzo periodo del corpus. La diminuzione e la successiva stabilizzazione delle ambientazioni rurali possono quindi essere interpretate come l'esito di un processo graduale, in cui la campagna passa progressivamente da spazio della comunità vissuta a oggetto di memoria, mentre la narrativa si orienta sempre più verso le nuove realtà urbane e industriali.

I luoghi commerciali sono praticamente assenti nell'Ottocento, aumentano nel primo Novecento e diminuiscono nuovamente negli anni successivi, un andamento che può essere messo in relazione con i cicli dell'economia mondiale descritti da O'Rourke e Williamson (1999, 33-43): l'incremento iniziale coincide con l'apice della cosiddetta *prima globalizzazione* (1870-1914), periodo in cui l'integrazione dei mercati e la riduzione dei costi di trasporto trasformano le città in nodi centrali degli scambi internazionali. Al contrario, la contrazione rilevata negli anni tra le due guerre può essere interpretata alla luce dei processi di *deglobalizzazione* e del ritorno a politiche protezionistiche e autarchiche, che riducono la centralità dello spazio commerciale e del mercato globale sia nella realtà economica sia nell'immaginario sociale dell'epoca (O'Rourke e Williamson 1999, 285-287). L'incremento dei riferimenti a luoghi legati al commercio e allo scambio può essere messo in relazione anche con la crescente centralità economica degli spazi urbani nella società moderna. Come osserva Williams (1973, 279-281), con lo sviluppo del capitalismo la città tende a configurarsi sempre più come snodo di un sistema economico che integra produzione, scambio e distribuzione su scala sempre più ampia. In questa prospettiva, la presenza più frequente di negozi, mercati e altri spazi commerciali nella narrativa può riflettere

²² Il progressivo spostamento delle ambientazioni dalla campagna alla città può essere letto anche alla luce delle trasformazioni economiche e sociali che hanno interessato l'Italia tra XIX e XX secolo. Come ricorda De Mauro, alla metà dell'Ottocento l'Italia era ancora un paese prevalentemente agricolo: gli addetti all'agricoltura costituivano il 59,6% della popolazione attiva e restavano ancora il 55,5% nel 1911; nel Novecento la loro percentuale scende al 42,6% nel 1951, al 33% nel 1961 e al 19,4% nel 1970, mentre cresce il peso delle attività industriali e terziarie (De Mauro 1991, 64). L'accelerazione del processo nel secondo dopoguerra è confermata dai dati sull'Italia repubblicana: nel 1951 la popolazione attiva era impegnata per il 42,2% in agricoltura, per il 32,1% nell'industria e per il 25,7% nei servizi; nel 1971 gli addetti all'agricoltura scendono al 17,2%, mentre l'industria sale al 44,3% e i servizi al 38,4% (De Mauro 2014, 60-61). La diminuzione delle ambientazioni rurali nella narrativa può essere quindi interpretata come il riflesso culturale di una più ampia transizione della società italiana da un'economia agricola a una industriale e urbana.

la progressiva importanza della città come luogo di organizzazione delle attività economiche e delle relazioni sociali.

I luoghi civili e quelli industriali hanno invece andamenti simili: entrambi subiscono una flessione e un aumento successivo, ma non mostrano grandi cambiamenti; al contrario, le ambientazioni pubbliche sono le uniche che tendono a crescere, anche se leggermente, almeno fino al 1975. Questo dato mostra una relativa stabilità delle rappresentazioni legate alla produzione e alla vita sociale organizzata, che non subiscono trasformazioni altrettanto marcate rispetto ad altri ambiti.

Le categorie delle opere tradotte hanno andamenti diversi e meno chiari, con più oscillazioni rispetto alle opere italiane. L'ambientazione più presente nell'Ottocento è quella rurale, che cala bruscamente nel periodo seguente e poi mostra un andamento poco lineare. Nel 1946-75 si osserva un incremento delle ambientazioni religiose, inizialmente meno rappresentate rispetto alle opere italiane. Tale andamento sembra dipendere principalmente dai temi di alcune opere del periodo: la parola *chiesa*, ad esempio, registra complessivamente 169 occorrenze, di cui 53 nel *Villaggio di vetro*, un romanzo incentrato sulla vita di una piccola comunità in cui la chiesa svolge un ruolo centrale come luogo di incontro e di organizzazione della vita collettiva.

Gli edifici militari sono invece molto meno presenti rispetto alla narrativa italiana e la loro frequenza è altalenante nei quattro sottoperiodi. I luoghi civili e quelli commerciali sono più presenti nel XIX secolo rispetto alla raccolta italiana, mentre quelli industriali mostrano percentuali simili e aumentano leggermente nella seconda metà del periodo preso in analisi. I luoghi pubblici non presentano invece un andamento chiaro: aumentano e poi diminuiscono nel corso del tempo.

A proposito della frequenza dei lemmi, per brevità saranno commentati solo gli aspetti più significativi nelle tendenze dei dieci luoghi con più occorrenze assolute in tutto il corpus: *albergo*, *appartamento*, *chiesa*, *cortile*, *giardino*, *negozio*, *piazza*, *scuola*, *ufficio*, *villa*. Da queste parole si possono già desumere alcuni temi evidentemente ricorrenti: la vita domestica in città (*appartamento* e *cortile*); il viaggio (*albergo*); il riferimento a un luogo sacro, ma anche di aggregazione (*chiesa*), e di nuovo l'aggregazione nella piazza; il commercio (*negozio* e *ufficio*) e così via. Nella Figura 5 sono rappresentati i dieci lemmi nei testi nativi, nella Figura 6 nelle traduzioni; la Figura 7 permette un confronto più immediato delle differenze tra sottoperiodi e testi italiani e tradotti.

