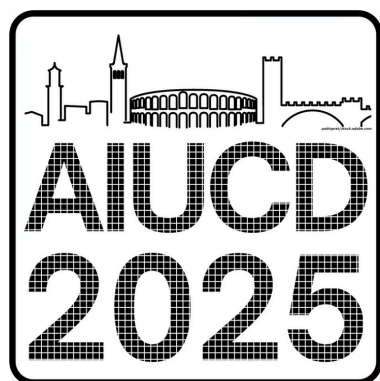




Atti del XIV Convegno Annuale

**Diversità, Equità e Inclusione: Sfide e
Opportunità per l'Informatica Umanistica
nell'Era dell'Intelligenza Artificiale**

Verona :: 11-13 giugno 2025

A cura di:

Simone Rebora • Marco Rospocher • Stefano Bazzaco



**UNIVERSITÀ
di VERONA**
Dipartimento
di LINGUE
E LETTERATURE STRANIERE



ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE

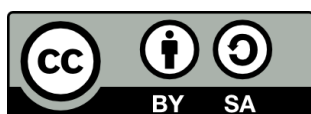


ISBN 978-88-942535-9-7



Copyright ©2025 AIUCD

Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution ShareAlike 4.0 International license (CC-BY-SA 4.0). Ogni altro diritto rimane in capo ai singoli autori.

This volume and all contributions are released under the Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). All other rights retained by the legal owners.

A cura di: Simone Rebora; Marco Rospocher; Stefano Bazzaco (2025). Diversity, Equity, and Inclusion: Challenges and Opportunities for Digital Humanities in the Age of Artificial Intelligence, Proceedings del XIV Convegno Annuale AIUCD, Verona 11-13 giugno 2025, Università di Verona.

Ultimo accesso agli URL in data 8 maggio 2025.

Si prega di notificare all'editore ogni omissione o errore si riscontri: segreteria [at] aiucd.org

Please notify the publisher of any omissions or errors found: segreteria [at] aiucd.org

Il programma della conferenza AIUCD 2025 è disponibile online

<https://aiucd2025.dlss.univr.it/detailed-schedule/>

The AIUCD 2025 Conference Program is available online

<https://aiucd2025.dlss.univr.it/en-gb/detailed-schedule/>

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante double-blind peer review, effettuata dai membri del Comitato di Programma sotto la supervisione del Comitato Scientifico di AIUCD 2025.

All the papers published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review, carried out by the members of the Programme Committee under the supervision of the Scientific Committee of AIUCD 2025.

Gli atti del convegno AIUCD 2025 sono pubblicati come raccolta di contributi in formato PDF forniti direttamente dagli autori e dalle autrici. I file sono stati raccolti e assemblati senza interventi redazionali da parte dei curatori.

The proceedings of the AIUCD 2025 conference are published as a collection of PDF contributions provided directly by the authors. The files have been collected and compiled without editorial intervention by the editors..

Il logo di AIUCD 2025 include l'immagine "Verona Dark Line Simple Minimalist Skyline With White Background" di @pabloprat/stock.adobe.com, ottenuta tramite la licenza Adobe Stock dell'Università di Verona.

The AIUCD 2025 logo includes the image "Verona Dark Line Simple Minimalist Skyline With White Background" by @pabloprat/stock.adobe.com, used under the Adobe Stock license of the University of Verona.

Il background della copertina è stato creato con tecniche di AI generativa con lo strumento "Magic Media" disponibile su Canva, usando un prompt con il tema del convegno.

The background of the cover was created using generative AI techniques with the "Magic Media" tool available on Canva, using a prompt based on the conference theme.

Linguistica dei corpora e informatica umanistica per la valorizzazione plurilingue del patrimonio culturale: implementazione del progetto UniVOCittà

Rita Gramellini¹, Valeria Zotti²

¹ Alma Mater Studiorum Università di Bologna, Italia - rita.gramellini2@unibo.it

² Alma Mater Studiorum Università di Bologna, Italia - valeria.zotti@unibo.it

ABSTRACT (ITALIANO)

La fruizione del patrimonio culturale da parte di un pubblico internazionale rappresenta un'esigenza sempre più rilevante nella società contemporanea. In tale prospettiva si colloca il progetto *UniVOCittà*, sviluppato dall'unità di ricerca LBC-CeSLiC del Dipartimento di Lingue, Letterature e Culture Moderne dell'Università di Bologna. Il progetto si inserisce tra le ricerche nel campo dell'informatica umanistica e propone una metodologia innovativa volta alla valorizzazione plurilingue del patrimonio artistico e culturale di Bologna attraverso la creazione di un corpus comparabile plurilingue annotato, composto da scritti di viaggio di personaggi illustri, dal Rinascimento all'epoca contemporanea.

