

Voci dall’Inferno: Dante per esprimere l’indicibile. Un’applicazione digitale per esplorare le testimonianze non letterarie dei sopravvissuti ai Lager

Elvira Mercatanti

Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italia
elvira.mercatanti@ilc.cnr.it

Angelo Mario Del Grosso

Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italia
angelo.delgrosso@ilc.cnr.it

Marina Riccucci

Università di Pisa, Pisa, Italia
marina.riccucci@unipi.it

Abstract

Il contributo presenta gli sviluppi del progetto *Voci dall’Inferno*, che si articola attorno a due obiettivi principali, tra loro integrati: a) la digitalizzazione e la codifica del primo corpus digitale di testimonianze *non letterarie* di sopravvissuti ai Lager; b) l’individuazione, la quantificazione e la valutazione della presenza di lessico e di immagini dantesche all’interno di quelle testimonianze. Da un lato si illustra la creazione del corpus digitale finalizzato alla raccolta, alla conservazione e alla tutela delle testimonianze *non letterarie*; dall’altro, si documenta l’impiego di strumenti e metodi informatici per l’analisi dei materiali digitalizzati. Tra i risultati più significativi del progetto si segnala l’applicazione web *Voci dall’Inferno*, pensata per catalogare, consultare e analizzare le testimonianze dell’archivio, nonché per censire e studiare sistematicamente la presenza del lessico dantesco all’interno del corpus. Il contributo si articola in tre sezioni. La prima analizza, da una prospettiva storico-letteraria e umanistica, le ragioni profonde per le quali molti sopravvissuti hanno scelto Dante per dire il Lager. Le due sezioni successive descrivono gli aspetti tecnici del progetto: la seconda è dedicata alla rappresentazione digitale XML/TEI delle testimonianze; la terza illustra lo stato attuale dell’applicazione web *Voci dall’Inferno*.

527

Copyright © 2025 The Author(s)

The text in this work is licensed under the Creative Commons BY License.

<https://creativecommons.org/licenses/by/4.0/>

Il contributo integra i risultati della tesi magistrale in Informatica Umanistica di Elvira Mercatanti [13].

Parole chiave: #AIUCD2024, XML-TEI, Testimonianze Olocausto, Archivi Digitali, XQuery, eXist-db, Applicazione Web, Dante, Divina Commedia, Lessico dantesco

This paper presents the ongoing developments of the project Voci dall'Inferno, which is structured around two integrated objectives: (a) the digitization and encoding of the first digital corpus of non-literary testimonies by Lager survivors; (b) the identification, quantification, and analysis of Dante's lexicon and imagery within those testimonies. On the one hand, the creation of the digital corpus aims to collect, preserve, and enhance of non-literary testimonies; on the other, the project documents the use of computational tools and methods for analyzing the digitized materials. Among the project's most significant outcomes is the web application Voci dall'Inferno, designed to catalogue, browse, and analyze the archived testimonies, as well as to systematically detect and study the presence of Dante's vocabulary in the corpus. The contribution is structured in three sections. The first investigates, from a historical-literary and humanistic perspective, the reasons why Dante became a privileged reference for many survivors to express the experience of the Lager. The second and third sections address the project's technical components: the second focuses on the XML/TEI encoding of the testimonies, while the third outlines the current state of development of the Voci dall'Inferno web platform. The contribution also incorporates results from Elvira Mercatanti's master's thesis in Digital Humanities [13].

Keywords: #AIUCD2024, XML-TEI, Holocaust Testimonies, Digital Archives, XQuery, eXist-db, Web Application, Dante's lexicon, Dante's allusions

1. Dante per dire il Lager

Si stima che a perdere la vita negli oltre 60.000 Lager nazisti furono circa 20 milioni di persone. [10] Di coloro che subirono la deportazione nei Lager, solo il dieci per cento è sopravvissuto ed è tornato a casa. In pochi hanno scelto di raccontare la propria esperienza e soltanto la minoranza ha iniziato a farlo subito. Per gli altri sono serviti decenni di silenzio. Molti non hanno mai raccontato. Perché il Lager è ineffabile.

Sento talora l'insufficienza dello strumento. Ineffabilità, si chiama, ed è una bellissima parola. Il nostro linguaggio è umano, è nato per descrivere cose a dimensioni umane, non cose disumane come Auschwitz. [12]

Se ricordare e riferire ciò che è stato vissuto è già di per sé un peso insostenibile, una volta trovata la forza per farlo, subentra il problema di trovare le parole per poter dire il Lager. Occorre dunque superare dapprima la barriera del silenzio, scegliendo di rivivere e di riattraversare il dolore ricordando e riferendo, e, successivamente, l'ostacolo della povertà di un vocabolario che non possiede dei termini attraverso i quali poter esprimere l'ineffabile.

Le tipologie testuali attraverso le quali in questi ottant'anni, il Lager e l'esperienza concentrazionaria sono stati descritti, riferiti e testimoniati, sono state individuate da Riccucci e sono essenzialmente due:

1. quella della testimonianza diretta – coeva e non – di chi ha vissuto il campo di sterminio e ne ha riferito in forme che solo di rado sono in tangenza con la letterarietà: il *modus dicendi* di questa fattispecie (che chiamerò di 'primo livello') si colloca nello spazio compreso tra il resoconto orale (l'intervista) e quello scritto (il diario, il racconto autobiografico / memoriale, la lettera);

2. quella della testimonianza diretta – coeva e non – di chi ha vissuto il campo di sterminio e ha scelto, per riferirne, la forma più spesso della prosa, meno della lirica, in ogni caso della narrativa (quindi della letteratura), volendo cioè che il proprio resoconto si presentasse sotto forma di racconto organizzato, tematicamente e stilisticamente strutturato [...]. [6]

Per definizione degli stessi superstiti il Lager è *indicibile*: quindi portarne testimonianza equivale a narrare l'inesprimibile. Coloro che hanno voluto provare a raccontare l'ineffabile sono stati colpiti da una crisi di impotenza in quanto hanno riscontrato un limite linguistico dovuto alla mancanza di parole per esprimere il Lager. Ma attraversando il corpus della letteratura concentrazionaria, seppure i modi, le forme e i toni siano differenziati, il minimo comun denominatore è la definizione universale e condivisa del campo di concentramento come di un "inferno". Più nello specifico, l'immagine del Lager che ne emerge è quella di un inferno sulla terra e in molti casi è proprio la prima cantica dantesca a costituirne il modello principale. [24] Il regno sotterraneo descritto da Dante, creato dalla sua immaginazione poetica, diventa qualcosa di così tangibile per i prigionieri dei campi di concentramento che, nell'immaginario collettivo, l'immagine letteraria dell'*Inferno* arriva a sovrapporsi completamente alla percezione reale di questa metafora infera. Le testimonianze dimostrano come i prigionieri, di fronte alla brutalità e all'assurdità dei campi, abbiano utilizzato le terzine di Dante per rappresentare a sé stessi la vita nel Lager, cercando di comprendere la loro condizione attraverso la metafora dell'*Inferno*.

Come riporta Valeria Traversi:

Il vocabolario e l'immaginario danteschi devono aver agito nelle menti dei deportati appena arrivati nel Lager come filtro per provare a decodificare una realtà per la quale non esisteva nessun altro strumento di conoscenza. [26]

La *Commedia* ha permesso di superare l'indicibilità e l'incomunicabilità dell'esperienza concentrazionaria, diventando un supporto cognitivo per agevolare la comprensione da parte di coloro che non hanno vissuto il Lager. Dante ha fornito quindi ai testimoni il vocabolario per sopperire alle mancanze di una lingua che non aveva lemmi per dare un nome a determinate cose e a determinati eventi, permettendo allo stesso tempo ai lettori o agli uditori di riuscire, seppur in parte, a comprendere quella realtà. La *Commedia* costituisce dapprima un ponte tra i sopravvissuti stessi e l'esperienza concentrazionaria e, in seguito, tra il loro racconto e gli uditori. È, dunque, «strumento di decodificazione per gli internati, medium di dicibilità della testimonianza e grammatica condivisa con il lettore». [24]

Per tutti i sopravvissuti il campo di concentramento è apparso come la realizzazione concreta di una realtà di cui Dante ci ha offerto il miglior esempio. Nonostante ciò, la differenza principale tra l'*Inferno* dantesco e il Lager nazista è che il primo è quanto di più moralmente distante si possa trovare da un Lager, in quanto in quest'ultimo la giustizia divina che caratterizza il poema dantesco è letteralmente capovolta. Come riportato dallo scrittore tedesco Martin Walser:

Paragonare Auschwitz all'*Inferno* dantesco è quasi un'insolenza, qualora non sia possibile accampare come attenuante l'ignoranza. [...] Nell'*Inferno*, in fin dei conti, i peccati venivano espiati dai colpevoli. All'*Inferno* seguivano pur sempre anche il Purgatorio e il Paradiso. Si sarebbe chiesto orrendamente troppo agli esseri umani ad Auschwitz, nel caso avessero dovuto declamare, a un Dante che passava di lì, per quali peccati venivano tormentati. E al tormento seguiva soltanto lo sterminio.[27].

Mentre l'*Inferno* di Dante è il luogo dei peccati commessi dai dannati e dunque luogo della giustizia divina, l'inferno del Lager è quello dell'ingiustizia assoluta in cui i campi non obbediscono in nessun modo alla relazione colpa-peccato. Nel Lager furono delle vittime innocenti a essere punite dai loro aguzzini. A rimarcare questa differenza fondamentale, Thomas Taterka scrive:

Laddove Dante sottolinea con la legge del contrappasso la commisurazione tra colpa e pena, nella scrittura dei testimoni si insiste sull'arbitrarietà di chi si è fatto giudice dei propri pari, fuori da ogni legge umana e divina, e ha distribuito pene gratuitamente senza fondamento giuridico e morale. [25]

Nonostante ciò, sono i testimoni stessi ad aggrapparsi al binomio Lager-Inferno attraverso riferimenti vaghi al regno infernale o rimandi più precisi al poema dantesco, nel tentativo di comprendere l'essenza del campo di concentramento. Come riportò Hanna Lévy-Hass nel diario che tenne durante il suo periodo di internamento a Bergen-Belsen:

Più di una volta [...] di fronte agli estremi tormenti di massa, mi sono vista davanti l'*Inferno* dantesco. Ma non per dilettermi con reminiscenze letterarie. Perché le rappresentazioni dell'*Inferno*, alle quali l'immaginazione è avvezza, erano l'unica impressione che il mio cervello sapesse ancora elencare. Non ero in grado di risvegliare un qualsiasi altro ricordo; era l'unica idea ancora viva nella mia mente. [25]

Questa connessione tra l'*Inferno* dantesco e il Lager nazista ha portato a una sovrapposizione semantica dei due termini, "inferno" e "Lager", divenuti ormai quasi sinonimi. Si è creata una simmetria tra il luogo immaginario di Dante e quello reale dei campi, entrambi trasmessi attraverso la scrittura. L'*Inferno* ha perso la sua connotazione teologica e il Lager la sua dimensione strettamente storica, fondendosi in un legame che ne ha rafforzato l'efficacia comunicativa ed evocativa.

1.1 Il progetto *Voci dall'Inferno*

L'opera dantesca ha sicuramente svolto un ruolo decisivo nelle testimonianze letterarie, ampiamente studiate, ma ciò su cui si è concentrato il primo studio che ha posto le basi per lo sviluppo del progetto *Voci dall'Inferno* sono le testimonianze di primo livello, cioè quelle che noi chiamiamo *non letterarie*. È in questi testi che il ricorso a Dante si configura come un atto maggiormente istintivo. [6]

Il progetto nasce in occasione del lavoro di tesi della dottoressa Sara Calderini (2006), la quale ha provato a rispondere alla domanda di base 'se' e, se sì, 'quanto e in che misura', Dante abbia fornito le parole per dire il Lager anche alle testimonianze definite come non letterarie. [5] La risposta alla domanda di base si è rivelata essere un'affermazione, ma non solo: dalle indagini che Marina Riccucci ha continuato a svolgere anche negli anni successivi, è emerso che la presenza di lemmi danteschi caratterizza le testimonianze indipendentemente non solo dall'età, dal genere e dalla nazionalità, ma anche dal livello culturale e di istruzione. Dante rompe il silenzio anche di chi non lo ha mai studiato a scuola, in quanto è diventato un patrimonio lemmatico collettivo fatto di parole entrate nell'uso comune, nella lingua del quotidiano.

È su questa linea che si colloca il progetto *Voci dall'Inferno*. L'informatica ha potuto offrire un supporto concreto, con l'obiettivo innanzitutto di costituire il primo corpus digitale di testimonianze non letterarie di sopravvissuti ai Lager e successivamente di poter analizzare,

interrogare e studiare il materiale digitalizzato in merito alla presenza di lessico e immagini dantesche, ampliandone ulteriormente i confini d'indagine.