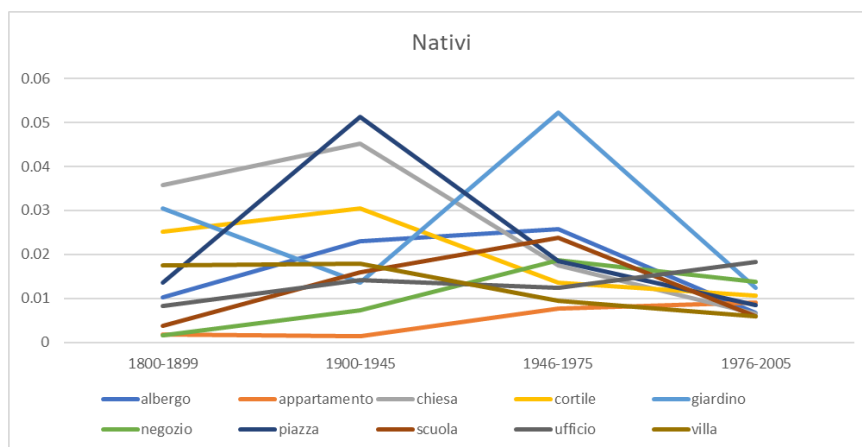


Figura 5: andamento dei dieci luoghi artificiali con più occorrenze nel subcorpus nativo

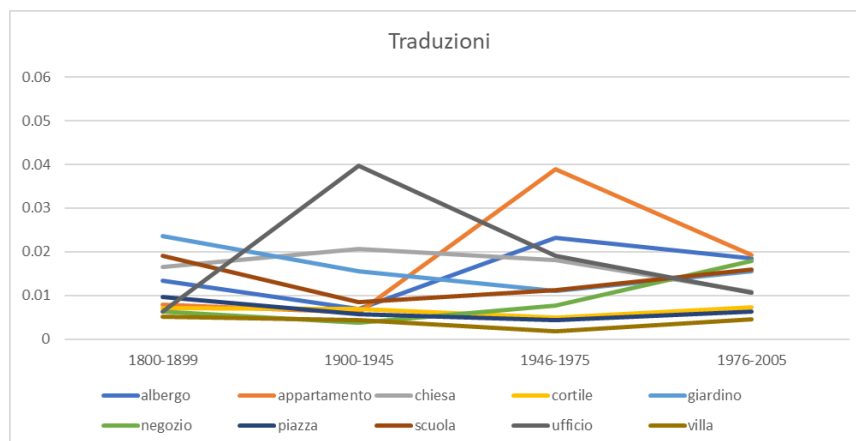


Figura 6: andamento dei dieci luoghi artificiali con più occorrenze nel subcorpus tradotto

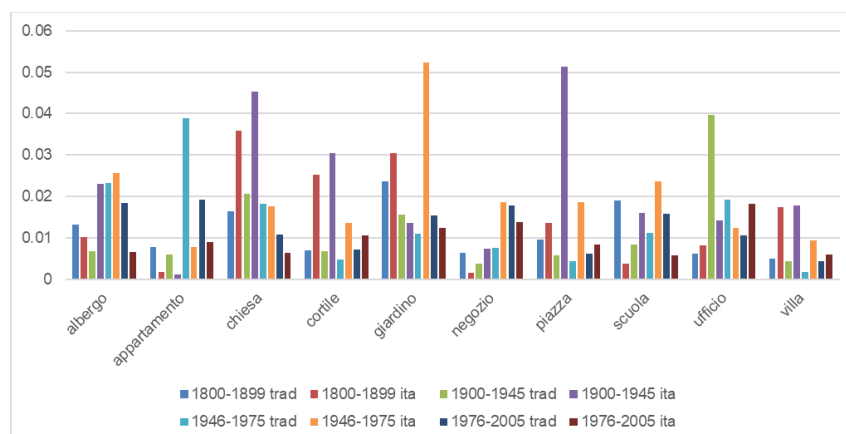


Figura 7: rappresentazione delle dieci parole più frequenti nel corpus nei sottoperiodi nativi e in quelli tradotti

Gli andamenti delle parole sono spesso altalenanti ed è difficile tracciare un profilo preciso; alcune tendenze risultano tuttavia riconoscibili e si registrano differenze tra opere di autori nativi e traduzioni. Nel primo caso, si osservano chiaramente nel periodo 1900-1945 i picchi di due luoghi pubblici e di incontro: *chiesa* e *piazza*, quest'ultima epicentro della vita cittadina (cfr. Collina 2003) e spazio fortemente radicato nell'immaginario italiano. Entrambe le ambientazioni tendono però a diminuire nei periodi successivi: in particolare, l'andamento di *chiesa* può essere interpretato come un indizio della graduale secolarizzazione della società italiana ed è coerente con la cronologia del fenomeno: il calo più marcato si registra nei sottoperiodi 1946-1975 e 1976-2005, che corrispondono alla fase in cui la secolarizzazione accelera il passo (Garelli 2007, cap. 2).