In collaborazione con specialisti di *Web Technologies* dell'Università di Bologna, è stato realizzato il prototipo di un'applicazione web mobile destinata al turismo culturale, che consente la consultazione della banca dati plurilingue tramite visualizzazioni sia lineari che dinamiche. Poiché le testimonianze letterarie sono attualmente disponibili solo nella lingua originale, è in corso un'implementazione della app nell'ambito di un progetto di dottorato finalizzato all'integrazione delle traduzioni dei contenuti in lingua italiana e nelle altre lingue del progetto. A tal fine, si prevede di testare un sistema di traduzione automatica neurale (DeepL), e un chatbot generativo di ultima generazione (ChatGPT 4o). Tuttavia, considerati i limiti di questi strumenti di intelligenza artificiale nella traduzione di lessemi e termini relativi alle belle arti, gli *output* verranno sottoposti a una fase di revisione manuale mediante la consultazione di un sotto-corpus comparabile di scritti di viaggio in lingua italiana, attualmente in fase di realizzazione. L'obiettivo finale è di offrire a utenti provenienti da diversi orizzonti linguistici una risorsa digitale che favorisca l'accesso e la valorizzazione del patrimonio artistico e culturale bolognese in una prospettiva plurilingue.

Parole chiave: patrimonio culturale digitale; corpus comparabile; applicazione web; turismo culturale; traduzione automatica

ABSTRACT (ENGLISH)

Corpus Linguistics and Digital Humanities for the Multilingual Valorisation of Cultural Heritage: Implementation of the UniVOCittà Project. The enjoyment of cultural heritage by an international audience represents an increasingly relevant need in contemporary society. Within this perspective, the *UniVOCittà* project was developed by the LBC-CeSLiC research unit of the Department of Modern Languages, Literatures and Cultures at the University of Bologna. The project is situated within the field of digital humanities and proposes an innovative methodology aimed at the multilingual valorization of Bologna's artistic and cultural heritage through the creation of an annotated multilingual corpus collecting travel writings by illustrious personalities from the Renaissance to the present day.

In collaboration with Web Technologies specialists from the University of Bologna, a prototype of a mobile web app for cultural tourism was developed. This application enables the consultation of the multilingual textual database through both linear and dynamic visualizations. As the literary testimonies are currently only available in the original languages, an ongoing implementation of the application within a doctoral project intends to integrate translations of all content into Italian and the other project languages. To this end, the project plans to test a neural automatic translator (DeepL) and a state-of-the-art generative chatbot (ChatGPT 4o). However, due to the limitations of these artificial intelligent systems in the translation of lexemes and terms related to the fine arts, the outputs will undergo a manual revision phase through the consultation of a comparable Italian sub-corpus of travel writings, currently under development. The ultimate goal is to provide users from diverse linguistic backgrounds with a digital resource that facilitates the valorization and access to Bologna's artistic and cultural heritage from a multilingual perspective.

Keywords: digital cultural heritage; comparable corpus; web application; heritage tourism; machine translation

1. INTRODUZIONE

Nella società contemporanea, l'ampia fruizione del patrimonio culturale in formato digitale costituisce un'esigenza di crescente rilevanza. In questo contesto, l'informatica umanistica si configura come un ambito promettente, in quanto in grado di offrire strumenti e metodologie efficaci per la realizzazione di contenuti culturali digitali fruibili in formato elettronico (Barbuti, 2018; Münster et al., 2019). Affinché le pratiche tecnologiche non si limitino a una semplice transizione in digitale dei beni culturali (Donati, 2019: 326), risulta imprescindibile adottare approcci orientati alla diffusione e fruizione del patrimonio digitale da parte di un pubblico eterogeneo e internazionale.

Il presente contributo illustra una nuova fase di implementazione del progetto *UniVOCIttà*, condotto nel biennio 2022-2023 dall'unità di ricerca LBC-CeSLiC¹ (Centro di Studi Linguistico-Culturale) del Dipartimento di Lingue, Letterature e Culture moderne (LILEC) dell'Università di Bologna. Il progetto si propone di promuovere la conoscenza e la diffusione internazionale del patrimonio artistico e culturale della città di Bologna e della sua provincia, mediante la creazione di una banca dati multilingue di scritti di viaggio, consultabile attraverso una piattaforma interattiva per il turismo culturale.