Come verificato da Federico Siragusa, [23] la maggior parte delle citazioni e dei riferimenti contenuti all'interno dei diari e delle opere esaminate, sono riconducibili alla prima cantica. Sebbene i luoghi danteschi dell'*Inferno* siano quelli che maggiormente vengono rievocati dalle parole degli ex deportati, si è osservato che anche il *Purgatorio* e il *Paradiso* sono stati citati dai testimoni, seppur in misura minore. Concentrando l'attenzione sulla prima cantica, tra le scene che ricorrono maggiormente nella memoria dei sopravvissuti c'è il momento dell'arrivo di Dante di fronte alla porta del regno infernale.

Ne sono un esempio le parole presenti nel diario del frate cappuccino Giannantonio Agosti da Romallo, internato prima a Flossenbürg e poi a Dachau, dove venne liberato dai soldati americani il 29 aprile 1945. Dopo il rimpatrio avvenuto nel maggio del 1945, Padre Giannantonio scrisse subito le sue memorie, pubblicate poi nel 1968, un anno dopo la sua morte, con il titolo *Nei Lager vinse la bontà. Memorie dell'Internamento nei campi di eliminazione tedeschi*. [1]

Nel ricordare il suo arrivo di fronte alla porta di ingresso del Lager, il cappuccino scrive:

Dopo un'ampia svolta si giunge di fronte al campo. V'è una scritta sull'arco della porta: «Arbeiterlager» «Campo dei lavoratori». Dopo l'esperienza dei primi giorni vi si poteva invece scrivere, e ben a ragione, la nota terzina di Dante: Per me si va nella città dolente, Per me si va nell'eterno dolore, Per me si va tra la perduta gente.

Chiaro è il riferimento all'epigrafe che Dante legge sopra la porta dell'*Inferno* in *Inferno* III:

Per me si va ne la città dolente,
per me si va ne l'eterno dolore,
per me si va tra la perduta gente.
Giustizia mosse il mio alto fattore;
fecemi la divina podestate,
la somma sapienza e 'l primo amore.
Dinanzi a me non fuor cose create
se non eterne, e io eterno duro.
Lasciate ogne speranza, voi ch'intrate.
(*Inferno* III, vv. 1-9) [2]

Riferendoci sempre a *Inferno* III, c'è un'altra immagine a cui la memoria dei testimoni si appella per cercare di descrivere più verosimilmente possibile il momento della loro deportazione: il traghettatore infernale Caronte.

Caron dimonio, con occhi di bragia
loro accennando, tutte le raccoglie;
batte col remo qualunque s'adagia.
(*Inferno* III, vv. 109-111)

Ancora Padre Giannantonio cita perfettamente il primo e il terzo verso di questa terzina all'interno del suo diario per descrivere l'accoglienza nel campo di concentramento di Flossenbürg:

Fummo accolti con carezze di bastoni. Otto o dieci incaricati, all'ingresso, o per sollecitarci o per metterci a posto o per nessun motivo, davano a

casaccio botte da orbi, che sulla nuda pelle si sentivano di più e lasciavano il segno. Come non ricordare la terzina dantesca: «Caron dimonio, con occhi di bragia... batte col remo qualunque s'adagia». [1]

Durante il suo cammino all'interno della selva, Dante si imbatte nelle tre fiere e, soprattutto a causa della lupa che lo respinge sempre più nell'oscurità, il poeta rinuncia a proseguire il proprio itinerario fin quando una figura indefinita gli appare davanti agli occhi:

«Miserere di me», gridai a lui,
«qual che tu sii, od ombra od omo certo!»
(Inferno I, vv. 65-66)

Questi versi annunciano l'ingresso nella *Commedia* del poeta latino Virgilio, ripresi da Luigi Giuntini all'interno del suo diario, seppur con una differenza:

«Miserere di me» gridai a lui «qual tu sii, od ombra od omo certo?». [8]

Come messo in luce da Irene Petrilli, [15] mentre in *If. I 66* troviamo un punto esclamativo alla fine del verso, Luigi Giuntini utilizza un punto interrogativo nello stesso luogo, trasformando così l'esclamazione dantesca in una domanda.

Un altro riferimento a Virgilio è rintracciabile all'interno dell'intervista del 10 marzo 2017 realizzata da Marina Riccucci a Liliana Segre (che qualche mese dopo sarebbe stata nominata senatrice a vita) nella quale, facendo riferimento a Goti Bauer, viene affermato:

Perché, per esempio, io la Goti Bauer l'ho conosciuta trent'anni fa, per combinazione [...]. [Goti] è una donna eccezionale: vi dico solo che per me è il mio Virgilio. L'ho sempre detto che lei per me era il mio Virgilio, perché lei mi ha preso per mano e mi ha portato a diventare testimone. [...] E un giorno, due, tre, a riparlarmi, a spingermi a fare, io le ho detto, ma te sei proprio il mio Virgilio [...]; sempre le ho detto te sei il mio Virgilio, perché io sono veramente in una selva oscura, come faccio a uscirne? E combinazione vuole che ne sono uscita, [...]. E lei mi ha aiutato moltissimo, la Goti. [19]

1.2 Costituzione dell'archivio digitale

Il progetto *Voci dall'Inferno* ha attraversato nella sua evoluzione tre fasi principali:

1. lo sviluppo di una banca dati per la gestione dell'anagrafica delle testimonianze: l'archivio ha preso il nome di *Memoriarchivio*; [18]
2. la creazione del corpus delle testimonianze in formato XML-TEI;
3. l'implementazione di un'applicazione web per la fruizione e l'interrogazione dei dati conservati nell'archivio digitale.

Il database *Memoriarchivio*, sviluppato dalla professoressa Frida Valecchi, ha consentito di creare un primo inventario delle testimonianze, conservando descrizioni catalografiche e letterarie nonché, qualora presenti, le trascrizioni della testimonianza.

In aggiunta, l'ambiente web permetteva anche il confronto del lessico della testimonianza con il testo della *Commedia*, l'aggiornamento dell'inventario, delle anagrafiche dei testimoni e dei curatori delle fonti, e, successivamente, anche la gestione di documenti in formato XML-TEI. Nel corso degli anni il corpus delle testimonianze codificate si è via via arricchito grazie al

contributo dei laureandi del Corso di Laurea in Informatica Umanistica dell'Università di Pisa, i quali si sono occupati della trascrizione e della marcatura seguendo le indicazioni fornite dalle linee guida del consorzio TEI.¹ Il Capitolo 2 di questo lavoro illustrerà nel dettaglio le scelte adottate per la rappresentazione digitale del corpus. Verranno inoltre descritte la struttura generale dei documenti TEI e gli elementi selezionati per la codifica dei principali fenomeni testuali.

L'archivio digitale delle testimonianze sarebbe meno efficace dal punto di vista funzionale e scientifico senza la presenza di un componente software dedicato all'estrazione, alla presentazione e alla fruizione dei dati prodotti durante la fase di codifica. Nel corso del progetto sono state sperimentate due differenti strategie di restituzione dei dati, facendo uso di due diversi approcci architetturali:

1. per il primo approccio, le applicazioni web prototipali sono state sviluppate facendo leva sulle funzionalità di una libreria client-side per l'elaborazione di documenti XML: SaxonJS;²
2. per il secondo, le applicazioni web prototipali sono state sviluppate adottando le tecnologie messe a disposizione dall'ambiente eXist-db mediante l'uso del modulo *HTML-templating framework* (server-side). eXist-db integra un modulo dedicato alla generazione dinamica di pagine HTML, partendo da collezioni di documenti XML e da procedure implementate tramite funzioni XQuery. Il funzionamento prevede l'uso di template in HTML, in cui si aggiungono chiamate a funzioni XQuery che permettono l'elaborazione dell'informazione conservata in collezioni di documenti XML e la generazione di frammenti HTML che completano la pagina web finale.

Fino ad oggi sono state implementate e proposte diverse versioni di applicazione web da parte di studenti dell'Università di Pisa nell'ambito dei rispettivi progetti di tesi. L'applicazione *Voci dall'Inferno*,³ nella sua versione attuale, consente la consultazione, l'analisi e la fruizione di tutte le tipologie di testimonianze che costituiscono il corpus. Le principali fasi di sviluppo dell'applet basata su eXist-db sono approfondite nel corso del capitolo 3, che presenta sia il modello architetturale sia alcuni esempi di implementazione. Inoltre, il capitolo illustra parte dei risultati generati e restituiti all'utente finale.

1.3 Il repertorio delle testimonianze del corpus

L'archivio digitale *Voci dall'Inferno* è attualmente composto da 23 testimonianze rilasciate da 18 sopravvissuti. Tutte le testimonianze sono state codificate secondo lo schema definito nel corso dello sviluppo del progetto, descritto in un documento ODD (One Document Does it All, come illustrato nel Capitolo 2). I testimoni provenienti da contesti diversi presentano esperienze e percorsi differenti e hanno lasciato le loro testimonianze attraverso modalità eterogenee. Questo ha contribuito a rendere il corpus particolarmente variegato, offrendo al contempo una visione ampia e articolata del Lager e del periodo del Nazifascismo. La maggior parte delle testimonianze è inedita, ma tutte sono estremamente preziose in quanto costituiscono un patrimonio storico di immenso valore che necessita di essere tutelato e studiato.

¹ Cfr. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

² Cfr. <https://www.saxonica.com/html/welcome/welcome.html>

³ Una demo dell'applicazione *Voci dall'Inferno* è raggiungibile al link https://48001.omega.ilc.cnr.it/exist/apps/voci_inferno/index.html

Alcune delle testimonianze che costituiscono l'archivio sono state anche delle vere e proprie scoperte. È il caso, per esempio, della storia di Bruno Cimoli. Tra il 1944 e il 1945 Cimoli fu tra i prigionieri che ogni giorno venivano portati da Wriezen all'atelier di Jäckelsbruch per servire e aiutare gli scultori sotto la guida di Arno Breker, scultore ufficiale del Terzo Reich, noto soprattutto per il contributo che portò al progetto della Nuova Berlino ideata da Hitler. Nella sua testimonianza Cimoli racconta di aver visto nell'atelier un'opera molto particolare a cui gli scultori stavano lavorando: una piramide dalle dimensioni gigantesche che aveva su ogni lato una figura maschile scolpita secondo i parametri di bellezza del regime e del gusto artistico di Breker. La piramide doveva essere collocata nella Nuova Berlino, ma ciò non avvenne mai perché fu distrutta dopo l'invasione dell'URSS nel 1945. Bruno Cimoli è stato uno dei testimoni del processo di creazione di molte delle opere architettoniche e scultoree più rappresentative dell'arte tedesca dell'epoca. Dopo l'invasione dell'URSS iniziata nel febbraio del 1945, venne cancellato qualsiasi tipo di modello, schizzo o testimonianza della piramide. Rimangono solo la testimonianza di Cimoli e un dépliant andato perduto (dell'esistenza del quale è Cimoli a dare notizia nella lettera del 2006). [14]

Interessante è anche il caso di Idek Wolfowicz, considerato un vero e proprio “testimone ritrovato”. Negli anni 2005-2006 le dottoresse Anna Segre e Gloria Pavoncello hanno raccolto le testimonianze di alcuni sopravvissuti che vivevano in Italia: questo lavoro esitò nel volume *Judenrampe*. [21] Nel 2006 Segre e Pavoncello intervistarono, presso la sua abitazione a Roma, anche Idek Wolfowicz, con il quale entrarono in contatto per puro caso, grazie alla figlia di Idek, Miriam Wolfowicz, poiché quest'ultima conosceva personalmente la dottoressa Segre. Idek rese testimonianza a Segre e Pavoncello, le quali erano ignare del fatto che all'interno della banca dati del Centro di Documentazione Ebraica Contemporanea di Milano non c'era nessuna traccia di Idek Wolfowicz. L'archivio non disponeva di alcuna informazione sull'uomo. L'intervista inedita è stata affidata da Riccucci alla dottoressa Diletta Lelli per il suo lavoro di tesi [11].