La progressiva marginalizzazione delle ambientazioni religiose non è un fenomeno isolato, ma si inserisce anche nel quadro dei processi di industrializzazione e del correlato spopolamento delle aree interne e montane del Paese (cfr. De Mauro 1991, 60). Le migrazioni interne del

secondo dopoguerra non implicano soltanto uno spostamento geografico della popolazione, ma contribuiscono alla dissoluzione delle tradizionali comunità rurali e dei loro sistemi di relazioni sociali (cfr. Ginsborg 2006, cap. VIII). Questo mutamento trova un possibile riflesso nelle tendenze letterarie del secondo dopoguerra, che registrano il passaggio da contesti sociali più circoscritti e riconoscibili a forme di organizzazione più complesse e frammentate. In questo senso, si può richiamare la nozione di *knowable community* proposta da Williams (1973, 165), che descrive comunità in cui le relazioni sociali risultano pienamente leggibili e interconnesse. La riduzione delle ambientazioni religiose può quindi essere letta anche in relazione alla perdita di centralità di questi spazi comunitari, progressivamente sostituiti da ambienti urbani e funzionali.

Nel 1946-1975 si assiste invece a un'ascesa repentina di *giardino*,²³ la “riproduzione artificiale di un paesaggio” (Basile 2003), quindi una sorta di spazio intermedio tra natura e città. *Scuola*, *albergo* e *negozio* crescono e subiscono poi una diminuzione nell'ultimo periodo; *ufficio* e *appartamento* sono invece gli unici due luoghi che vedono un aumento piuttosto costante, e possono essere letti come un'affermazione dell'ambientazione urbana e più borghese. All'aumento di *appartamento* si contrappone invece l'andamento speculare di *villa*, abitazione signorile e spesso collocata in contesti più rurali.

Una prima differenza tra opere italiane e tradotte si può evincere dalla minore presenza in queste ultime dei dieci luoghi presi in analisi, che può trovare spiegazione nel fatto che le ambientazioni delle opere tradotte sono diverse o diversamente distribuite, anche a causa della maggiore eterogeneità di culture e di temi da cui derivano le opere tradotte. I picchi più evidenti sono diversi rispetto a quelli italiani: riguardano *ufficio* (nel 1900-45) e *appartamento* (nel 1946-75). Nello stesso periodo, anche *albergo* ricorre con una certa frequenza; *chiesa* ha invece un andamento relativamente costante che non mostra lo stesso calo delle opere italiane, e *scuola*, *negozio* e *giardino* riprendono terreno dopo cali di diversa durata. Infine, nei testi tradotti, il cortile, la piazza e la villa sono luoghi marginali che non mostrano particolari variazioni.

È opportuno precisare che alcune delle variazioni rilevate tra sottoperiodi sono di entità estremamente ridotta e che non è stata condotta una verifica formale della loro significatività statistica. Le differenze più ampie e sistematiche, come il calo delle ambientazioni religiose nelle opere native o la progressiva affermazione di *ufficio* e *appartamento* in entrambi i subcorpora, appaiono comunque robuste rispetto alla dimensione dei sottocorpora e coerenti con le tendenze storico-sociali descritte. Le oscillazioni minori potrebbero invece riflettere la composizione specifica del corpus, in particolare la presenza di autori o opere con tematiche ricorrenti, più che tendenze strutturali della narrativa del periodo.

In generale, si può rilevare che gli autori italiani tendono a menzionare più di frequente luoghi pubblici e di incontro nelle loro opere, mentre quelle tradotte sono caratterizzate da ambientazioni più private. Entrambe le raccolte condividono comunque l'ambientazione prevalentemente urbana, un dato che contrasta con quanto osservato da Grisot e Herrmann (2023) per la narrativa svizzera in lingua tedesca, in cui prevalgono nettamente gli spazi rurali e naturali.

Un'altra osservazione importante riguarda la minore presenza dei dieci luoghi analizzati nell'ultimo sottoperiodo, sia nella narrativa italiana che in quella tradotta, e la loro sovrapposizione nei due subcorpora: si tratta di un aspetto che richiede ulteriori approfondimenti, ma un'ipotesi potrebbe ricollegarsi all'avvento di nuovi luoghi. Un'altra

²³ *Giardino* compare in tutti i dodici romanzi che compongono il periodo, con un totale di 239 occorrenze (dalle 5 di *Tempo di uccidere* alle 45 del *Conformista*).

spiegazione può riguardare una maggiore varietà negli argomenti dei testi più recenti, che non fa prevalere nessuna ambientazione in particolare.

Un'analisi dell'insieme dei lemmi estratti, al di là dei dieci più frequenti discussi nei paragrafi precedenti, permette di osservare un *pattern* ricorrente che riguarda il rapporto tra traduzioni e nativi nel tempo. Diversi lemmi mostrano frequenze più elevate nelle traduzioni in un dato sottoperiodo, per poi affermarsi nei testi nativi nel periodo successivo, il che suggerisce un possibile ruolo di anticipazione della letteratura tradotta rispetto a quella nativa nell'introduzione di nuovi spazi narrativi. È il caso di *ufficio*, già presente con alta frequenza nelle traduzioni ottocentesche e in crescita nei nativi nel Novecento; di *negozi*, con picco nelle traduzioni del 1900-1945 seguito da crescita nei nativi del dopoguerra; di *cinema* e *garage*, che compaiono nelle traduzioni del primo Novecento e si affermano nei nativi solo nel secondo. Il caso più estremo è *supermercato*, che appare esclusivamente nelle traduzioni del periodo 1946-1975, senza occorrenze nei testi nativi dello stesso periodo. Non tutti i luoghi seguono questo schema: *fabbrica* e *ospedale* mostrano frequenze più alte nei nativi, un dato che potrebbe dipendere sia dalla specificità della realtà sociale italiana (con la fabbrica al centro del romanzo operaista e neorealista) sia dalla composizione generica del corpus tradotto, in cui generi come il giallo e la fantascienza angloamericana, ben rappresentati nel dopoguerra, non includono tipicamente queste ambientazioni.