Nonostante l'applicazione sia stata ideata come una risorsa plurilingue destinata a un pubblico internazionale, la fruizione dei testi è attualmente circoscritta agli utenti in grado di comprendere le lingue originali dei documenti. L'attuale fase del progetto prevede, pertanto, la traduzione dei segmenti testuali in lingua italiana e nelle altre lingue del progetto, attraverso la sperimentazione di due strumenti di intelligenza artificiale: il traduttore automatico neurale DeepL, e il chatbot generativo di ultima generazione ChatGPT 4o. Considerati i limiti comprovati di tali "sistemi artificiali intelligenti" (Andler, 2023: 23) nella traduzione di lessemi e termini afferenti al dominio delle belle arti, gli *output* verranno sottoposti a una revisione linguistica e concettuale, supportata dalla consultazione di un sotto-corpus comparabile italiano di letteratura di viaggio, attualmente in fase di elaborazione. L'obiettivo finale del progetto è la realizzazione di una risorsa digitale in grado di facilitare la fruizione del patrimonio culturale e artistico bolognese da parte di utenti provenienti da diversi contesti linguistici e culturali.

Dopo aver delineato la genesi del progetto *UniVOCIttà*, il presente contributo ne illustra in primo luogo la metodologia adottata e i risultati finora conseguiti, con un focus specifico sul processo di costruzione del corpus testuale plurilingue, sull'annotazione linguistica dei dati e sullo sviluppo di un'applicazione web mobile destinata al turismo culturale. Nella parte conclusiva, si presentano le prospettive di sviluppo futuro, che includono l'integrazione di traduzioni multilingue dei contenuti dell'app mediante il test di due sistemi di intelligenza artificiale.

2. GENESI DEL PROGETTO

Tra le diverse iniziative di informatica umanistica finalizzate alla diffusione del patrimonio tangibile e intangibile italiano, si distingue il progetto "Lessico dei Beni Culturali"² (LBC), avviato nel 2013 da Annick Farina presso l'Università di Firenze. Tale progetto, che attualmente coinvolge diverse università italiane e straniere, nasce dall'esigenza di sviluppare risorse linguistiche digitali plurilingui (banche dati testuali, corpora comparabili, corpora paralleli, lessici) volte a favorire l'accesso internazionale alla conoscenza del patrimonio culturale italiano, iniziando da quello toscano (Farina, 2016; Farina & Martínez, 2020: 8). Nell'ambito di questo progetto interuniversitario, un gruppo di linguiste del Dipartimento di Lingue, Letterature e Culture moderne (LILEC) dell'Università di Bologna ha costituito nel 2018 l'unità di ricerca LBC-CeSLiC (Centro di Studi Linguistico-Culturali), responsabile della costruzione di un sotto-corpus comparabile in quattro lingue (inglese, francese, spagnolo, russo) dedicato al patrimonio architettonico, artistico, museale e naturalistico della città di Bologna e della regione Emilia-Romagna, denominato BER. Questa risorsa raccoglie testi appartenenti a diverse tipologie testuali (diari di viaggio, romanzi, blog turistici, manuali tecnici, saggi di critica d'arte), i quali vengono suddivisi in tre macrocategorie definite dal progetto madre fiorentino: testi letterari, divulgativi e tecnici (Perotto & Zotti, 2022: 168). Sulla scia delle attività condotte nell'ambito del progetto LBC, l'unità di ricerca LBC-CeSLiC ha avviato il progetto di disseminazione *UniCittà: Riscoprire Bologna attraverso l'Alma Mater Studiorum*, finanziato tra il 2019 e il 2021 dalla Fondazione Cassa di Risparmio di Bologna all'interno del bando "RiScopriamo la città". *UniCittà* è stato concepito con l'obiettivo di diffondere a un pubblico internazionale la conoscenza del patrimonio immobile (musei, biblioteche, archivi, palazzi storici) e mobile (opere d'arte, collezioni librerie) dell'Università di Bologna mediante un corpus plurilingue digitale di testimonianze presenti e passate che

¹ L'unità di ricerca è composta da Valeria Zotti (referente, francese), Ana Pano Alamán (spagnolo), Monica Turci e Antonella Luporini (inglese), Monica Perotto (russo).

² <https://www.lessicobeniculturali.net/> (cons. 10/04/2025)

attestano la ricchezza storica e culturale dell'Alma Mater (Zotti, 2023: 199). I testi selezionati sono stati dapprima sottoposti a un processo di annotazione con etichette (*tags*) che identificano otto categorie rappresentative del patrimonio universitario³. Successivamente, sono stati resi disponibili tramite una banca dati "annotata", consultabile gratuitamente in versione web⁴ e sul repository istituzionale dell'Università di Bologna.⁵

La metodologia e i risultati ottenuti nell'ambito di *UniCittà* hanno posto le basi per un progetto derivato condotto dall'unità di ricerca LBC-CeSLiC, volto alla valorizzazione del potenziale turistico-culturale dell'intero patrimonio bolognese: il progetto *UniVOCittà*.