Degna di menzione è anche la storia del diario di prigionia di Nicola Ricci, un internato militare italiano. Ricci fu catturato pochi chilometri dopo il confine francoitaliano dall'esercito nazista e fu inizialmente deportato come Internato Militare Italiano (I.M.I.) in un campo di smistamento per prigionieri nel sud della Germania e, successivamente, ad Armestein, un campo di concentramento vicino al confine con la Cecoslovacchia. Ricci raccontò i fatti vissuti dal giorno dell'armistizio, l'8 settembre 1943, al giorno della liberazione, il 14 marzo 1945, scrivendoli su fogli volanti sotto forma di appunti. Assemblerà tutto il diario in modo più strutturato e cronologicamente coerente negli anni successivi alla liberazione, nel suo ufficio alla questura di Ancona, con una macchina da scrivere che è ancora proprietà della famiglia Ricci. Nicola non fece mai parola con nessuno di questo suo scritto, che fu rinvenuto in un cassetto della sua abitazione pochi giorni dopo la sua morte. [17]

Fanno parte del corpus *Voci dall'Inferno* anche le memorie di una persona che ha offerto un punto di vista diverso sulla persecuzione nazifascista: Emanuele Artom, partigiano torinese catturato e torturato dai nazisti perché riconosciuto come commissario politico delle bande azioniste in Val Pellice e Val Germanasca, nonché ebreo. Nonostante Artom non sia stato deportato, il suo contributo è fondamentale per comprendere la varietà e la complessità delle esperienze di coloro che hanno subito la persecuzione nazifascista. Il suo Inferno non è rappresentato dal Lager, ma è costituito dai provvedimenti antisemiti, dalla guerra, dal terrore di essere catturato dai fascisti, dalla paura della deportazione. La sua storia arricchisce il progetto, offrendo una prospettiva essenziale che permette di documentare in maniera più completa le molteplici forme di oppressione vissute dagli ebrei durante la Seconda Guerra Mondiale. Nelle pagine del suo *Diario* Artom fa uso di citazioni provenienti dall'*Inferno* dantesco per descrivere la quotidianità della vita partigiana. Dante offre quindi a chiunque le parole per descrivere il proprio Inferno. [23]

Infine, arricchiscono l'archivio anche le testimonianze di due prigionieri che furono obbligati a collaborare con le SS⁴ svolgendo il ruolo di *Sonderkommando*,⁵ i gruppi speciali di prigionieri che erano incaricati di occuparsi delle operazioni di rimozione dei corpi dalle camere a gas e di quelle successive di cremazione. È il caso di Enrico Vanzini e di Shlomo Venezia. Vanzini, attualmente l'ultimo *Sonderkommando* italiano ancora in vita, fu un internato militare italiano che svolse tale ruolo nel campo di concentramento di Dachau. Venezia, deportato ebreo, fece lo stesso nel campo di concentramento di Auschwitz-Birkenau. Le loro testimonianze sono particolarmente preziose in quanto di tali squadre, in pochissimi sono sopravvissuti: i *Sonderkommando*, infatti, restavano in vita solamente per tre mesi, dopodiché venivano eliminati e sostituiti da nuovi prigionieri per impedire la circolazione di notizie su ciò che avveniva nelle camere a gas e nei forni crematori. Vanzini e Venezia sono gli unici sopravvissuti e memorialisti italiani di tali squadre.

1.4 Definizione di una tassonomia

L'archivio dispone quindi sia del racconto di chi visse l'esperienza del Lager sia di chi non fu mai deportato. Tra coloro che hanno conosciuto il campo di concentramento, sono presenti testimonianze rese da deportati ebrei e da internati militari italiani. Infine, di chi non visse il Lager, è presente la testimonianza del partigiano ebreo Emanuele Artom. È stato quindi fondamentale individuare e definire una prima ipotesi di tassonomia del progetto *Voci dall'Inferno* (Figura 1), a oggi non ancora del tutto esaustiva e quindi ancora in via di definizione, che consentisse di identificare e di distinguere le testimonianze rese dai sopravvissuti. In secondo luogo tale operazione è stata necessaria al fine di gestire e organizzare i dati estremamente disomogenei di cui si dispone all'interno dell'applicazione. Il capitolo 3 presenta al lettore maggiori informazioni rispetto alla tassonomia implementata in *Voci dall'Inferno*.

I 18 testimoni dell'archivio digitale, dei quali sono state codificate le testimonianze, sono così ripartiti: 17 testimoni vissero l'esperienza del campo di concentramento mentre, attualmente, un solo testimone non fu deportato. Di coloro che videro il Lager, 12 sono deportati ebrei, 5 sono internati militari italiani. Della prima categoria, sono presenti le testimonianze di Edith Bruck, Romana Feld, Nedo Fiano, Goti Herskovitz Bauer, Ida Marcheria, Samuel Modiano, Liliana Segre, Alessandro Smulevich, Piero Terracina, Shlomo Venezia, Arminio Wachsberger, Idek Wolfowicz. Della categoria degli I.M.I. fanno parte Bruno Cimoli, Luigi Giuntini, Alberto Pacini, Nicola Ricci, Enrico Vanzini. L'unico testimone che non fu deportato e che è attualmente presente nell'archivio è il partigiano ebreo Emanuele Artom.

⁴ Cfr. <https://www.treccani.it/enciclopedia/ss/>.

⁵ Cfr. <https://www.treccani.it/vocabolario/sonderkommando/>.

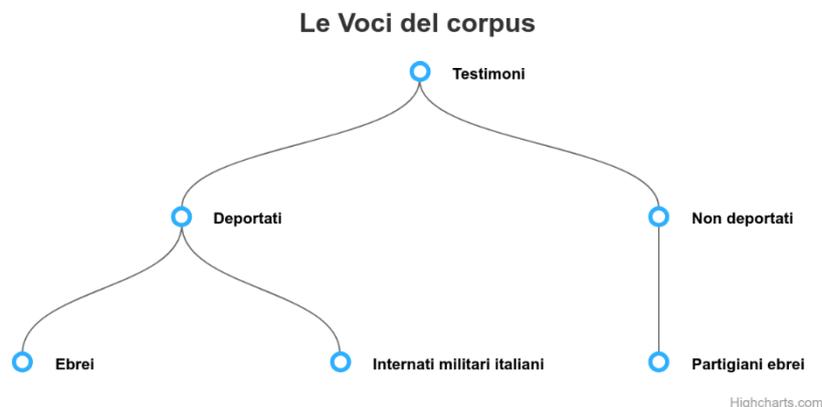


Figura 1: Tassonomia di Voci dall’Inferno

1.5 I numeri di Voci dall’Inferno

Le testimonianze che attualmente costituiscono l’archivio digitale si dividono in due macro-classi: da un lato le testimonianze orali e dall’altro le testimonianze scritte. All’interno di ciascuna macro-classe sono individuabili ulteriori tipologie in base a come si presentano le fonti primarie. Le testimonianze appartenenti alla categoria delle fonti orali si presentano in formato video e in formato audio, mentre le testimonianze appartenenti alla categoria delle fonti scritte sono costituite da diari, lettere, fogli manoscritti e fascicoli.

Il numero di testimonianze codificate che costituiscono l’archivio digitale è attualmente pari a 23. Di questo numero, 13 sono testimonianze orali, di cui 3 interviste in formato video e 10 in formato audio. La maggior parte delle fonti orali proviene dal fondo di interviste inedite realizzate dalle dottoresse Anna Segre e Gloria Pavoncello negli anni 2005-2006. [21] Le interviste, registrate su micro-cassette, sono state messe a disposizione del progetto e opportunamente convertite in formato mp3. Dei 25 sopravvissuti che hanno rilasciato interviste a Segre e Pavoncello, 7 sono stati oggetto di studio del progetto Voci dall’Inferno: Edith Bruck, Ida Marcheria, Idek Wolfowicz, Liliana Segre,⁶ Nedo Fiano, Piero Terracina e Shlomo Venezia.

Le altre due registrazioni audio che costituiscono l’archivio sono quelle di Goti Herskovits Bauer e Arminio Wachsberger.⁷ La testimonianza di Herskovitz Bauer è stata realizzata il 7 marzo 2020 da Marina Riccucci e da Laura Ricotti. [19] La testimonianza di Wachsberger appartiene alla Fondazione Centro di Documentazione Ebraica Contemporanea di Milano (CDEC)⁸ ed è stata realizzata a Milano da Liliana Picciotto nell’ambito del progetto “Ricerca sulla deportazione”. Le

⁶ Di Liliana Segre il corpus dispone di due interviste realizzate da Segre e Pavoncello: una del 2006 e una del 2007.

⁷ Di Wachsberger sono state digitalizzate due testimonianze: una del 1987 e una del 1998. L’intervista in questione è quella datata 1998.

⁸ Cfr. <https://digital-library.cdec.it/cdec-web/>.

tre testimonianze orali in formato video provengono tutte da YouTube: si tratta delle interviste di Enrico Vanzini, Samuel Modiano e Arminio Wachsberger.⁹

Delle 23 testimonianze totali, le restanti 10 sono testimonianze scritte e si compongono di 4 diari, 1 testimonianza in forma di lettera, 4 fogli manoscritti e 1 fascicolo. La parte più cospicua delle testimonianze scritte proviene da collezioni familiari private. È il caso della testimonianza di Alessandro Smulevich, della quale sono stati digitalizzati 4 fogli manoscritti, di quelle di Nicola Ricci e di Luigi Giuntini, dei quali sono stati codificati i 2 diari, e di quella di Alberto Pacini, di cui è stato trascritto e digitalizzato il corpus delle 50 lettere di prigionia. Delle rimanenti 3 testimonianze scritte, le testimonianze di Emanuele Artom e di Romana Feld appartengono all'archivio digitale del CDEC, mentre la testimonianza di Bruno Cimoli proviene dalla sede centrale dell'ANPI di Massa.¹⁰

1.6 *Dante nelle testimonianze di Voci dall'Inferno*

La presenza di tessere dantesche è stata rinvenuta anche tra le testimonianze dell'archivio *Voci dall'Inferno*, le quali si sono però presentate attraverso diverse forme: citazioni (sia in forma implicita che in forma esplicita) di versi specifici della *Commedia*, allusioni o riferimenti più generici all'opera e vocaboli danteschi. Durante la fase di codifica si sono stabilite diverse modalità di marcatura che permettessero di esplicitare e di distinguere opportunamente le differenti tipologie.

Dei 18 testimoni che costituiscono l'archivio, 11 fanno uso di citazioni, allusioni, o termini danteschi provenienti dalla *Commedia*, per un totale di 68 tessere dantesche. Di queste, 15 sono citazioni esplicite, 4 sono citazioni implicite, 10 sono allusioni e 39 sono vocaboli danteschi.

I dati che emergono dall'archivio digitale di *Voci dall'Inferno* confermano quanto osservato anche da Siragusa, ovvero che la maggior parte delle citazioni e delle allusioni presenti all'interno delle testimonianze, sono riconducibili alla prima cantica, [23] la quale presenta infatti 16 citazioni su 19 totali. Nonostante ciò, anche il *Purgatorio* e il *Paradiso* sono stati citati dai testimoni, seppur in misura minore: il primo due volte, il secondo una volta.

Tra i testimoni che ricorrono maggiormente a citazioni dantesche, c'è Luigi Giuntini, che nel suo diario [8] cita a memoria in maniera esplicita la *Commedia* per ben 12 volte. All'epoca dei fatti narrati Giuntini aveva solo la licenza elementare. Nonostante avesse interrotto gli studi, si era appassionato alla letteratura e aveva studiato da autodidatta dal 1934 fino al suo arruolamento. Tra i suoi autori preferiti c'era proprio Dante, dal quale Giuntini ha spesso preso in prestito le parole per riuscire ad esprimere ciò che stava vivendo. [15]

Uno dei canti citati da Giuntini è, ad esempio, *Inferno* I, di cui Luigi, nel passo del diario riportato di seguito, cita il verso 21, e i versi 37-40 e 65-66:

Mi soffermo con grande interesse sul primo Canto dell'Inferno perché a me pare denso di allegorie e di similitudini. [...] Dirò che, prima di essere chiamato alle armi, avevo letto gran parte del divino Poema, che il mio babbo comprava a dispense e che era illustrato con indubbio talento artistico da Gustavo Doré. Questa era la seconda o la terza volta che mi avvicinavo ai primi canti dell'Inferno e potevo confrontare le interpretazioni del testo precedenti, confrontandole con quelle attuali.

⁹ L'intervista in questione è quella datata 1987.

¹⁰ Cfr. <https://www.anpi.it/sedi?titolo=massa&area=>.