6. Conclusioni e prospettive di ricerca

I dati evidenziano uno spostamento progressivo degli spazi rappresentati: dalla centralità di luoghi collettivi e comunitari, come la chiesa e la piazza, verso ambienti sempre più privati, funzionali e urbani, come l'appartamento e l'ufficio. Le variazioni nella distribuzione dei luoghi artificiali risultano coerenti con le trasformazioni storico-sociali del periodo, anche se con un'inerzia propria della rappresentazione letteraria, che non segue meccanicamente i cambiamenti del mondo reale ma li rielabora con ritardi e resistenze. Il lessico delle ambientazioni si conferma quindi un indicatore linguistico sensibile, capace di registrare non solo ciò che cambia, ma anche ciò che persiste e ciò che declina più lentamente.

I mutamenti si inseriscono nei più ampi processi di urbanizzazione e riorganizzazione della società tra XIX e XX secolo, che non incidono soltanto sulla struttura economica, ma trasformano in profondità anche le forme della vita quotidiana e le modalità di rappresentazione dello spazio. Il calo che si osserva nelle ambientazioni rurali e religiose e il parallelo aumento di luoghi legati alla sfera domestica e professionale riflettono il passaggio da una società fondata su comunità locali relativamente stabili e di stampo rurale a una realtà più mobile, frammentata e urbana. La narrativa registra, pur con tempi e modalità proprie, lo spostamento del baricentro della vita sociale dalla dimensione comunitaria a quella individuale e funzionale e i risultati sono coerenti con l'ipotesi iniziale di una progressiva riduzione delle ambientazioni rurali e religiose a favore di spazi urbani, privati e funzionali.

Il caso dei luoghi commerciali è particolarmente emblematico: la loro crescita nel primo Novecento e la successiva contrazione negli anni tra le due guerre si ricollegano ai cicli della prima globalizzazione e della deglobalizzazione descritti da O'Rourke e Williamson (1999); ciò suggerisce che la narrativa possa funzionare come archivio dei mutamenti dell'economia e dell'immaginario sociale.

Allo stesso tempo, le differenze tra opere italiane e tradotte suggeriscono che questi processi non si distribuiscono in modo uniforme: mentre la narrativa italiana conserva più a lungo una

forte presenza di luoghi pubblici e di incontro, soprattutto di tipo religioso, le opere tradotte mostrano una maggiore incidenza di spazi privati, che si ricollegano a modelli narrativi e culturali in parte diversi. Le ambientazioni riflettono anche differenze tra tradizioni letterarie e nelle modalità con cui queste vengono recepite e rielaborate nel contesto italiano. Il fatto che la narrativa italiana mantenga più a lungo una forte presenza di luoghi pubblici e religiosi, mentre le opere tradotte privilegino spazi privati, non è riducibile a una semplice differenza tematica: evidenza che le due componenti del corpus costruiscono lo spazio narrativo secondo logiche in parte diverse, legate a tradizioni letterarie, contesti culturali e aspettative di pubblico che non convergono necessariamente anche quando condividono la stessa lingua di pubblicazione. Inoltre, il confronto tra nativi e traduzioni fa anche emergere un possibile ruolo di anticipazione della letteratura tradotta nell'introduzione di nuovi spazi narrativi, come discusso nella sezione precedente. Il confronto tra narrativa nativa e tradotta non è quindi solo un controllo metodologico, ma una finestra su come l'italiano letterario abbia assorbito o rifiutato, rispecchiato o trasformato modelli spaziali provenienti da altre tradizioni.

I risultati qui presentati sollevano anche alcune questioni interpretative che meritano di essere problematizzate. La correlazione tra variazioni lessicali e trasformazioni storico-sociali è necessariamente ipotetica: il lessico delle ambientazioni riflette le scelte narrative degli autori, che sono a loro volta mediate da convenzioni letterarie, mode editoriali e aspettative di pubblico, e non traducono direttamente la realtà sociale. Allo stesso modo, il confronto tra narrativa nativa e tradotta è influenzato dalla composizione del corpus e dalla distribuzione non uniforme delle lingue di partenza nei diversi sottoperiodi. I dati suggeriscono tendenze, non spiegazioni causali, e la loro interpretazione richiede cautela.

Dal punto di vista metodologico, i risultati evidenziano al tempo stesso le potenzialità e i limiti dell'approccio adottato. Se infatti, da un lato, l'analisi quantitativa a partire da una lista lessicale permette di individuare tendenze difficili da cogliere attraverso una lettura ravvicinata, dall'altro richiede ulteriori verifiche, sia sul piano statistico²⁴ sia su quello interpretativo. L'analisi mostra come anche un approccio lessicale possa contribuire all'individuazione di *pattern* ricorrenti nella rappresentazione dello spazio narrativo su larga scala. È tuttavia opportuno sottolineare anche alcuni limiti: il lessico delle ambientazioni misura infatti la presenza esplicita di luoghi nel testo, ma non la loro centralità narrativa né il ruolo che assumono nella costruzione del significato: un romanzo ambientato prevalentemente in una chiesa può menzionare il termine una sola volta, mentre un testo che vi fa riferimento più volte può trattarlo come sfondo marginale. In altri termini, il metodo coglie tendenze di superficie, che richiedono di essere integrate con analisi più fini della funzione narrativa degli spazi.