3. METODOLOGIA

Il progetto di ricerca *UniVOCittà: Voci digitali sull'unicità del patrimonio bolognese*, finanziato con fondi NextGenerationEU nell'ambito del bando AlmaIdea 2022 e coordinato da Valeria Zotti, si propone di preservare e promuovere il patrimonio tangibile (luoghi di culto, piazze, mura, torri) e intangibile (parlata, cucina, tradizioni) di Bologna e della sua provincia attraverso un corpus di testimonianze letterarie in lingua originale redatte da personaggi illustri (letterati, artisti, scienziati, politici) che hanno visitato il capoluogo emiliano nel corso dei secoli, il corpus BER-LET.

Come menzionato da Zotti (2025), la selezione dei testi è stata effettuata seguendo i criteri definiti da L'Homme (2004: 126-127) per la costituzione di un corpus specialistico, così articolati:

- Ambito di specialità: patrimonio culturale tangibile e intangibile della città di Bologna e della sua provincia;
- Lingue: inglese, francese, spagnolo, russo;
- Livello di specializzazione: specializzazione media;⁶
- Tipo di documento: testi appartenenti al genere della letteratura di viaggio, quali resoconti, diari di viaggio, corrispondenze epistolari, romanzi;
- Supporto: testi scritti, in formato cartaceo o digitale;
- Periodo di pubblicazione: dal XVII al XXI secolo.

Per garantire una copertura esaustiva delle fonti, la ricerca dei testi è stata principalmente guidata dalle numerose ricerche disponibili sul fenomeno del *Grand Tour*, concentrandosi sugli studi relativi ai viaggiatori che hanno esplorato l'Emilia-Romagna (Zotti, 2025). Tra le opere di riferimento si distinguono l'antologia *Viaggi e viaggiatori del Settecento in Emilia e in Romagna* curata da Giorgio Cusatelli (1986) e i contributi di Attilio Brilli (2005; 2014). Le opere di pubblico dominio sono state reperite dai principali archivi e biblioteche digitali internazionali.⁷ Quando disponibili esclusivamente in formato cartaceo, i testi sono stati digitalizzati mediante la tecnologia OCR (Riconoscimento Ottico dei Caratteri) dai membri dell'unità e da un gruppo di tirocinanti.

Al termine del progetto, il corpus risulta composto da oltre 200 testi, di cui circa il 45% redatti in lingua francese, il 35% in inglese, il 15% in spagnolo e il 10% in russo. Nonostante l'intento di garantire una rappresentazione equilibrata delle quattro lingue coinvolte, la distribuzione effettiva dei materiali testuali riflette le diverse tradizioni di viaggio nel corso dei secoli. Non sorprende, ad esempio, la predominanza delle opere di viaggiatori inglesi e francesi, che furono tra i principali protagonisti del *Grand Tour*.

I documenti raccolti sono stati successivamente annotati manualmente con il software di analisi qualitativa Atlas.ti⁸ (Muhr, 1991) seguendo un approccio *bottom-up*. L'osservazione empirica dei testi nelle quattro lingue ha consentito di identificare le macrocategorie più rilevanti del patrimonio artistico e culturale dell'area di interesse, come Luoghi del Patrimonio ("LU"); Opere d'arte ("OPA"); Usi e costumi ("USIeCOSTUMI"), e le sotto-aree ("Museo"; "Zona Universitaria"; "Portico", "Pittura"; "Architettura"; ecc.). Complessivamente, sono stati annotati 1860 frammenti testuali per le quattro lingue del progetto.

³ Artisti ("ART"); Libri ("LIB"); Luoghi ("LU"); Oggetti ("OGG"); Opere d'arte ("OPA"); Periodo storico ("PE"); Personaggi storici ("PERS"); Tematica ("TEM"); Titoli di libri ("TIL").

⁴ <https://site.unibo.it/unicitta/it> (cons. 24/01/2025)

⁵ <https://amsacta.unibo.it/id/eprint/7166/> (cons. 24/01/2025)

⁶ "La pertinenza di un testo a livello alto, medio o basso è determinata dalle caratteristiche del pubblico, del mezzo e dello scopo. Ad esempio, un testo prodotto da uno specialista per gli studenti può essere definito di "livello medio" (Cabré, 2007-2008: 40, trad. mia). Nel caso del corpus BER-LET, i testi di letteratura di viaggio raccolti sono stati scritti da autori colti e destinati a un pubblico di lettori di cultura medio-alta (Zotti, 2025).