Allora penso di poter affermare che, sceverando tra i 136 versi di questo complesso proemio, almeno una decina di essi colpivano di nuovo la mia sensibile fantasia e, cioè, i seguenti: «La notte ch'io passai con tanta pietà», «Temp'era dal principio del mattino e 'l sol montava in su con quelle stelle ch'eran con lui quando l'amor divino mosse di prima quelle cose belle», «Miserere di me» gridai a lui “qual tu sii, od ombra od omo certo?”» [...]. ([8]:242)

È sempre Giuntini a citare in forma esplicita anche *Purgatorio* e *Paradiso*. Della seconda cantica, per esempio, Giuntini utilizza nel suo diario il verso 12 del canto I per descrivere l'azzurro cielo estivo privo di nuvole:

Stamani, guardando il cielo, potrei finalmente descriverlo come padre Dante: «Dolce colore d'oriental zaffiro» tanto è azzurro e senza una nuvola. ([8]:437)

Del *Paradiso*, invece, riprende il verso 141 del canto XII:

Mentre ci auguriamo che ciò non accada, intorno al nostro tavolo, nel primo pomeriggio si aduna, si fa per dire, “la bella scola” di coloro che per ragioni diverse, sono interessati a discutere di letteratura, di filosofia e di storia e, cioè, lo studente pugliese e quello umbro, Riccardi, Nicolato, Breazzano, Basso, un calabrese, che sa molte cose su Gioacchino da Fiore «di spirito profetico dotato», De Marzi, un napoletano, innamoratissimo di Salvatore Di Giacomo. ([9]:365)

Le citazioni implicite rinvenute nelle testimonianze del corpus appartengono per la maggior parte a Emanuele Artom, il quale nel suo diario fa riferimento esclusivamente alla prima cantica. Interessante, per esempio, è il rimando a *Inferno* VI 10-12:

Almeno mi pareva di essere come Ciacco stasera, pochi minuti fa, quando tornavo in albergo dopo aver camminato per ore al buio, perdendomi e disperdendomi per la campagna, sotto una pioggia continua, affondando nel fango e nella fanghiglia [...]. [3]

Per quanto riguarda le 10 allusioni, provengono tutte dalla prima cantica della *Commedia: l'Inferno*. Un esempio di allusione è presente nell'intervista che Goti Bauer ha rilasciato a Marina Riccucci il 7 marzo 2020:

Era pieno il campo e noi siamo arrivati in una bolgia, dove tutti gli altri sapevano tutto e ci informavano. Era una cosa da infarto, venire a sapere una cosa così nel giro di un minuto. [19]

Un'altra allusione si ritrova in un passo del diario di Nicola Ricci:

Se il divino poeta, appostato su una di quelle colline, ci avesse potuto vedere passare, ci avrebbe certamente scambiato per qualche stuolo di quelle anime perse, da lui collocate in qualche cerchio dell'inferno, che costantemente girano portando il fardello dei loro peccati. Stimolati da feroci mastini, visi pallidi, stracciati, emaciati, camminiamo. [17]

Interessanti sono anche le parole di Romana Feld, la quale, oltre a definire come un inferno il luogo in cui si trova, testimonia anche la difficoltà del sopravvissuto nel trovare le parole per dire il Lager:

È impossibile con le povere parole ridare quadro esatto dell'inferno vissuto.

Anche i 39 vocaboli danteschi rinvenuti nelle testimonianze rimandano alla prima cantica. Il termine che ricorre in assoluto più di tutti è proprio *inferno*, che si presenta per ben 21 volte. Gli altri vocaboli individuati sono: *dante*, che occorre 4 volte, *bolgia*, *virgilio* e *divino poeta* che occorrono 2 volte, *infernale*, *fiere*, *divina commedia*, *anime dannate*, *purgatorio*, *anime*, *cerchio dell'inferno* e *peccati* che occorrono 1 volta.

2. L'archivio digitale

Le due macro-classi in cui si ripartisce il costituendo corpus ne determinano gli aspetti rappresentazionali, funzionali e fruizionali: testimonianze scritte e testimonianze orali.

Pur mantenendo le specifiche differenze, entrambe le classi seguono le indicazioni fornite dalle linee guida del consorzio TEI. In particolare, nel corso del progetto, è stato creato - e via via sempre più raffinato - un *One Document Does it All* (ODD)¹¹ che dichiara i moduli, gli elementi, gli attributi e i possibili valori ammessi per la codifica del repertorio delle testimonianze. Nonostante le differenze formali dovute alla tipologia delle fonti, sono stati stabiliti degli obiettivi di codifica comuni presenti in entrambi i modelli. L'obiettivo principale di codifica del progetto *Voci dall'Inferno* è quello di attenersi il più possibile alla fonte primaria, rappresentandone fedelmente la struttura e il contenuto.

2.1 Codifica di fonti scritte

Le testimonianze scritte seguono uno schema di edizione *image-based* di tipo diplomatico-interpretativo con approccio *parallel-transcription*¹² alla rappresentazione del testo-documento. A tale scopo sono stati utilizzati gli elementi definiti nel modulo *Representation of Primary Sources*,¹³ per la trascrizione della fonte primaria; modulo *Names, Dates, People, and Places*,¹⁴ per la rappresentazione delle entità nominate; modulo *Linking, Segmentation, and Alignment*,¹⁵ per le analisi di particolari strutture del testo; modulo *Simple Analytic Mechanisms*,¹⁶ per l'annotazione semantica e linguistico-lessicale delle unità testuali. Per quanto riguarda la descrizione della fonte

¹¹ I file ODD sono documenti XML-TEI utilizzati per la personalizzazione di qualsiasi schema di codifica conforme alle linee guida TEI.

¹² Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/it/html/PH.html#PH-bov>.

¹³ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.

¹⁴ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>.

¹⁵ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html>.

¹⁶ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html>.

primaria, sono stati impiegati, come norma, gli elementi definiti nel modulo *Manuscript Description*¹⁷ delle linee guida TEI.

Nel Listing 1 è riportato un esempio dello schema di codifica per le fonti scritte.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="it">
  <teiHeader>
    <fileDesc>
      <titleStmt></titleStmt>
      <editionStmt></editionStmt>
      <publicationStmt></publicationStmt>
      <sourceDesc>
        <msDesc></msDesc>
      </sourceDesc>
    </fileDesc>
    <encodingDesc></encodingDesc>
    <profileDesc>
      <creation></creation>
      <langUsage></langUsage>
      <textDesc></textDesc>
    </profileDesc>
  </teiHeader>
  <facsimile>
    <surface>
      <graphic/>
      <zone></zone>
    </surface>
  </facsimile>
  <text>
    <front></front>
    <body></body>
    <back></back>
  </text>
  <standOff>
    <listPerson></listPerson>
    <listPlace></listPlace>
    <listOrg></listOrg>
    <listBibl></listBibl>
  </standOff>
</TEI>
```

Listing 1: Schema di codifica di una fonte scritta

La rappresentazione della fonte primaria fa uso del tagset <facsimile> che permette la rappresentazione della fonte scritta sotto forma di un insieme di immagini, allineando la regione di interesse con la relativa porzione di trascrizione contenuta all'interno dell'elemento <text>. Ciascuna pagina del documento è rappresentata da un elemento <surface> (Listing 2) ed è stata

¹⁷ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>.

mappata riga per riga tramite lo strumento TEI Zoner,¹⁸ che ha permesso di generare molteplici elementi <zone> disegnando punti sull'immagine. Ogni elemento <surface> contiene infatti gli elementi <graphic> e <zone>: il primo rappresenta l'immagine, il secondo la zona specifica della superficie scrittoria, collegata alla singola riga codificata con l'elemento <lb/> all'interno del <text>. Questo tipo di marcatura è utilizzato per creare edizioni digitali interattive. Le immagini facsimilari delle pagine vengono annotate con zone che corrispondono a parti specifiche del testo per consentire agli utenti di interagire con le immagini e ottenere informazioni dettagliate sulle diverse parti del documento.

```

<surface xml:id="pag1" n="1"> <!-- PAG1 -->
  <graphic url="fotodiario/pag1.jpeg" width="1024px"
height="781px"/>
  <zone n="0" ulx="20" uly="14" lrx="1009" lry="762"
rendition="HotSpot" xml:id="pag1_l0"/>
  <zone n="1" ulx="259" uly="258" lrx="677" lry="290"
rendition="Line" xml:id="pag1_l1"/>
  <zone n="2" ulx="261" uly="297" lrx="630" lry="322"
rendition="Line" xml:id="pag1_l2"/>
</surface>

  <surface xml:id="pag2" n="2"> <!-- PAG2 -->
  <graphic url="fotodiario/pag2.jpeg" width="1024px"
height="697px"/>
  <zone n="0" ulx="12" uly="16" lrx="1009" lry="682"
rendition="HotSpot" xml:id="pag2_l0"/>
  <zone n="1" ulx="147" uly="113" lrx="767" lry="151"
rendition="Line" xml:id="pag2_l1"/>
  <zone n="2" ulx="143" uly="165" lrx="917" lry="190"
rendition="Line" xml:id="pag2_l2"/>
  <!-- [...] -->
</surface>

```

Listing 2: Esempio struttura facsimile nel Diario di Emanuele Artom

La trascrizione vera e propria della testimonianza è riportata all'interno del tag <body>, figlio di <text>, dove sono presenti elementi destinati alla rappresentazione del contenuto della fonte primaria, attenendosi ad essa il più possibile. Di conseguenza, la trascrizione della testimonianza contenuta in <body> può essere articolata in elementi <div>, il cui contenuto testuale può essere raggruppato in paragrafi (<p>) e suddiviso in pagine o in righe tramite i tag <pb/> (page beginning) e <lb/> (line beginning), i quali indicano rispettivamente l'inizio di una nuova pagina e l'inizio di una nuova riga nel testo originale (Listing 3).

¹⁸ Cfr. <http://teicat.huma-num.fr/zoner.php>

```

<text xml:lang="it">
  <front>
    <pb xml:id="pre1" facs="#pag1"/>
    <p></p>
    <pb xml:id="pre2" facs="#pag2"/>
    <p></p>
  </front>
  <body >
    <div type="giorno" xml:id="nov_43"></div>
    <div type="giorno" xml:id="g23nov_43"></div>
    <div type="giorno" xml:id="g24nov_43"></div>
    <div type="giorno" xml:id="g25nov_43"></div>
    <div type="giorno" xml:id="g27nov_43"></div>
    <div type="giorno" xml:id="g28nov_43"></div>
    <div type="giorno" xml:id="g29nov_43"></div>
    <div type="giorno" xml:id="g30nov_43"></div>
    <! -- [...] -->
  </body>
</text>

```

Listing 3: Esempio struttura <text> codifica fonte scritta

Durante la fase di codifica della testimonianza, sono stati individuati e marcati tutti gli elementi che permettessero una ricostruzione il più corretta possibile rispetto al valore semantico veicolato. Sono state quindi codificate le porzioni di testo che presentano omissioni o passaggi meno leggibili, tramite gli elementi <gap> e <supplied>, le modifiche apportate al testo dall'autore come le cancellazioni, le aggiunte e le sostituzioni, rispettivamente mediante i tag , <add> e <subst>. Tramite l'uso del tag <abbr> sono state segnalate le parole abbreviate, fornendo le loro espansioni complete con <expan>, e gli errori ortografici o di altro tipo, tramite il tag <sic>, accompagnandoli con la relativa forma corretta marcata con il tag <corr>. Per segnalare le parole o le espressioni normalizzate, ovvero riportate nell'italiano standard, si è fatto ricorso all'uso dell'elemento <orig> per indicare la grafia originale e <reg> per la versione normalizzata. In caso di frasi o parole diverse dalla lingua prevalente del testo, è stato utilizzato l'elemento <foreign> per marcare le espressioni riportate in lingua straniera. Infine, è stato utilizzato l'elemento <distinct> per identificare le parole ritenute linguisticamente distinte, come ad esempio le parole arcaiche o le parole tecniche.

2.2. Codifica di fonti orali

Gli elementi XML/TEI più significativi utilizzati per la descrizione delle fonti orali sono di seguito analizzati: le informazioni relative al supporto sono state registrate mediante l'elemento <recordingStmt>, contenuto a sua volta nell'elemento <sourceDesc>, appartenente al modulo *Transcriptions of Speech*.¹⁹ L'elemento <recording> infine rappresenta una singola registrazione e contiene tutte le informazioni necessarie a specificare il contesto e le responsabilità della registrazione. Ogni elemento <recording> è accompagnato dall'attributo di tipo (@type) per specificare la natura audio o video, e di una durata (@dur) per ogni singola registrazione. Infine, all'interno del blocco <profileDesc> sono stati impiegati gli elementi definiti nel modulo

¹⁹ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html>.

Language Corpora,²⁰ in particolare l'elemento <particDesc> offre una descrizione accurata delle persone che hanno preso parte alla conversazione (Listing 4).