Sarà comunque necessario lavorare su corpora più ampi e bilanciati²⁵ e integrare l'analisi con strumenti più raffinati, così da rendere più solide le ipotesi che si possono desumere dai risultati. In questa prospettiva, una prima direzione di ricerca riguarda l'estensione del metodo ad altre categorie di luoghi, in particolare quelli naturali, per analizzare in modo sistematico la relazione tra spazi artificiali e paesaggio, luoghi urbani e rurali (cfr. Grisot e Herrmann 2023; 2025) e

²⁴ Una possibile direzione riguarda l'applicazione di metodi di analisi delle serie temporali lessicali, come il *curve clustering*, già utilizzati in studi su corpora diacronici (cfr. Trevisani 2016). Approcci di questo tipo permettono di modellizzare l'andamento delle frequenze nel tempo e di individuare *pattern* evolutivi più robusti, tenendo conto sia delle differenze di dimensione tra sottocorpora sia della variabilità intrinseca dei dati testuali.

²⁵ Per esempio, quello proposto nell'ambito del progetto RIND (V. nota 6).

ricostruire nel tempo configurazioni più complesse dello spazio narrativo. Allo stesso tempo, l'approccio potrebbe essere applicato ad altri ambiti semantici, come i personaggi o le professioni, così da mettere in relazione le trasformazioni dello spazio con altri livelli della rappresentazione. Un ulteriore sviluppo consisterà in analisi più mirate su singoli generi o sottogeneri, che permettano di verificare se le tendenze osservate su larga scala si mantengano anche in contesti più omogenei. Ci si potrebbe chiedere, per esempio, se la narrativa poliziesca, ben rappresentata nel corpus tradotto del secondo dopoguerra, mostri una distribuzione degli spazi sistematicamente diversa rispetto al romanzo psicologico o a quello storico, e in che misura queste differenze siano attribuibili al genere piuttosto che alla provenienza culturale dell'opera. Rispondere a queste domande richiede un'integrazione tra l'approccio quantitativo qui adottato e strumenti di classificazione testuale più fini, così da mettere i dati in collegamento più diretto con la storia letteraria e con le interpretazioni critiche esistenti.

Riferimenti

- AA. VV. 2011. *I luoghi nella letteratura del Novecento*. I quaderni di Chichibìo 2. Palermo: Palumbo.
- Alfieri, Gabriella. 1994. «La lingua di consumo». In *Storia della lingua italiana*, vol. 2, a cura di Luca Serianni e Pietro Trifone. Torino: Einaudi.
- Anselmi, Gian Mario, e Gino Ruozzi, a cura di. 2003. *I luoghi della letteratura*. Milano: Mondadori.
- Antonelli, Giuseppe. 2011. «Lingua». In *Modernità italiana. Cultura, lingua e letteratura dagli anni Settanta a oggi*, a cura di Afriso, Andrea, e Emanuele Zinato, 15–52. Roma: Carocci.
- Aproso, Alessio Palmero, e Giovanni Moretti. 2018. «Tint 2.0: An All-Inclusive Suite for NLP in Italian». In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-It 2018): 10-12 December 2018, Torino*, a cura di Elena Cabrio, Alessandro Mazzei e Fabio Tamburini, 311-17. Torino: Accademia University Press. <https://doi.org/10.4000/books.aaccademia.3571>.
- Atlante Calvino. n.d. <https://atlantecalvino.unige.ch/>
- Bachtin, Michail. 1979. *Estetica e romanzo*. Torino: Einaudi.
- Bamman, David, Sejal Popat e Sheng Shen. 2019. «An Annotated Dataset of Literary Entities». In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2138-2144.
- Basile, Bruno. 2003. «Giardino». In *I luoghi della letteratura*, a cura di Gian Mario Anselmi e Gino Ruozzi, 213-21. Milano: Mondadori.
- Bornet, Cyril, e Frédéric Kaplan. 2017. «A Simple Set of Rules for Characters and Place Recognition in French Novels». *Frontiers in Digital Humanities* 4. <https://www.frontiersin.org/articles/10.3389/fdigh.2017.00006>.
- Collina, Beatrice. 2003. «Piazza». In *I luoghi della letteratura*, a cura di Gian Mario Anselmi e Gino Ruozzi, 285-97. Milano: Mondadori.
- De Mauro, Tullio. 1991. *Storia linguistica dell'Italia unita*. Bari: Laterza.
- De Mauro, Tullio. 2014. *Storia linguistica dell'Italia repubblicana*. Bari-Roma: Laterza.
- ELTeC (European Literary Text Collection). n.d. <https://distantreading.github.io/ELTeC/>.