⁷ British Library, Library of Congress, LexisNexis, Internet Archive, HathiTrust, BNF Gallica, Bibliothèque et Archives nationales du Québec, Persée, Biblioteca russa di stato, Biblioteca presidenziale B. Eltsin, Biblioteca Digital Hispánica della Biblioteca Nacional de España, Biblioteca Virtual Miguel de Cervantes.

⁸ <https://atlasti.com/> (cons. 24/01/2025)

4. RISULTATI

I frammenti annotati sono stati resi disponibili tramite un'applicazione web mobile multilingue⁹ sviluppata in collaborazione con il Dipartimento di Ingegneria e Scienze Informatiche (DISI) e il Dipartimento di Filologia Classica e Italianistica (FICLIT) dell'Università di Bologna, sotto la supervisione di Fabio Vitali, professore ordinario di Informatica ed esperto di *Web Technologies*. Questa piattaforma, disponibile in tutte e quattro le lingue del progetto, intende mettere i testi digitalizzati al servizio del grande pubblico e degli operatori coinvolti nell'industria del turismo, al fine di promuovere Bologna come una meta di turismo culturale esperienziale (Apostolakis, 2003: 799; Zotti, 2023: 385). L'applicazione si articola in quattro sezioni principali ("Progetto", "Su di noi", "Corpus" e "Mappe") e offre due modalità distinte di esplorazione del corpus letterario. Nella sezione "Corpus", l'utente ha accesso ai frammenti testuali attraverso una visualizzazione lineare, con la possibilità di effettuare ricerche incrociate mediante l'impiego dei metadati etichettati come filtri di ricerca (vd. Fig. 1) (Zotti, 2025). Al fine di garantire una migliore fruizione del patrimonio letterario digitalizzato, ogni citazione è accompagnata da un collegamento ipertestuale al testo integrale in formato elettronico, permettendo la consultazione approfondita delle fonti originali.

La sezione "Mappe" propone invece una visualizzazione dinamica di itinerari culturali georeferenziati che consentono ai visitatori odierni di esplorare in tempo reale i luoghi del patrimonio bolognese seguendo le "voci" dei viaggiatori che li hanno preceduti nel corso dei secoli (vd. Fig. 2).

L'applicazione UniVOCittà contribuisce così ad arricchire l'esperienza del territorio mediante lo *storytelling* virtuale, una strategia narrativa che collega racconti ed emozioni personali alla visita dei siti culturali (Roque, 2022: 26-27).

Filtri

Trova

Tutte le lingue ▼ Tutti i luoghi ▼ Tutti gli autori ▼ Tutte le epoche ▼

Tutte le esperienze ▼ Tutti i personaggi ▼ Tutti gli artisti ▼ Tutte le città ▼

All Nature ▼ All Themes ▼ All Artworks ▼

Risultati per pagina: 5 ▼

Frammento	Fonte
<i>Il nous tardait d'arriver à Bologne, de saluer la ville où le même siècle vit naître Annibal Carrache, le Dominiquin, le Guide, l'Albane et le Guerchin.</i>	Ducos, Basile Joseph Link
<i>Son éminence n'assistait point aujourd'hui à la représentation de Richard Cœur-de-Lion. La musique italienne de cet opéra est médiocre. Grétry en a rendu la répétition difficile. Il serait donné à peu de compositeurs de faire oublier les motifs variés et spirituels du musicien liégeois, la légèreté et la grâce de ses accompagnements, et surtout les accens de la touchante romance de Blondel.</i>	Ducos, Basile Joseph Link

Figura 1. La visualizzazione lineare nella sezione "Corpus"

⁹ <https://univocitta.github.io/UniVOCittà/> (cons. 24/01/2025). Poiché sviluppata nell'ambito di un progetto finanziato con durata annuale, l'applicazione web mobile del progetto è stata concepita come prototipo per testare l'approccio metodologico. Data la natura sperimentale di questa fase, sono previsti sviluppi futuri per migliorare l'accessibilità e conformare la piattaforma agli standard previsti dalle direttive UE e alle linee guida definite dall'Agenzia per l'Italiana Digitale (AgID).