```

<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="it">
  <teiHeader>
    <fileDesc>
      <titleStmt></titleStmt>
      <editionStmt></editionStmt>
      <publicationStmt></publicationStmt>
      <sourceDesc>
        <recordingStmt>
          <recording></recording>
        </recordingStmt>
      </sourceDesc>
    </fileDesc>
    <encodingDesc></encodingDesc>
    <profileDesc>
      <particDesc></particDesc>
      <langUsage></langUsage>
      <abstract></abstract>
    </profileDesc>
  </teiHeader>
  <standOff>
    <timeline></timeline>
  </standOff>
  <standOff>
    <listPerson></listPerson>
    <listPlace></listPlace>
    <listOrg></listOrg>
    <listBibl></listBibl>
  </standOff>
  <text>
    <body></body>
  </text>
</TEI>

```

Listing 4: Schema di codifica di una fonte orale

La differenza principale tra una fonte scritta e una fonte orale è la dimensione temporale di un discorso e l'ordine in cui i vari enunciati si alternano. Durante una conversazione orale tra più parlanti o durante un discorso proferito da un singolo, possono esserci momenti di sovrapposizione tra gli interlocutori, non sempre facilmente definibili. Sono quindi stati definiti specifici elementi <timeline>, inseriti all'interno del tag <standOff>, finalizzati alla sincronizzazione temporale degli argomenti trattati dal testimone, con lo scopo di allineare il minutaggio degli enunciati con le relative trascrizioni dei parlanti. A complemento della timeline è stata inoltre introdotta una sezione di sintesi ragionata, all'interno dell'elemento <abstract>, figlio di <profileDesc>, che contiene un riassunto del contenuto della trascrizione suddiviso in una lista di segmenti <item>.

²⁰ Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CC.html>.

In fase di codifica sono quindi state definite quattro <timeline>:

1. Nella prima timeline gli elementi <when/> individuano i momenti in cui si introducono i vari argomenti della testimonianza, quelli cioè riassunti all'interno degli elementi <item> presenti all'interno di <abstract>, i quali, a loro volta, possiedono un attributo @synch allo scopo di collegare il minutaggio specificato dal relativo tag <when/>
2. Nella seconda timeline, invece, sono stati raccolti tutti i segmenti in cui le voci si sovrappongono. Essendo questa timeline collegata al contenuto di <body>, all'interno della trascrizione le parole o le frasi pronunciate in contemporanea con un altro parlante sono state delimitate da due elementi <anchor>, uno all'inizio della sovrapposizione e uno alla fine. Questo permette di stabilire un doppio collegamento timeline-trascrizione e trascrizione-timeline proprio sfruttando gli elementi <when/>, presenti nella timeline, e gli elementi <anchor> presenti nel <body>
3. La terza timeline è stata realizzata per raggruppare gli istanti in cui avviene un cambio di parlante
4. La quarta permette di registrare gli istanti in cui agli enunciati si sovrappongono rumori di sottofondo

La trascrizione delle testimonianze orali, contenuta nell'elemento <text>, viene suddivisa in unità testuali dette *utterances* (enunciati) mediante l'uso dell'elemento <u>. Ogni enunciato è accompagnato dall'attributo @who che permette di associare ad esso la persona che lo ha formulato. In aggiunta, gli attributi @xml:id e @synch sono funzionali ad una corretta sincronizzazione con le timeline discusse in precedenza. L'attributo @trans, invece, specifica se gli enunciati dei partecipanti si susseguono oppure si sovrappongono.

Una conversazione orale presenta anche una serie di elementi che possono essere considerati come una componente del testo parlato, tutti identificati e marcati in fase di codifica. Durante la produzione di un enunciato è infatti possibile incontrare ripetizioni, false partenze e parole troncate, fenomeni segnalati attraverso l'elemento accompagnato dall'attributo @type, che può assumere i diversi valori *repetition*, *falseStart* o *truncation*. Altri fenomeni registrati sono ad esempio le pause vocali (elemento <pause> con attributo @type per indicare la lunghezza), i suoni non lessicali (elemento <vocal>), gli eventi prossemici (elemento <kinesic>), i rumori di sottofondo (elemento <incident>), i cambiamenti di caratteristiche paralinguistiche quali intonazione, volume, ritmo, velocità mediante l'elemento <shift/> accompagnato come si conviene dagli attributi @feature e @new. Infine, le attività di rappresentazione digitale di risorse orali presentano alcune difficoltà dal punto di vista dell'udibilità, poiché le registrazioni effettuate in situazioni di comunicazione naturale sono spesso imperfette, il che influisce sulla precisione della trascrizione. Queste situazioni sono state marcate con gli elementi <gap> e <unclear>.

3. Creazione dell'applicazione web

L'applicazione web *Voci dall'Inferno* è stata realizzata aprendo il proprio bacino di utenza all'intero corpus di testimonianze che compongono l'archivio digitale, cercando di fornire una prima struttura generale in grado di gestire correttamente anche l'aggiunta futura di nuovi dati. Essa consente la gestione di tutte le tipologie di testimonianze che costituiscono il corpus e ha permesso di effettuare le prime analisi e i primi studi considerando tutti i dati a disposizione dell'archivio digitale.

3.1 L'applicazione web *Voci dall'Inferno*

L'applicazione web *Voci dall'Inferno* è stata realizzata mediante le tecnologie messe a disposizione dall'ambiente eXist-db. La piattaforma integra il modulo *HTML templating framework* che permette la generazione dinamica di pagine HTML partendo da collezioni di documenti XML e da funzioni implementate mediante il linguaggio di interrogazione XQuery. Il funzionamento prevede l'uso di documenti-templates in HTML, in cui si aggiungono opportune chiamate a funzioni XQuery che navigando la collezione di dati, favoriscono l'estrazione dell'informazione relativa ai nodi XML di interesse e la generazione di frammenti HTML che permettono di completare la pagina restituita all'utente finale. Il principale vantaggio dell'uso dei templates è quello di mantenere separato il codice di presentazione da quello di elaborazione dei dati.

Nel Listing 5 è riportato un esempio di una funzione XQuery attraverso la quale viene creato l'elenco dei testimoni che appartengono ad una specifica collezione.

```

declare function app:ElencoTestimoniPerCollezione($node as node(),
$model as map(*) ){
  let $nome_collezione :=request:get-parameter("collezione","")
  let $collezione:=
    if ($nome_collezione="deportati") then
      let $path:= "/db/apps/Voci_Dall_Inferno/xml/Deportati"
      return $path
    else
      let $path:= "/db/apps/Voci_Dall_Inferno/xml/NonDeportati"
      return $path

  let $testimoni:=
    for $xml in collection($collezione)
    let $find_testimone := $xml//tei:person[@role="testimone"]
    let $nome_testimone := fn:normalize-
space($find_testimone/tei:persName/tei:forename)
    let $cognome_testimone:=fn:normalize-space(string-
join($find_testimone/tei:persName/tei:surname, " "))
    let $testimone :=concat($cognome_testimone," ",$nome_testimone)