- Even-Zohar, Itamar. 1990. «The Position of Translated Literature within the Literary Polysystem». *Poetics Today* 11 (1): 45-51.
- Feroldi, Donata, e Elena Dal Pra. 2011. *Dizionario analogico della lingua italiana*. Bologna: Zanichelli.
- Fowler, Alastair. 1982. *Kinds of Literature: An Introduction to the Theory of Genres and Modes*. Oxford: Clarendon Press.
- Garelli, Franco. 2007. *La Chiesa in Italia*. Bologna: il Mulino.
- Ginsborg, Paul. 2006. *Storia d'Italia dal dopoguerra a oggi: società e politica 1943-1988*. Torino: Einaudi.
- Grisot, Giulia, e Berenike Herrmann. 2023. «Examining the Representation of Landscape and Its Emotional Value in German-Swiss Fiction between 1840 and 1940». *Journal of Cultural Analytics* 8 (1). <https://doi.org/10.22148/001c.84475>.
- Grisot, Giulia, e Berenike Herrmann. 2025. “Is Heidi Really Happier in the Mountains? A Mixed-Methods Investigation of Spatial Affect in Fiction.” *Journal of Cultural Analytics*. <https://doi.org/10.22148/001c.124640>.
- Heuser, Ryan, e Long Le-Khac. 2012. *A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method*. Pamphlets of the Stanford Literary Lab 4. Stanford: Stanford Literary Lab. <https://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.
- Heuser, Ryan, Franco Moretti ed Erik Steiner. 2016. *The Emotions of London*. Stanford: Stanford Literary Lab.
- Iacoli, Giulio. 2018. «Spazio, romanzo e modernità». In *Il romanzo in Italia. 1. Forme, poetiche, questioni*, a cura di Giancarlo Alfano e Francesco de Cristofaro, 257-280. Roma: Carocci.
- Kababgi, Daniel, Giulia Grisot, Federico Pennino e Berenike Herrmann. 2024. «Recognising Non-Named Spatial Entities in Literary Texts: A Novel Spatial Entities Classifier». In *Computational Humanities Research 2024*, 472-481. CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3834/paper59.pdf>.
- Moretti, Franco. 1997. *Atlante del romanzo europeo*. Torino: Einaudi.
- Moretti, Franco. 2000. «Conjectures on World Literature». *New Left Review* 1: 54-68.
- Moretti, Franco. 2005. *La letteratura vista da lontano*. Torino: Einaudi.
- Moretti, Franco. 2007. *Graphs, Maps, Trees: Abstract Models for Literary History*. London e New York: Verso.
- Moretti, Franco. 2013. *Distant Reading*. London e New York: Verso.
- NerIta. n.d. https://huggingface.co/bullmount/it_nerIta_trf
- O'Rourke, Kevin H., e Jeffrey G. Williamson. 1999. *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy*. Cambridge, MA: MIT Press.
- Ondelli, Stefano. 2018. «Treat Texts as Data but Remember They Are Made of Words: Compiling and Pre-processing Corpora». In *Tracing the Life Cycle of Ideas in the Humanities and Social Sciences*, a cura di Arjuna Tuzzi, 133-50. Cham: Springer. https://doi.org/10.1007/978-3-319-97064-6_7.

- Ondelli, Stefano. 2019. «I verbi procomplementari nella prosa letteraria tra diacronia e traduzione». *Rivista internazionale di tecnica della traduzione / International Journal of Translation*. <https://doi.org/10.13137/2421-6763/29125>.
- Ricci, Laura. 2013. *Paraletteratura. Lingua e stile dei generi di consumo*. Roma: Carocci.
- Roventini, Adriana, Rita Marinelli, e Francesca Bertagna. 2016. *ItalWordNet v.2*. ILC-CNR for CLARIN-IT Repository, Institute for Computational Linguistics “A. Zampolli”. <https://dspace-clarin-it.ilc.cnr.it/items/3bcdffb9-ea4f-45f3-a321-6c32aa2f9f06>.
- Ryan, Marie-Laure, Kenneth Foote e Maoz Azaryahu. 2016. *Narrating Space / Spatializing Narrative: Where Narrative Theory and Geography Meet*. Columbus: Ohio State University Press.
- Sciandra, Andrea, Matilde Trevisani e Arjuna Tuzzi. 2021. «Sulle tracce dell’espressione dell’interiorità: analisi diacronica di un corpus di narrativa italiana del XIX-XX secolo». *Rivista internazionale di tecnica della traduzione / International Journal of Translation*. <https://doi.org/10.13137/2421-6763/33469>.
- Sciumbata, Floriana Carlotta e Luca Tringali. n.d. *Bran*. <https://github.com/zorbaproject/Bran>
- Sciumbata, Floriana Carlotta, Paolo Nadalutti e Luca Tringali. 2021. «“Trovare lavoro” in un corpus di narrativa del XIX-XX secolo. Procedure, aspetti e problemi di creazione, estrazione e rappresentazione dei dati». *Rivista internazionale di tecnica della traduzione / International Journal of Translation*. <https://doi.org/10.13137/2421-6763/33468>.
- Sciumbata, Floriana Carlotta. 2023. «Titles Associated with Power in a Corpus of Fictional Prose in Italian (XIX-XXI Century)». In *Jezik, Književnost, Moć / Language, Literature, Power*, a cura di Biljana Mišić Ilić, Vesna Lopičić, Ivana Mitić e Sanja Ignjatović, 279-97. Niš: Filozofski fakultet <https://dx.doi.org/10.46630/jkm.2023.17>
- Sciumbata, Floriana C. 2024. “Il traduttese nella narrativa.” In *Capitoli sul traduttese*, a cura di Stefano Ondelli e Floriana C. Sciumbata, 137–170. Trieste: EUT Edizioni Università di Trieste. <https://www.openstarts.units.it/handle/10077/37101>
- Silva, Matheus Oliveira e Mirella M. Moro. 2024. «PPORTAL_ner: An Annotated Corpus of Portuguese Literary Entities». In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 12927-12937.
- Straka, Milan. 2018. «UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task». In *Proceedings of the 22nd Conference on Computational Natural Language Learning (CoNLL 2018)*, 197-207. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/K18-2020>.
- Tally, Robert T. 2018. «The Space of the Novel». In *The Cambridge Companion to the Novel*, a cura di Eric Bulson, 152-67. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316659694.011>.
- Tint. n.d. <https://dh.fbk.eu/research/tint/>
- Trevisani, Matilde. 2016. “Analisi di dati testuali cronologici in corpora diacronici: effetti della normalizzazione sul curve clustering.” In *JADT 2016. Proceedings of the 13th International Conference on Statistical Analysis of Textual Data*, edited by Dominique Mayaffre, Christophe Poudat, Lorenzo Vanni, Valeria Magri, and Pauline Follette.
- Ubbidente, Roberto e Marie-Hélène Rybicki. 2010. *Luoghi di rappresentazione sociale nella letteratura italiana e francese (800/Novecento)*. Firenze: Franco Cesati.