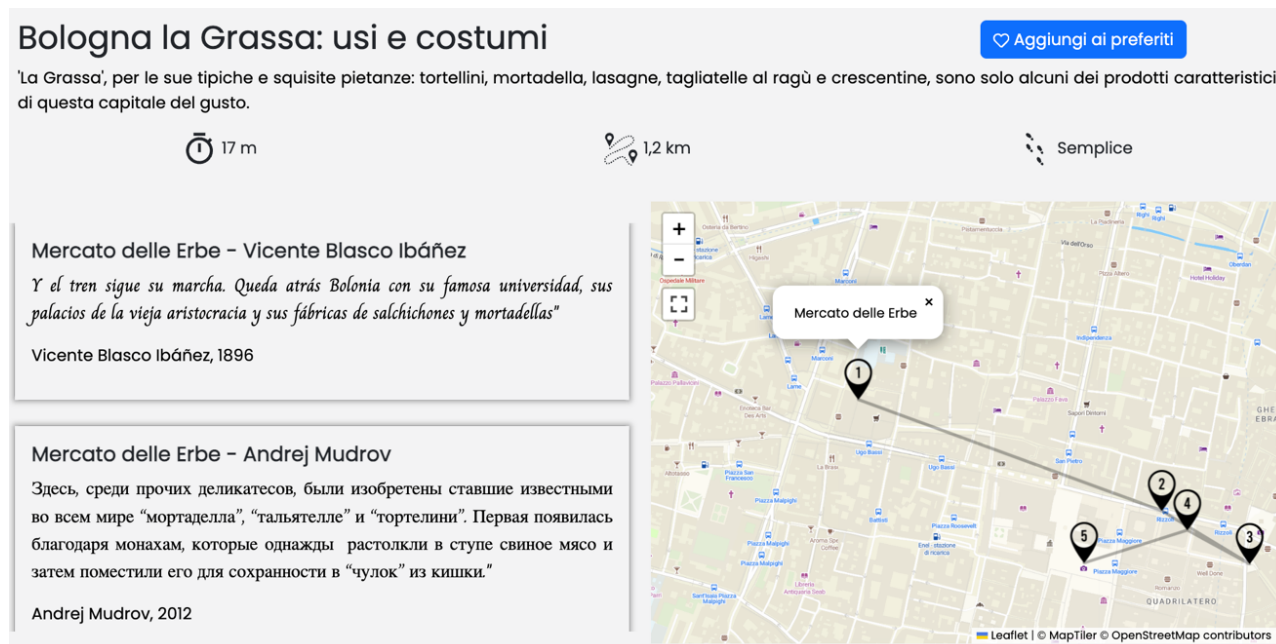


Figura 2. La visualizzazione dinamica nella sezione "Mappe"

5. IMPLEMENTAZIONI IN CORSO E FUTURE

Sebbene l'applicazione sia stata concepita come una risorsa plurilingue destinata a un pubblico internazionale, l'accesso ai frammenti testuali del corpus comparabile BER-LET risulta attualmente limitato a coloro che conoscono le lingue di partenza. Le testimonianze letterarie nelle sezioni "Corpus" e "Mappe" sono infatti disponibili esclusivamente in lingua originale (es. Dickens in inglese, Stendhal in francese, Cervantes in spagnolo, Tolstoj in russo), come si evince dalla Figura 2.

Per ovviare a questo limite, una nuova fase è attualmente in corso nell'ambito di un progetto di dottorato e prevede l'integrazione delle traduzioni delle citazioni in lingua italiana e nelle altre lingue del progetto. Tale implementazione mira a favorire una maggiore fruibilità dei contenuti dell'applicazione, rendendo questa esperienza inedita del patrimonio disponibile alla cittadinanza locale, al pubblico italiano, e a turisti e studiosi di diversi orizzonti linguistici e culturali.

In mancanza di traduzioni edite della maggior parte dei testi del corpus e quindi dell'impossibilità di costituire un corpus parallelo, è stato stabilito di testare l'efficacia di un software di traduzione automatica neurale, DeepL, e di un chatbot generativo di ultima generazione, ChatGPT versione 4o. Questi sistemi presentano delle differenze significative sia nella loro architettura che nei dati utilizzati per l'addestramento (Hendy et al., 2023). I modelli di traduzione automatica neurale, come DeepL, si basano tipicamente su un'architettura *encoder-decoder* e vengono addestrati su corpora paralleli di grandi dimensioni allineati a livello di frase (Koehn, 2020: 125; Pérez-Ortiz et al., 2022: 143; Yvon, 2023: A19). Diversamente, GPT, così come altri modelli generativi di grandi dimensioni (*Large Language Models* - LLM), si fonda su un'architettura *decoder-only* e apprende in maniera automatica come generare porzioni di testo coerenti e ben strutturate attraverso l'esposizione a enormi quantità di dati principalmente monolingui, estratti da fonti eterogenee disponibili sul Web (Ciotti, 2023: 299). Sebbene non sia stato specificatamente progettato per svolgere compiti di traduzione, recenti studi dimostrano che ChatGPT è in grado di produrre traduzioni efficaci e accurate, pur presentando ancora diverse criticità (Jiao et al., 2023; Hendy et al., 2023). Questi risultati suggeriscono dunque la necessità di condurre ulteriori ricerche che mettano a confronto e valutino le prestazioni di diversi modelli di intelligenza artificiale, soprattutto in settori specialistici tradizionalmente meno esplorati, come quello delle belle arti.