    order by $testimone
    return $testimone

  for $testimone in distinct-values($testimoni)
  return
    <div class="elenco_testimoni">
      <section class="testimone">
        <p>
          <a class="nome_testimone"
onclick='mostra_testimonianze("{ $testimone}")'>{($testimone)}</a>
        </p>
      </section>
    </div>
};

```

Listing 5: Funzione XQuery per la creazione delle liste dei testimoni in base alla collezione

Attualmente l'applicazione non è ancora disponibile per l'uso da parte del pubblico in quanto è necessario adottare opportune licenze per gestire l'accesso ai dati sensibili contenuti nelle testimonianze trattate.

L'applicazione web *Voci dall'Inferno* si compone di 9 sezioni distinte, ciascuna raggiungibile mediante le voci presenti nel menù di navigazione: *Home*, *Il progetto*, *Le voci*, *Cerca un testimone*, *Dante*, *Statistiche*, *Trascrizione automatica*, *Eventi*, *Bibliografia*.

La sezione *Il progetto* descrive i motivi per i quali i sopravvissuti si sono serviti del vocabolario offertogli da Dante per raccontare l'esperienza ineffabile e indicibile del Lager e per tentare di stabilire, mediante la *Commedia*, un ponte tra il loro racconto e gli uditori. Viene poi inquadrato in tale contesto il progetto *Voci dall'Inferno* e ne viene ripercorsa la storia, dai primi anni ad oggi, sia dal punto di vista delle prime ricerche effettuate, sia dal punto di vista dell'evoluzione delle metodologie informatiche delle quali si è fatto uso nel corso del tempo.

La voce *Eventi* propone una rassegna di convegni e seminari in occasione dei quali sono stati presentati i risultati raggiunti, mentre in *Bibliografia* sono raccolti i materiali prodotti contestualmente al progetto.

La sezione *Cerca un testimone* offre la possibilità di effettuare la ricerca di un testimone all'interno dell'archivio inserendo le informazioni all'interno del *form* (Figura 2). È possibile ricercare la persona in base al nome, al cognome, alla data o al luogo di nascita oppure alla data o al luogo di morte.

Cerca un testimone
Compila uno dei campi con le relative informazioni per cercare un testimone all'interno del corpus

Nome:	<input type="text"/>
Cognome:	<input type="text"/>
Anno di nascita:	<input type="text"/>
Luogo di nascita:	<input type="text"/>
Anno di morte:	<input type="text"/>
Luogo di morte:	<input type="text"/>

Figura 2: Ricerca di un testimone mediante il form

Il cuore vero e proprio dell'applicazione è la sezione *Le voci*, dalla quale si accede alla consultazione dei testimoni presenti nell'archivio e alle relative testimonianze. La sezione presenta una panoramica sui dati relativi ai testimoni e alle categorie che compongono il corpus, offrendo anche a livello grafico la possibilità di consultare e interagire con un *treegraph chart* che illustra la struttura gerarchica dei testimoni all'interno della tassonomia attualmente definita. Cliccando su uno dei due pulsanti disponibili, *Deportati* o *Non deportati*, si accede all'elenco dei testimoni che costituiscono la categoria scelta (Figura 3).

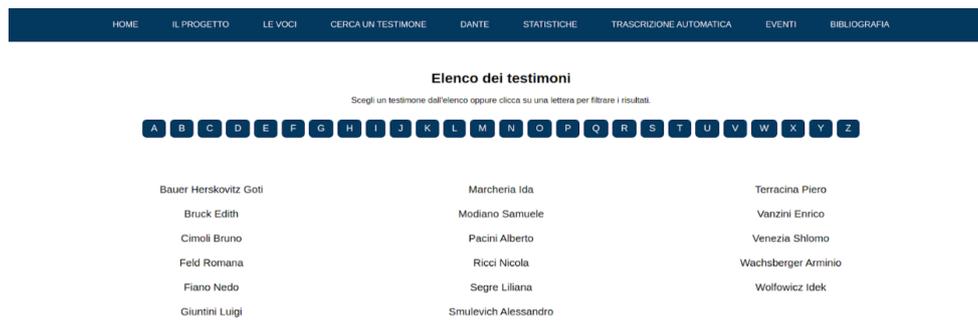


Figura 3: Elenco dei testimoni che costituiscono l'archivio digitale

Nella parte in alto della pagina è stata inserita la possibilità di cliccare su una delle lettere del menù per filtrare i risultati in base all'iniziale del cognome. La scelta di tale implementazione è dettata dal tentativo di agevolare la scrematura dei dati e di migliorarne la consultazione, considerando la possibilità di avere, un domani, un archivio molto più ampio. Cliccando sul nome del testimone di cui si vuole consultare una testimonianza, si accede ad una pagina in cui è mostrata una breve biografia del testimone e l'elenco delle relative testimonianze codificate e quindi disponibili all'utente (Figura 4).



Figura 4: Scheda anagrafica del testimone ed elenco delle relative testimonianze consultabili

Di recente sono state integrate nell'applicazione due funzionalità, attualmente in fase di affinamento [22]. Le implementazioni riguardano un grafo orientato ed etichettato che rappresenta le relazioni tra le persone menzionate nella testimonianza (Figura 5) e due mappe che illustrano rispettivamente i luoghi menzionati nelle testimonianze e il flusso degli spostamenti effettuati dal testimone prima, durante e dopo la deportazione (Figura 6).

mettere in evidenza nel testo gli elementi marcati (Figura 7). Sia la trascrizione della testimonianza che la legenda cambieranno la struttura in base al tipo di fonte consultata.

The screenshot shows a web-based transcription interface. At the top, there is a dark navigation bar with menu items: HOME, IL PROGETTO, LE VOCI, CERCA UN TESTIMONE, DANTE, STATISTICHE, TRASCRIZIONE AUTOMATICA, EVENTI, BIBLIOGRAFIA. The main content area is split into two panels. The left panel displays a transcription of a testimonial with various markers in red, green, and blue. The right panel shows an active legend titled 'Fenomeni marcati' (Marked Phenomena) with a list of markers and their corresponding colors: GAP XXX (red), UNCLEAN (green), PAUSE (-) (blue), VOCAL (cyan), INCIDENT (magenta), KROMEIC (orange), DEL (yellow), SIC (light blue), CORR (dark blue), ORIG (purple), REG (pink), ABBR (brown), EXPAN (grey), FOREIGN (dark red), PERSONAME (light green), PLACE NAME (light purple), and ORONAME (light yellow). A 'SICURO TUTTA (FRANCESCO)' button is visible at the bottom of the legend panel.

Figura 7: Trascrizione di una testimonianza con legenda attiva

La trascrizione del linguaggio parlato è una componente fondamentale per la ricerca umanistica: è infatti molto utile per discipline come la fonetica, la sociolinguistica, la dialettologia e l'analisi delle conversazioni e del discorso.

Negli ultimi anni sono state sviluppate molteplici convenzioni di trascrizione inerenti a lingue, interessi e tradizioni diverse. Le convenzioni vengono implementate per mezzo dell'utilizzo di una varietà di strumenti informatici, ognuno con il proprio modello di dati e formati. Non esiste quindi un metodo dominante o standardizzato per la trascrizione del linguaggio parlato. [20]

Per ridurre la variazione di standardizzazione è fondamentale tenere in considerazione alcuni fattori:

1. I trascrittori devono lavorare con strumenti informatici adeguati;
2. La standardizzazione dovrebbe distinguere attentamente tra la variabilità presente nelle convenzioni di trascrizione e nei formati degli strumenti che può derivare da differenze di interesse di ricerca e approcci teorici o da idiosincrasie individuali e proporre unificazioni solo per il primo caso;
3. Ogni sforzo di standardizzazione deve lasciare spazio a negoziazioni tra le parti coinvolte [20]

Tuttavia, attualmente non esiste uno standard comune a tutti gli ambiti e necessità. Questa funzionalità è attualmente in corso di sviluppo. Per il momento, riprendendo il lavoro svolto da Diletta Lelli, [11] per la trascrizione delle testimonianze orali si sono utilizzate le convenzioni offerte da CHAT,²¹ adottando l'identificazione di due fenomeni più ricorrenti nelle testimonianze: le lacune e le pause. Per le prime, in CHAT la convenzione utilizzata è xxx; per le seconde è (...).

Grazie alla marcatura di numerosi fenomeni in fase di codifica, per ciascuna testimonianza è possibile consultare alcune analisi statistiche. I dati relativi a tali fenomeni sono stati estratti e utilizzati per la creazione di tre grafici, ciascuno riguardante un determinato aspetto: il primo

²¹ Cfr. <https://talkbank.org/manuals/CHAT.pdf>

grafico mostra i fenomeni che caratterizzano la fonte in questione, che differiscono in base al tipo di testimonianza consultata; il secondo grafico, pone invece l'attenzione sul modo di esprimersi del testimone, mostrando il dato espresso sia in forma percentuale che in forma assoluta (Figura 8 e Figura 9); il terzo grafico, infine, prende in analisi le entità nominate.

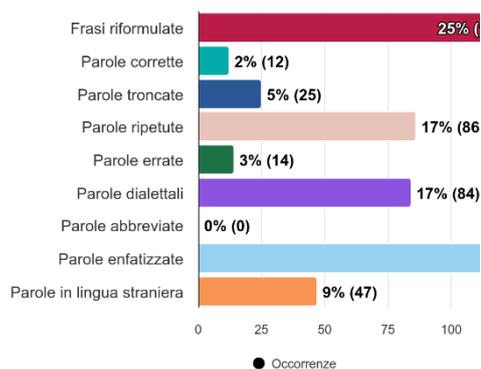


Figura 8: come si esprime Ida Marcheria

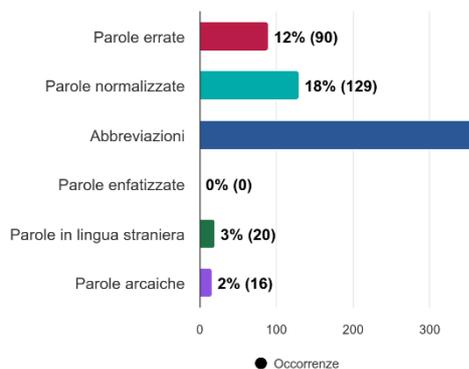


Figura 9: come scrive Emanuele Artom

Infine, è presente la sezione che consente di visualizzare se nella testimonianza consultata sono presenti riferimenti alla *Divina Commedia* e, in caso, quanti e quali sono. Durante la fase di trascrizione e codifica sono state individuate e distinte, mediante un'opportuna marcatura, le differenti tipologie di tessere dantesche: le citazioni implicite, le allusioni e i vocaboli danteschi. Cliccando sulla citazione o allusione riportata nella parte sinistra della schermata (Figura 10) sarà possibile mettere in evidenza i versi o la porzione di testo della *Commedia* alla quale si fa riferimento.²²

²² L'edizione dell'opera utilizzata come riferimento per il testo indicizzato della *Commedia* e per le ricerche comparative nell'applicazione *Voci dall'Inferno* è la versione digitalizzata disponibile sulla Biblioteca Italiana (Collezione BibIt, codice: bibit000019). Questa edizione digitale si basa sull'edizione critica curata da Giorgio Petrocchi pubblicata nel 1994, riconosciuta per il suo valore scientifico [28].

Il testo è consultabile al seguente indirizzo: <http://www.bibliotecaitaliana.it/scheda/bibit000019>.

La scelta di questa risorsa è motivata da diversi fattori, tra cui: la disponibilità del testo in formato standard XML/TEI, l'autorevolezza dell'edizione a stampa di riferimento e la facilità di accesso alla risorsa nel periodo preso in esame.

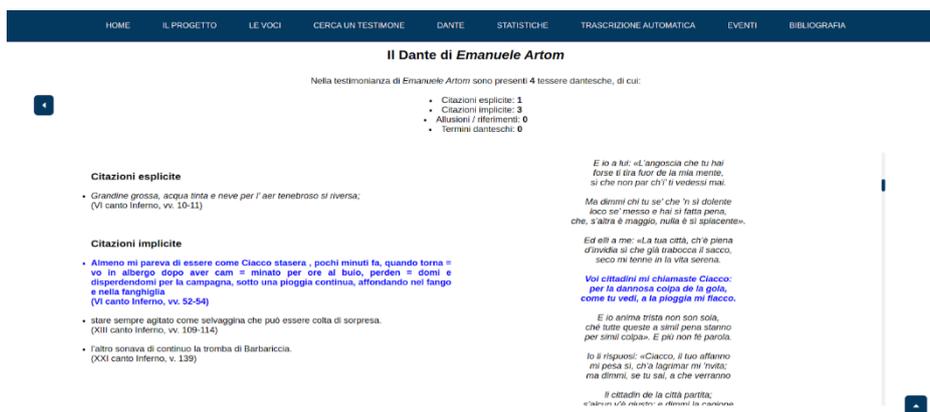


Figura 10: Visualizzazione delle tessere dantesche nella testimonianza di Emanuele Artom

Per il momento questa funzionalità di collegamento tra le citazioni presenti nella testimonianza e l'opera dantesca è disponibile solamente per la testimonianza di Emanuele Artom. Inoltre, l'identificazione delle tessere dantesche presenti nelle testimonianze fino a ora è stata realizzata manualmente. Attualmente è in fase di sviluppo un prototipo di applicazione web denominata *Voci dall'Inferno Verse Similarity Search*, un sistema progettato per individuare citazioni e allusioni al lessico dantesco mediante approcci computazionali alla ricerca di frasi presenti nelle testimonianze e il confronto di essi con i versi presenti nella *Divina Commedia* di Dante Alighieri. La metodologia per la realizzazione di tale applicazione verrà approfondita nella sezione 3.2.

Un'altra interessante e importante sezione dell'applicazione web è quella chiamata *Dante*, che presenta i risultati emersi dalla prima analisi svolta sull'intero corpus circa la presenza di tessere dantesche nelle testimonianze di *Voci dall'Inferno*. Come già scritto precedentemente, l'archivio digitale di *Voci dall'Inferno* è costituito da 18 testimoni, per un totale di 23 testimonianze. Da una prima analisi generale è emerso che su 18 testimoni, ben 11 si servono del vocabolario dantesco per dire *l'ineffabile*.