Williams, Raymond. 1973 [1975]. *The Country and the City*. New York: Oxford University Press.

Appendice - Composizione e consistenza dei subcorpora

Tabella 1: autore, opera, anno di pubblicazione e consistenza dei testi italiani (1800-1899)

Autore	Titolo	Anno	Occorrenze
I. Nievo	<i>Il barone di Nicastro</i>	1860	54.453
G. Verga	<i>Storia di una capinera</i>	1871	31.529
C. Collodi	<i>Pinocchio</i>	1881	40.670
R. Fucini	<i>Le veglie di Neri: paesi e figure della campagna toscana</i>	1882	43.829
E. De Amicis	<i>Cuore</i>	1886	84.055
E. De Marchi	<i>Il cappello del prete</i>	1888	57.486
G. D'Annunzio	<i>Il piacere</i>	1889	105.195
L. Capuana	<i>Profumo</i>	1892	59.565
F. De Roberto	<i>I viceré</i>	1894	145.781
A. Fogazzaro	<i>Piccolo mondo antico</i>	1895	112.456
E. Salgari	<i>I pirati della Malesia</i>	1896	58.934
G. Deledda	<i>La via del male</i>	1896	67.091

Tabella 2: autore, opera, anno di pubblicazione, consistenza e traduttore (1800-1899)

Autore	Titolo	Anno	Occorrenze	Traduttore
J.P. Claris de Florian	<i>Numa Pompilio secondo re di Roma</i> (Tomo I)	1811	19.795	?
L. Sterne	<i>Viaggio sentimentale di Yorick lungo la Francia e l'Italia</i>	1813	43.500	D. Chierico
F.A. Chateaubriand	<i>Atala ovvero Gli amori di due selvaggi nel deserto</i>	1814	23.900	?
F. Schiller	<i>L'oste del sole ovvero Il delinquente per onore perduto</i>	1831	7.891	C. Belligoni
W. Scott	<i>Ivanhoe</i> (vol. III)	1832	34.612	A. Clerichetti
H. de Balzac	<i>Papà Goriot</i>	1835	85.448	L. Masieri

V. Hugo	<i>Le ore estreme di un sentenziato a morte</i>	1835	39.367	G.B. Carta
W. Scott	<i>Il castello pericoloso</i>	1843	38.537	L.M.
A.R. Lesage	<i>Storie di Gil Blas di Santillano</i> (vol. I)	1853	137.639	Q. Viviani
A. Dumas	<i>I tre moschettieri</i> (vol. I)	1855	43.844	C. Coriani
C. Dickens	<i>Le campane ovvero Il capo d'anno</i>	1855	50.924	P. Bettoni
W. Goethe	<i>I dolori del giovane Werther</i>	1858	52.044	R. Ceroni
E.A. Poe	<i>Storie incredibili (L'uomo della folla)</i>	1869	55.365	B.E. Maineri
L. Carroll	<i>Alice nel paese delle meraviglie</i>	1872	22.687	T. Pietrocola Rossetti
I.S. Turgenev	<i>Acque di primavera</i>	1872	45.718	S. de Gubernatis-Besobrasoff
J. Verne	<i>Dalla terra alla luna. Tragitto in 97 ore e 20 minuti</i>	1872	52.248	?
V. Hugo	<i>Novantatré</i> (vol. II)	1874	40.412	C. Pizzigoni
I.S. Turgenev	<i>Il primo amore e Assia</i>	1876	39.276	E. Zucchelli
C. Dickens	<i>Cantico di Natale</i>	1888	26.815	F. Verdinois

Tabella 3: autore, opera, anno di pubblicazione e consistenza dei testi italiani (1900-1945)

Autore	Titolo	Anno	Occorrenze
A. Fogazzaro	<i>Piccolo mondo moderno</i>	1901	92.355
G. D'Annunzio	<i>Le novelle della Pescara</i>	1902	50.123
L. Pirandello	<i>Il fu Mattia Pascal</i>	1904	75.010
E. Salgari	<i>Gli ultimi filibustieri</i>	1908	90.030
E. Salgari	<i>Il figlio del Corsaro Rosso</i>	1908	109.410
Vamba	<i>Il giornalino di Gian Burrasca</i>	1912	74.340
E. Salgari	<i>La rievocata di Yanez</i>	1912	80.225
G. Deledda	<i>Canne al vento</i>	1913	60.430
L. Pirandello	<i>I vecchi e i giovani</i>	1913	109.430
F. Tozzi	<i>Bestie</i>	1917	15.624

F. Tozzi	<i>Con gli occhi chiusi</i>	1919	47.445
F. Tozzi	<i>Tre croci</i>	1920	28.272
I. Svevo	<i>La coscienza di Zeno</i>	1923	145.433
L. Pirandello	<i>Uno nessuno centomila</i>	1926	46.414
I. Svevo	<i>Una burla riuscita</i>	1926	21.330
L. Pirandello	<i>Una giornata</i>	1937	30.319
C. Levi	<i>Cristo si è fermato a Eboli</i>	1945	81.122
E. Vittorini	<i>Uomini e no</i>	1945	48.668

Tabella 4: autore, opera, anno di pubblicazione, consistenza e traduttore (1900-1945)

Autore	Titolo	Anno	Occorrenze	Traduttore
L. Tolstoj	<i>Anna Karenina</i>	1901	327.334	E. V. Foulques
E. Brontë	<i>Cime tempestose</i>	1926	117.562	R. Binetti
F. Dostoevskij	<i>L'idiota</i>	1927	245.393	F. Verdinois
C. Dickens	<i>Il circolo Pickwick</i>	1930	296.865	F. Verdinois
N. Hawthorne	<i>La lettera scarlatta</i>	1930	80.331	F.M. Martini
R. Stout	<i>La lega degli uomini spaventati</i>	1937	57.799	A. Pitta
A. Christie	<i>Il Natale di Poirot</i>	1940	50.929	E. Piceni
Stendhal	<i>La badessa di Castro</i>	1942	34.392	P.P. Trompeo