Alcuni test preliminari condotti sul sotto-corpus francese BER-LET hanno evidenziato numerosi limiti di ChatGPT 4o e DeepL nella resa in italiano di lessemi e termini relativi al patrimonio artistico e culturale a causa della mancanza di addestramento su dati di dominio specializzati. A titolo esemplificativo, entrambi i sistemi hanno mostrato delle approssimazioni nella traduzione di combinazioni lessicali specializzate, come nel caso di *piliers carrés et ronds*, che è stato tradotto letteralmente come "pilastri quadrati e rotondi", senza cogliere la distinzione semantica e concettuale propria della lingua di arrivo. In italiano esistono infatti due termini distinti nella terminologia architettonica: i "pilastri" sono a sezione quadrata o poligonale, mentre le "colonne" sono a sezione circolare (Zotti & Gramellini, 2025).

Per correggere gli errori generati da questi sistemi artificiali intelligenti nella traduzione del linguaggio specialistico dei beni culturali (Zotti & Turci, 2022), è stata avviata la costituzione di un sotto-corpus comparabile italiano di letteratura di viaggio che sarà consultabile tramite l'applicazione. Questa risorsa, originariamente non prevista nel progetto e attualmente in corso di realizzazione, rappresenta un'implementazione significativa volta al miglioramento delle prestazioni delle tecnologie di traduzione automatica. Il sotto-corpus sarà dapprima impiegato come riferimento per la revisione e la correzione manuale degli *output* inaccurati generati dai sistemi di intelligenza artificiale (*post-editing*). Successivamente, verrà utilizzato per l'apprendimento supervisionato di questi strumenti attraverso una fase di *fine-tuning* finalizzata a ottimizzare la qualità traduttiva in relazione all'ambito di specialità del progetto. Il sotto-corpus italiano faciliterà inoltre la conservazione digitale di un patrimonio letterario attualmente poco noto, data la scarsità di resoconti di viaggiatori e viaggiatrici italiane in Emilia-Romagna digitalizzati e consultabili su biblioteche e archivi online. In questo modo, l'attuale implementazione del progetto di ricerca consentirà a un pubblico sempre più ampio di conoscere il patrimonio artistico e culturale bolognese.