Per studiare nello specifico le tessere dantesche individuate e marcate in fase di codifica, si sono svolte due differenti analisi:

1. nella prima, si è analizzata la tipologia dei riferimenti alla *Commedia* che in fase di codifica sono stati opportunamente individuati e distinti mediante diversa marcatura. Questa operazione ha permesso di approfondire l'indagine estraendo e analizzando in dettaglio i dati relativi alle diverse forme attraverso le quali le tessere dantesche si sono presentate nelle testimonianze.
2. nella seconda, si sono analizzati i testimoni che utilizzano vocaboli danteschi, studiando coloro che ricorrono maggiormente a Dante per dire il Lager e attraverso quale modalità.

Nelle testimonianze che costituiscono l'archivio sono state rinvenute 61 tessere dantesche, di cui 4 citazioni implicite, 15 citazioni esplicite, 7 allusioni e 35 vocaboli danteschi (Figura 11).

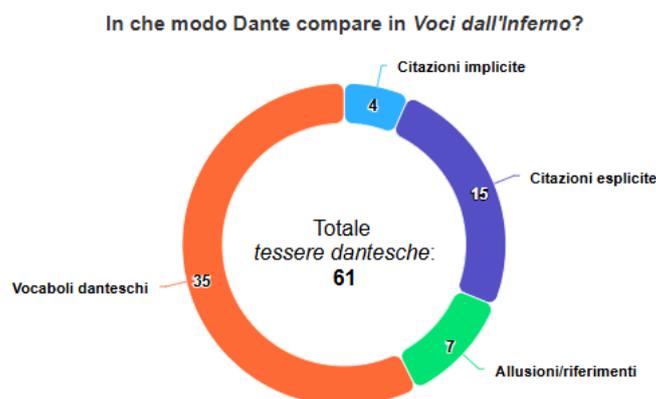


Figura 11: Tessere dantesche rilevate nel corpus

Ponendo l'attenzione esclusivamente sulle citazioni, i dati che sono emersi dallo studio delle testimonianze dell'archivio digitale dimostrano che la maggior parte delle citazioni sono riconducibili alla prima cantica, la quale presenta infatti 16 citazioni su 19 totali. Nonostante ciò, anche il Purgatorio e il Paradiso sono stati citati dai testimoni, seppur in misura minore: il primo 2 volte, il secondo 1. Il canto a cui i testimoni fanno maggiore riferimento risulta essere *Inferno* I, che presenta 3 citazioni, seguito da *Inferno* II, *Inferno* IV, *Inferno* VI che presentano 2 citazioni ciascuno. Gli altri canti dai quali i testimoni attingono una citazione sono *Inferno* XVII, *Inferno* XXVII, *Inferno* VII, *Inferno* XIII, *Inferno* XXI, *Inferno* III e *Inferno* V. Del *Purgatorio* i 2 canti citati sono *Purgatorio* I e *Purgatorio* IX, mentre della terza cantica viene citato *Paradiso* XII (Figura 12).

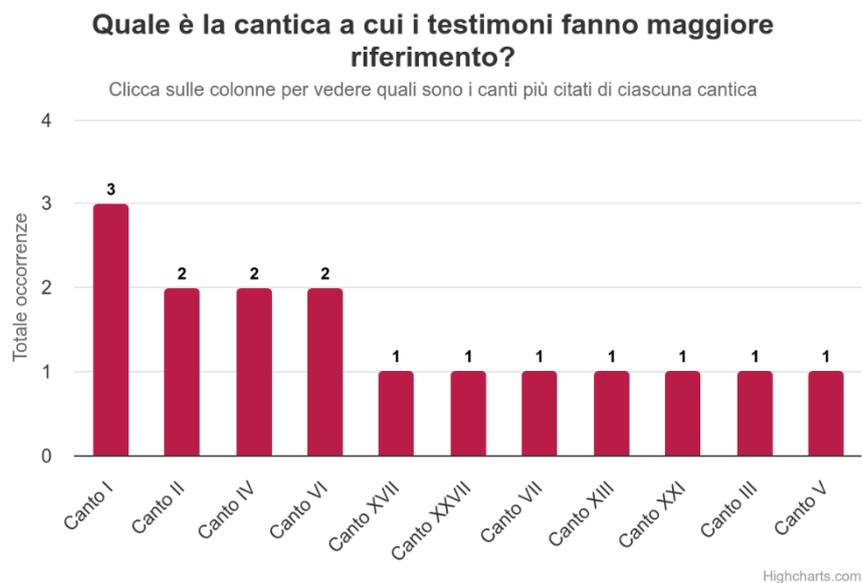


Figura 12: Ripartizione citazioni per canti dell'Inferno

I vocaboli danteschi che in fase di codifica sono stati individuati e marcati sono 39. Il termine che è risultato più frequente nelle parole dei testimoni è *inferno*, che occorre ben 21 volte. Gli altri vocaboli dei quali si fa maggior uso sono *dante*, che occorre 4 volte, *bolgia*, *virgilio* e *divino poeta*, con 2 occorrenze ciascuno.

La seconda analisi si è concentrata su chi sono coloro che fanno uso della *Commedia* per esprimere l'ineffabile. Dei 18 testimoni che costituiscono l'archivio, in 10 sono presenti tessere dantesche. Luigi Giuntini è in assoluto colui che si serve maggiormente delle parole di Dante, nel cui diario si sono registrate 19 tessere dantesche (Figura 13).

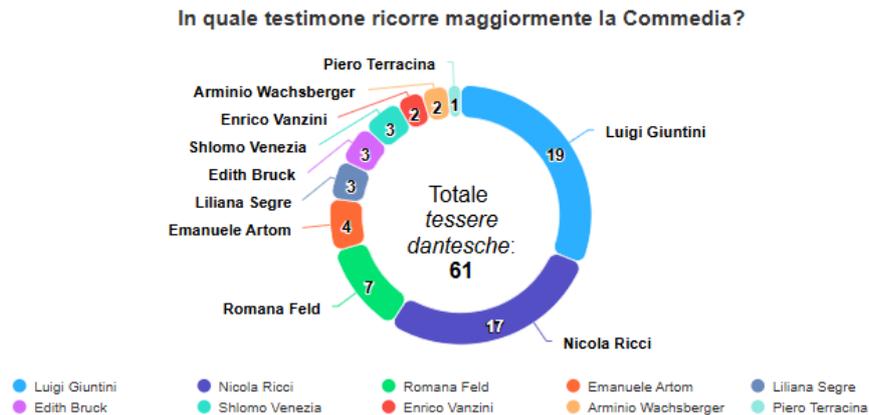


Figura 13: Testimoni che attingono maggiormente dalla Commedia

Dopo una prima analisi più generale su quali testimoni e in che misura si servono del lessico e delle immagini dantesche, si è svolta un'indagine più approfondita, andando ad analizzare ciascuna modalità attraverso cui Dante si presenta nelle testimonianze. Per ciascuna tipologia è stato valutato nel dettaglio quanto ciascun testimone attinge dalla *Commedia*, ed è stata realizzata una tabella che riporta rispettivamente le citazioni, le allusioni e i termini identificati (Figura 14).

Cantica	Canto	Testimone	Titolo testimonianza	Citazione esplicita
Inferno	Canto XVII	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	« Le ripe eran grommate d'una muffa per l'alto di giù che vi s' appasta che con gli occhi e col naso faceva zuffa »
Inferno	Canto XXVII	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	" contraddizioni che noi consente "
Inferno	Canto I	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	« La notte ch'io passai con tanta pièta »
Inferno	Canto I	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	« Temp'era dal principio del mattino e 'l sol montava 'n sù con quelle stelle ch'eran con lui quando l'amor divino mosse di prima quelle cose belle ; »
Inferno	Canto I	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	« "Miserere di me" gridai a lui "qual tu sii, od ombra od omo certo?" »
Inferno	Canto II	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	« Lo giorno se n'andava, e faere bruno toglieva gli animal che sono in terra dalle fatiche loro »
Inferno	Canto IV	Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	"a bella corda "

Figura 14: Tabella delle citazioni esplicite

Analizzando le citazioni espresse in forma esplicita (15), si è osservato che la maggior parte (12) fa riferimento alla prima cantica della *Commedia*. Appartengono a questa tipologia le citazioni provenienti da *Purgatorio* e *Paradiso*: della seconda cantica sono presenti infatti due citazioni, della terza, una citazione. Luigi Giuntini risulta colui che cita maggiormente in modo esplicito. Seguono, con una sola citazione, Emanuele Artom, Liliana Segre e Nicola Ricci.

Per quanto riguarda, invece, le citazioni implicite (4), fanno tutte riferimento esclusivamente all'*Inferno*. Tre citazioni provengono dal diario di Emanuele Artom e 1 da quello di Nicola Ricci.

Le allusioni registrate nelle testimonianze del corpus sono 10: 3 sono presenti nell'intervista di Goti Bauer, 2 nella testimonianza di Romana Feld, 2 nel diario di Nicola Ricci, 1 nell'intervista di Arminio Wachsberger, 1 in Enrico Vanzini e 1 in Liliana Segre.

Per quanto riguarda i vocaboli, ben 10 testimoni ricorrono al lessico propriamente dantesco. Dei 39 termini individuati nel corpus, 13 provengono dal diario di Nicola Ricci, 7 da quello di Luigi Giuntini e 5 dalla testimonianza di Romana Feld. Nell'intervista di Goti Bauer ne sono stati rilevati 4, in quelle di Edith Bruck e Shlomo Venezia 3 e 1 solo termine nelle testimonianze di Liliana Segre, Piero Terracina, Enrico Vanzini ed Arminio Wachsberger.

L'applicazione web presenta anche la sezione *Statistiche*, nella quale sono riportati i "numeri" di *Voci dall'Inferno* e alcuni risultati emersi da diverse indagini svolte. Per prima cosa ci si è concentrati su chi sono i testimoni che attualmente popolano l'archivio: è stato quindi realizzato un *treegraph chart* (Figura 15) che riporta la tassonomia completa con i nomi di tutti i testimoni che costituiscono il corpus. Si è scelto di utilizzare una struttura gerarchica per favorire visivamente la comprensione della ripartizione interna dei dati che compongono l'archivio.

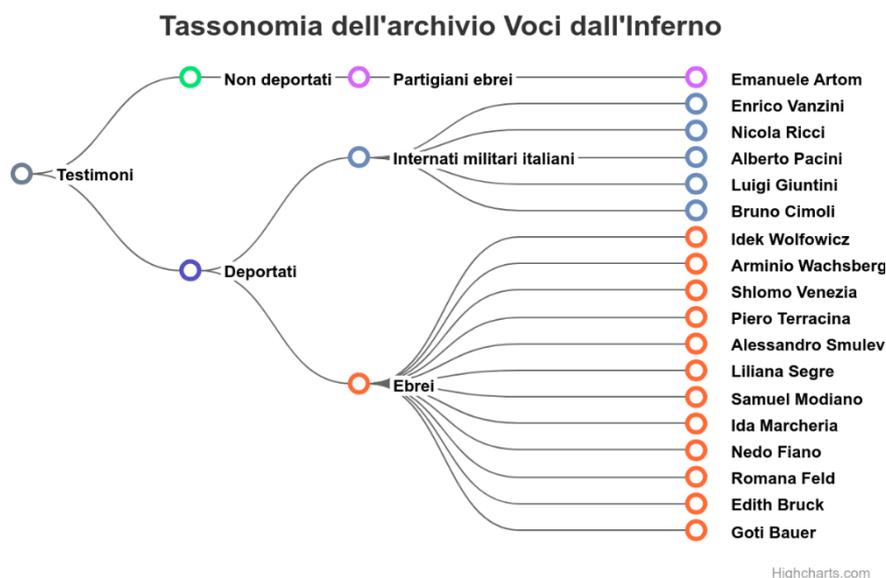


Figura 15: Treegraph chart che illustra la tassonomia completa dell'archivio

Un'ulteriore indagine è stata svolta circa la provenienza delle fonti. Per quanto riguarda la macro-classe delle testimonianze orali, è emerso che la maggior parte proviene dalle interviste inedite realizzate dalle dottoresse Segre e Pavoncello: su 13 testimonianze orali, 8 provengono da questo fondo. Delle restanti, le 3 in formato video provengono da YouTube,²³ una è stata realizzata da Marina Riccucci e una appartiene all'archivio della Fondazione CDEC (Figura 16). Delle 23 testimonianze totali, le restanti 10 sono testimonianze scritte (Figura 17).

²³ Sono le interviste rilasciate a Ivan Scalfarotto, Telepace e Associazione Figli della Shoah.



Figura 16: provenienza delle fonti orali



Figura 17: provenienza delle fonti scritte

3.2 Funzionalità in via di sviluppo

Tra le implementazioni più recenti, *Voci dall'Inferno* ha visto un primo impiego di modelli di intelligenza artificiale e approcci computazionali per la realizzazione di funzionalità volte ad automatizzare specifiche operazioni che al momento sono realizzate manualmente.

Una prima implementazione di una funzionalità che vede l'utilizzo di un modello di intelligenza artificiale è presente all'interno della sezione *Trascrizione automatica*. Cliccando sul tasto *Upload file* è possibile caricare un file audio per ottenere mediante l'*automatic speech recognition* la trascrizione nell'area sottostante. Nell'esempio mostrato in Figura 18 è stata trascritta una frase della testimonianza di Sami Modiano.

Figura 18: trascrizione automatica di un file audio

Per la realizzazione di tale funzionalità si è utilizzato *Whisper*,²⁴ un modello di intelligenza artificiale sviluppato da OpenAI²⁵ rilasciato a settembre 2022. Whisper è un sistema di riconoscimento vocale automatico (ASR) addestrato su 680.000 ore di dati supervisionati multilingue e multitask raccolti dal web. È quindi anche un modello multi-task in grado di eseguire il riconoscimento vocale multilingue, nonché la traduzione vocale e l'identificazione della lingua. Si distingue per la sua capacità di trascrivere non solo audio di alta qualità ma anche quelli che presentano rumore di fondo, accenti particolari o altre difficoltà comuni nel riconoscimento vocale tradizionale. A dicembre 2022 è stata rilasciata la versione *Whisper Large v2*,²⁶ addestrata su un numero di epoche 2,5 volte superiore rispetto alla prima versione. A novembre 2023, il modello *Whisper Large v3*,²⁷ è stato rilasciato nella sua terza versione, che ha migliorato ulteriormente la precisione delle trascrizioni rispetto a *Whisper v2*, grazie a un'ulteriore ottimizzazione del modello e a una maggiore quantità di dati di addestramento, migliorando la capacità di riconoscere accenti meno comuni e di gestire situazioni con maggiore rumore di fondo, mostrando una riduzione degli errori dal 10% al 20% rispetto a *large-v2*.

All'interno dell'applicazione web *Voci dall'Inferno* è stato utilizzato il modello *Whisper Large v3* integrandolo mediante le API disponibili su HuggingFace.²⁸

La possibilità di poter effettuare la trascrizione automatica del parlato è soltanto un supporto parziale e un punto di partenza per il progetto *Voci dall'Inferno*. L'archivio digitale è costituito infatti da documenti XML nei quali sono stati manualmente individuati e rappresentati, mediante un'opportuna codifica descritta nel capitolo 2, tutti i fenomeni che caratterizzano una fonte orale. Di conseguenza, strumenti per ottenere una trascrizione automatica sono sicuramente un supporto, ma è comunque necessario l'intervento del codificatore per annotare tutti i fenomeni al fine di ottenere tutti i dati utili da inserire all'interno dell'archivio. Tra gli obiettivi futuri del progetto, una volta ampliato maggiormente il corpus, c'è quindi quello di addestrare un modello che, oltre ad effettuare la trascrizione automatica del parlato, è anche in grado di realizzare la marcatura di tutti i fenomeni che caratterizzano una fonte orale.