Tabella 5: autore, opera, anno di pubblicazione e consistenza dei testi italiani (1946-1975)

Autore	Titolo	Anno	Occorrenze
E. Flaiano	<i>Tempo di uccidere</i>	1947	83.113
C. Malaparte	<i>La pelle</i>	1949	111.759
A. Moravia	<i>Il conformista</i>	1951	99.957
I. Calvino	<i>Il barone rampante</i>	1957	71.189
F. Tomasi di Lampedusa	<i>Il Gattopardo</i>	1958	74.657
O. Ottieri	<i>Donnarumma all'assalto</i>	1959	60.914

P. Chiara	<i>Il piatto piange</i>	1962	45.531
N. Ginzburg	<i>Lessico familiare</i>	1963	65.496
A. Bevilacqua	<i>L'occhio del gatto</i>	1968	86.219
L. Romano	<i>Le parole tra noi leggere</i>	1969	94.795
G. Piovene	<i>Stelle fredde</i>	1970	21.030
D. Maraini	<i>Memorie di una ladra</i>	1972	99.340

Tabella 6: autore, opera, anno di pubblicazione, consistenza e traduttore (1946-1975)

Autore	Titolo	Anno	Occorrenze	Traduttore
K. Millar	<i>La città del diavolo</i>	1950	53.264	A. Tedeschi
E. Queen	<i>Il re è morto</i>	1954	57.556	B. Tasso
R. Stout	<i>Nero Wolfe e i ragni d'oro</i>	1954	42.760	I. Omboni
E.S. Gardner	<i>Perry Mason e la sveglia sotterrata</i>	1955	44.599	G. Gogioso
E. Queen	<i>Il villaggio di vetro</i>	1957	78.082	G. Camia
U. Curtiss	<i>Ombre di cera</i>	1958	48.071	P. Soleri
C.E. Maine	<i>Rischio calcolato</i>	1961	45.829	A. Negretti
F. Brown	<i>Gli strani suicidi di Bartlesville</i>	1962	47.756	M. Galli
M. Leinster	<i>Questo è un Gizmo</i>	1962	38.019	B. Russo
R. Stout	<i>Nero Wolfe fa la spia</i>	1964	48.133	L. Grimaldi
J.H. Chase	<i>Canaglia cervasi</i>	1964	53.790	H. Brinis
I. Levin	<i>Rosemary's Baby</i>	1967	60.716	A. Veraldi
D. Koontz	<i>Jumbo-10. Il rinnegato</i>	1969	32.488	B. Della Frattina
M. Zimmer Bradley	<i>Il drago di bronzo</i>	1972	45.663	M.B. De Castiglione
C.D. Simak	<i>Pescatore di stelle</i>	1973	75.892	R. Rambelli
M.G. Eberhart	<i>Il giudice buonanima</i>	1974	68.482	C. Merlo
R. Silverberg	<i>Vacanze nel deserto</i>	1975	77.696	G. Tamburini

Tabella 7: autore, opera, anno di pubblicazione e consistenza dei testi italiani (1976-2005)

Autore	Titolo	Anno	Occorrenze
F. Cialente	<i>Le quattro ragazze Wieselberger</i>	1976	67.486
F. Tomizza	<i>La miglior vita</i>	1977	94.441
A. De Carlo	<i>Uccelli da gabbia e da voliera</i>	1982	70.610
C. Sgorlon	<i>L'armata dei fiumi perduti</i>	1985	88.049
U. Eco	<i>Il pendolo di Foucault</i>	1988	205.964
A. Baricco	<i>Castelli di rabbia</i>	1991	53.052
S. Veronesi	<i>Venite venite B-52</i>	1995	95.519
N. Ammaniti	<i>Fango</i>	1996	33.936
M. Buticchi	<i>Le pietre della luna</i>	1997	120.641
G. Faletti	<i>Io uccido</i>	2002	199.268
M. Mazzantini	<i>Non ti muovere</i>	2002	76.093
V.M. Manfredi	<i>L'impero dei draghi</i>	2005	121.571

Tabella 8: autore, opera, anno di pubblicazione, consistenza e traduttore (1976-2005)

Autore	Titolo	Anno	Occorrenze	Traduttore
A. Christie	<i>Addio miss Marple</i>	1976	55.688	D. Fonticoli
M. Crichton	<i>Congo</i>	1980	89.529	E. Capriolo
S. King	<i>Pet Sematary</i>	1985	132.829	H. Brinis
P. Süskind	<i>Il profumo</i>	1988	81.071	G. Agabio
R. Dahl	<i>Storie impreviste</i>	1993	90.942	A. Veraldi
H. Murakami	<i>Dance dance dance</i>	1998	116.264	G. Amitrano
H. Corbin	<i>Cadaveri senza volto</i>	2000	104.456	F. Sessi
K. Follett	<i>Le ragazze ladre</i>	2001	136.296	A. Raffo
I. Allende	<i>Il regno del drago d'oro</i>	2003	86.493	E. Liverani
C. Ruiz Zafón	<i>L'ombra del vento</i>	2004	134.436	L. Sezzi
T. Yamada	<i>Estranei</i>	2005	45.949	A. Martini
P. Coelho	<i>Lo Zahir</i>	2005	81.235	R. Desti