BIBLIOGRAFIA

- Andler, D. (2023). *Intelligence artificielle, intelligence humaine : la double énigme*, Paris: Gallimard.
- Apostolakis, A. (2003). The convergence process in heritage tourism. *Annals of Tourism Research*, 30(4), 795-821. [https://doi.org/10.1016/S0160-7383\(03\)00057-4](https://doi.org/10.1016/S0160-7383(03)00057-4).
- Barbuti, N. (2018). From Digital Cultural Heritage to Digital Culture: Evolution in Digital Humanities. DTUC '18: Proceedings of the 1st International Conference on Digital Tools & Uses Congress, Paris 3-5 Ottobre 2018. Reyes, E., Szoniecky, S., Mkadmi, A., Kembellec, G., Fournier-S'niehotta, R., Siala-Kallel, F., Ammi, M. & Labelle, S. (Eds.). 1-3. <https://doi.org/10.1145/3240117.3240142>.
- Brilli, A. (2005). *Bologna nei taccuini e negli scritti dei viaggiatori stranieri dell'Ottocento*. Banca Etruria. ISBN: 201-81-0100-168-8.
- Brilli, A. (2014). *Viaggiatori stranieri fra Romagna e Marche: 19.-20. Secolo*. Bologna: Minerva Edizioni. ISBN: 978-88-7381-717-8.
- Ciotti, F. (2023). Minerva e il pappagallo. IA generativa e modelli linguistici nel laboratorio dell'umanista digitale. *Testo e Senso*, 26, 289-315. <https://doi.org/10.58015/2036-2293/671>
- Cusatelli, G. (1986). *Viaggi e viaggiatori del Settecento in Emilia e in Romagna*. Bologna: Il mulino.
- Donati, D. (2019). La digitalizzazione del patrimonio culturale. Caratteri strutturali e valore dei beni, tra disciplina amministrativa e tutela opere d'ingegno. *P. A. Persona e Amministrazione*, 2, 323-337. <https://doi.org/10.14276/2610-9050.2093>.
- Farina, A. & Martínez, M. C. N. (2020). La banca dati LBC. In Billero, R., Farina, A. & Martínez, M. C. N. (Eds.), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni culturali*. Firenze: Firenze University Press. 7-17. <https://doi.org/10.36253/978-88-5518-253-9>.
- Farina, A. (2016). Le portail lexicographique du Lessico plurilingue dei Beni Culturali, outil pour le professionnel, instrument de divulgation du savoir patrimonial et atelier didactique. *Publif@rum*, 25, <https://riviste.unige.it/index.php/publiforum/article/view/1811>.
- Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H. & Awadalla HH. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *arXiv preprint arXiv:2302.09210*.
- Jiao, W., Wang, W., Huang, JT., Wang, X. & Tu, Z. (2023). Is ChatGPT a good translator? A preliminary study. *arXiv preprint arXiv:2301.08745*.
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108608480>.
- L'Homme, MC. (2004). *La terminologie: principes et techniques*. Presses de l'Université de Montréal. <https://doi.org/10.4000/books.pum.10693>.
- Muhr, R. (1991). ATLAS/ti — A prototype for the support of text interpretation. *Qualitative Sociology*, 14(4), 349-371.
- Münster, S., Apollonio, F. I., Bell, P., Kuroczynski, I., Di Lenardo, I., Rinaudo, F. & Tamborrino, R. (2019). Digital Cultural Heritage meets Digital Humanities. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2-W15, 813-820. <https://doi.org/10.5194/isprs-archives-XLII-2-W15-813-2019>.

- Pérez-Ortiz, JA., Forcada, ML. & Sánchez-Martínez, F. (2022). How neural machine translation works. In Dorothy Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, 141-164. <https://doi.org/10.5281/zenodo.6653406>.
- Perotto, M. & Zotti, V. (2022). Il patrimonio culturale e artistico di Bologna nel corpus LBC (Lexikon of Cultural Heritage) e nel corpus UniCittà (UniCity). In Conterno, C. & Sabba, F. (Eds.), *Il patrimonio culturale della Biblioteca Universitaria di Bologna e della città allo specchio dei viaggiatori europei. Esplorazioni tra la prima modernità e l'era contemporanea*, Bologna: Bologna University Press, 161-165.
- Roque, M. I. (2022). Storytelling in Cultural Heritage: Tourism and Community Engagement. In Campos A. C. & Almeida, S. (Eds.), *Global Perspectives on Strategic Storytelling in Destination Marketing*. IGI Global. 22-37. ISBN: 978-16-6843-436-9.
- Yvon, F. (2023). La traduction multilingue : analyse d'une prouesse technologique. *MediAzioni*, 39. A17-A34. <https://doi.org/10.6092/issn.1974-4382/18785>.
- Zotti, V. & Gramellini, R. (2025). Intelligenza umana e intelligenza artificiale per la traduzione del lessico artistico: valutazioni preliminari sul confronto tra DeepL e ChatGPT 4.0 . Atti della XXa Giornata Scientifica REALITER e della XXIIIa Giornata AETER, San Millán de la Cogolla: Cilengua. In fase di revisione.
- Zotti, V. & Turci, M. (Eds.). (2022). *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali*. Firenze: Firenze University Press. <https://doi.org/10.36253/979-12-215-0061-5>.
- Zotti, V. (2023). Didactique de la méthodologie de corpus et applications pratiques dans le tourisme patrimonial et en lexicographie bilingue spécialisée : le projet UniVOCittà. Actes des 11èmes Journées Internationales de la Linguistique de Corpus, Université de Grenoble Alpes, Lidilem. 384-389. https://jlc2023.sciencesconf.org/data/pages/abstracts_JLC_2023.pdf
- Zotti, V. (2025). Lessico artistico plurilingue e valorizzazione del patrimonio culturale italiano: i progetti UNICittà e UniVOCittà. XXXIII Convegno dell'Associazione Italiana per la Terminologia (Ass.I.Term), Napoli 7 novembre 2023, Peter Lang, in stampa.
- Zotti, V. (2023). Traduire en français le lexique du patrimoine artistique de la ville de Bologne : le sous-corpus comparable BER du projet LBC. In Zotti, V. & Turci, M. (Eds.), *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali*. Firenze: Firenze University Press. 197-229. <https://doi.org/10.36253/979-12-215-0061-5.14>.