Un primo utilizzo, invece, di modelli computazionali riguarda lo sviluppo di un prototipo di applicazione web denominata *Voci dall'Inferno Verse Similarity Search*, un sistema progettato per individuare citazioni e allusioni al lessico dantesco mediante approcci computazionali alla ricerca di frasi presenti nelle testimonianze e il confronto di essi con i versi presenti nella *Divina Commedia* di Dante Alighieri.

L'applicazione è stata sviluppata usando il linguaggio di programmazione Python e ha richiesto l'utilizzo di due tecnologie:

- (a) *Weaviate*,²⁹ una piattaforma di vector search open-source che, sfruttando modelli di Machine Learning allo stato dell'arte, trasforma dati non strutturati

²⁴ Cfr. <https://openai.com/index/whisper/>

²⁵ Cfr. <https://openai.com/>

²⁶ Cfr. <https://github.com/openai/whisper/discussions/661>

²⁷ Cfr. <https://github.com/openai/whisper/discussions/1762>

²⁸ Cfr. <https://huggingface.co/openai/whisper-large-v3>.

²⁹ Cfr. <https://weaviate.io/>

in un database vettoriale ricercabile sul quale è possibile eseguire interrogazioni e analisi semantiche.

- (b) *Streamlit*,³⁰ un framework open-source dell'ecosistema Python che facilita la creazione e lo sviluppo di applicazioni web personalizzate.

L'approccio adottato per l'individuazione dei tasselli danteschi nel corpus di testimonianze fa leva sull'uso di metriche per il calcolo della "Sentence Similarity", [16] ovvero il processo di valutazione della similarità tra due testi costituiti da poche parole che abbiano una coerenza sintattica/semantica. Ciò infatti viene realizzato utilizzando un modello matematico di *sentence similarity*, che permetta di trasformare i dati da testo piano in punti di uno spazio vettoriale misurabili, detti *embeddings* [7]. Questo permette di misurare la similarità tra frammenti di testo, calcolando la loro vicinanza semantica.

Al momento è stato adottato il modello di *sentence similarity* denominato SentenceTransformers *all-mpnet-base*, pubblicamente accessibile per mezzo della piattaforma HuggingFace.³¹

Per il corretto funzionamento dell'applicazione è stato necessario prevedere diverse fasi di sviluppo. Dopo aver raccolto le terzine delle cantiche della *Divina Commedia* in una struttura tabellare, mediante il modello di SentenceTransformers sono stati calcolati i corrispondenti *embeddings*. L'utilizzo di Weaviate permette la configurazione di una collezione per archiviare le informazioni strutturate relative alla *Divina Commedia* e la creazione di un database per categorizzare e archiviare i dati relativi a cantiche, canti, versi e terzine. In questo modo è possibile archiviare gli *embeddings* ed eseguire ricerche semantiche sui dati. Dopo aver chiamato la funzione che consente il calcolo del punteggio di similarità tra la query ricercata e gli *embeddings* delle terzine il sistema restituisce i passi più simili ordinati secondo un punteggio di similarità (*similarity score*). Data una sequenza di parole il sistema restituisce le terzine semanticamente più simili mediante il calcolo del *similarity score* (Figura 19).

³⁰ Cfr. <https://streamlit.io/>

³¹ Cfr. <https://huggingface.co/>

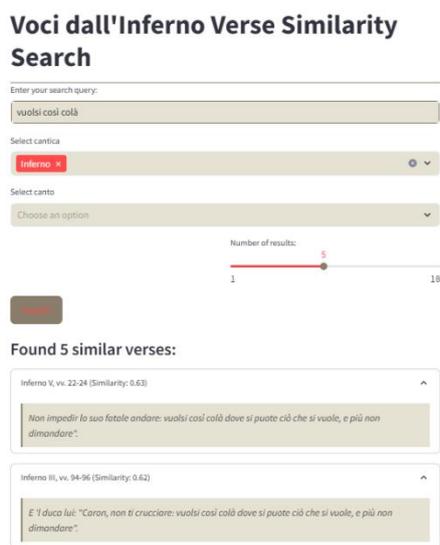


Figura 19: Screenshot dell'applicazione Voci dall'Inferno Verse Similarity Search

4. Conclusioni

Dal 2016 ad oggi il progetto *Voci dall'Inferno* ha raggiunto notevoli risultati. Mediante il supporto di strumenti informatici e un nutrito gruppo di studenti, è stato possibile innanzitutto costituire il primo *corpus* digitale di testimonianze *non letterarie* di sopravvissuti ai Lager e, successivamente, analizzare, interrogare e studiare il materiale digitalizzato in merito alla presenza di lessico e immagini dantesche all'interno delle testimonianze. L'applicazione permette di raccogliere, di consultare, di analizzare e di interrogare tutte le testimonianze che costituiscono l'archivio. Consente inoltre di censire la presenza di tessere dantesche all'interno del corpus e di analizzare le differenti tipologie attraverso le quali Dante si presenta nelle testimonianze.

L'analisi preliminare condotta sull'intero corpus, composto da 23 testimonianze di 18 sopravvissuti, ha evidenziato un dato significativo: oltre la metà dei testimoni ricorre a riferimenti all'opera dantesca per esprimere e rappresentare l'esperienza concentratoria. In totale sono stati individuati 61 riferimenti alla Commedia, articolati in 15 citazioni esplicite, 4 implicite, 7 allusioni e 35 vocaboli di matrice dantesca. Questi risultati incoraggiano non solo l'ampliamento del corpus, ma anche ulteriori riflessioni, indagini e analisi computazionali sul ruolo della memoria letteraria e sull'intertestualità come strumenti di comunicazione.

L'applicazione, nella sua attuale implementazione, è un punto di partenza per lo sviluppo di ulteriori funzionalità. Tra le implementazioni più recenti, *Voci dall'Inferno* lo sviluppo di un primo prototipo per il riconoscimento automatico del parlato basata su tecnologie ASR (Automatic Speech Recognition). Sebbene il sistema di trascrizione automatica offra un supporto iniziale molto utile, il suo contributo rimane parziale: è ancora necessario l'intervento del codificatore per annotare accuratamente i fenomeni rilevanti secondo gli standard adottati al fine di ottenere ulteriori dati significativi da integrare nell'archivio. Una fonte orale è composta da un insieme complesso di elementi verbali, non verbali e prossemici che ne definiscono l'unicità e che devono essere opportunamente identificati e marcati durante la fase di codifica. Tra questi si annoverano,

ad esempio, frasi interrotte o lasciate incompiute, ripetizioni, anacoluti, pause vocali, rumori di fondo, variazioni di tono e, nel caso di registrazioni audiovisive, anche gesti e movimenti del corpo. Fenomeni codificabili mediante gli elementi TEI <vocal>, <kinesic> o <pause>.

Ad esempio, segue un frammento di trascrizione del parlato adottando le convenzioni dello standard CHAT.

IM: **xxx** c'era una resistenza nel campo, i miracoli [?] non li potevano fare, cercavano di salvare (...) qualche cosa. Sapevano ↑ da cinque anni.

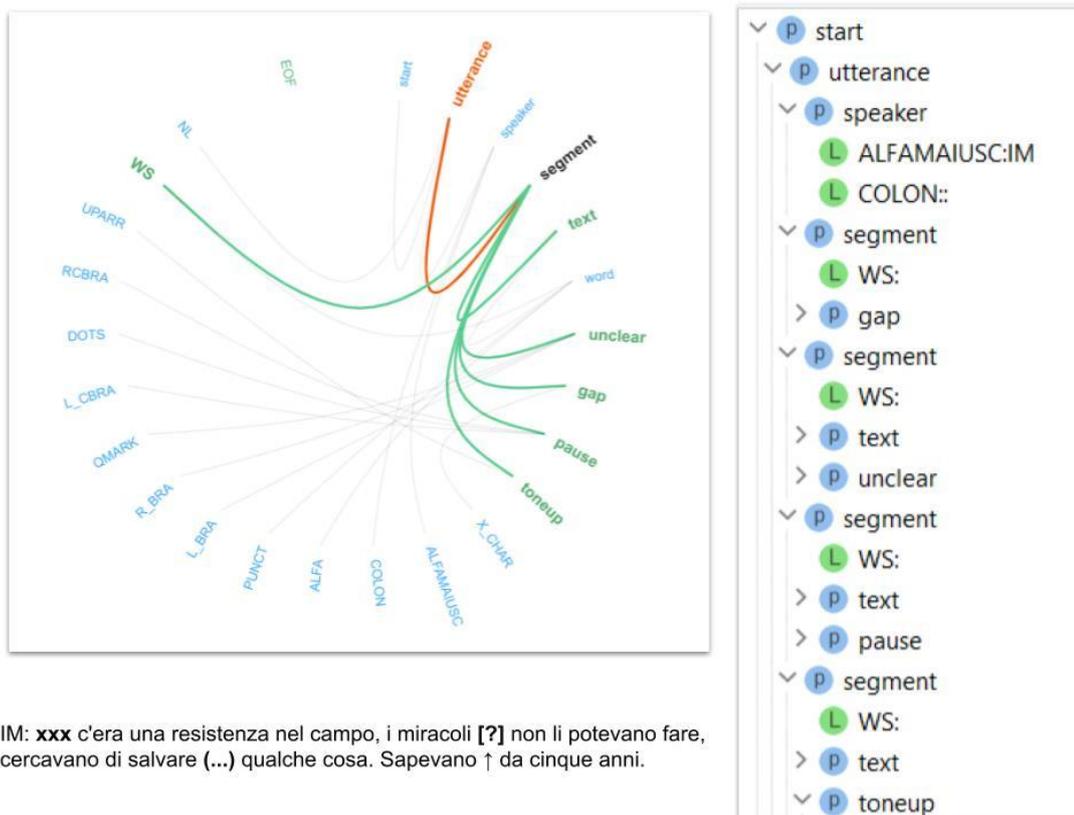


Figura 20: Esempio di trascrizione da testimonianza orale secondo lo standard CHAT e una grammatica formale (CFG). A sinistra, visualizzazione delle relazioni testuali; a destra, struttura ad albero del testo riconosciuto.

Il frammento è riconoscibile attraverso una grammatica libera da contesto, che lo rende trattabile come un linguaggio specifico di dominio (DSL), come mostrato in Figura 20. In essa sono evidenziati i fenomeni individuati automaticamente dalla procedura di analisi, tra cui speaker (IM), interruzioni (*gap*), segmenti non chiari (*unclear*), variazioni di tono (*toneup*), e altri.

Tra gli obiettivi futuri del progetto, una volta ampliato maggiormente il corpus, c'è quindi quello di addestrare un modello che, oltre ad effettuare la trascrizione automatica del parlato, è anche in grado di realizzare la marcatura di tutti i fenomeni che caratterizzano una fonte orale. Altre recenti implementazioni effettuate riguardano la realizzazione di un grafo per rappresentare le relazioni tra le persone menzionate nella testimonianza e due mappe che illustrano rispettivamente i luoghi citati e il flusso degli spostamenti effettuati dal testimone prima, durante

e dopo la deportazione. Per quanto riguarda il modulo *Voci dall'Inferno Verse Similarity Search*, progettato per rilevare citazioni e allusioni al lessico dantesco mediante tecniche computazionali, l'applicazione è stata sviluppata e testata con risultati promettenti. In particolare, ha dimostrato un'elevata accuratezza nell'individuazione delle citazioni esplicite (il 100% delle citazioni sono state riconosciute), ovvero quelle che riproducono fedelmente, o con minime variazioni, il testo dantesco. Tuttavia, il riconoscimento delle citazioni implicite, più sfumate e meno direttamente riconducibili al testo originale, risulta ancora poco efficace. È attualmente in corso una fase preliminare di fine-tuning del modello, con l'obiettivo di potenziare la capacità del sistema nel rilevamento automatico delle occorrenze intertestuali all'interno del corpus.

Infine, il progetto mira a consolidare collaborazioni scientifiche con enti e gruppi di ricerca nazionali e internazionali, tra cui il progetto *Voices from Ravensbrück*, coordinato dalla prof.ssa Silvia Calamai[29][30][31], il Centro di Documentazione Ebraica Contemporanea (CDEC), l'infrastruttura europea European Holocaust Research Infrastructure (EHRI), nonché il progetto H2IOSC, con particolare riferimento all'infrastruttura di ricerca CLARIN, attraverso il consorzio italiano CLARIN-IT.

5. Ringraziamenti

Il progetto procede grazie al lavoro di numerosi laureandi, principalmente dei Corsi di Laurea in Informatica Umanistica e in Italianistica, e si è avvalso del supporto del Dipartimento di Filologia, Letteratura e Linguistica dell'Università di Pisa, del Centro di Documentazione Ebraica Contemporanea (CDEC),³² del Centro Interdipartimentale di Studi Ebraici (CISE),³³ dell'infrastruttura di ricerca CLARIN-IT,³⁴ del laboratorio CoPhiLab,³⁵ dell'Istituto di Linguistica Computazionale del CNR di Pisa e del CLARIN Knowledge Centre DiPText-KC,³⁶ *Voci dall'Inferno* fa infine parte dell'European Holocaust Research Infrastructure (EHRI).³⁷

References

- [1] Agosti, Giannantonio. 1960. *Nei lager vinse la bontà: memorie dell'internamento nei campi di eliminazione tedeschi. Nostre missioni. Missioni estere dei Cappuccini.*
- [2] Alighieri, Dante. 1980. *La Divina Commedia. Inferno.* A cura di Umberto Bosco e Giovanni Reggio. Firenze: Le Monnier.

³² Cfr. <https://www.cdec.it/>

³³ Cfr. <https://www.cise.unipi.it/>

³⁴ Cfr. <https://www.clarin-it.it/it>

³⁵ Cfr. <https://cophilab.ilc.cnr.it/>

³⁶ Cfr. <https://diptext-kc.clarin-it.it/>

³⁷ Cfr. <https://www.ehri-project.eu/>

- [3] Artom, Emanuele. 2008. *Diari di un partigiano ebreo, gennaio 1940-febbraio 1944*. A cura di Guri Schwarz. Torino: Bollati Boringhieri.
- [4] Baiolo, Chiara. 2024. «La codifica del Diario partigiano di Emanuele Artom: sulle tracce di una biblioteca da studiare». Tesi Triennale, Pisa: Università di Pisa.
- [5] Calderini, Sara. 2015. «L'ineffabilità della nefandezza». Università di Pisa.
- [6] Calderini, Sara, e Marina Riccucci. 2020. «L'ineffabilità della nefandezza : Dante “per dire” il Lager : un sondaggio preliminare nelle testimonianze non letterarie». *Italianistica : rivista di letteratura italiana : XLIX, 1, 2020*. <https://doi.org/10.19272/202001301011>.
- [7] Cao, Hongliu. 2024. «Recent Advances in Text Embedding: A Comprehensive Review of Top-Performing Methods on the MTEB Benchmark». arXiv.Org. 27 maggio 2024. <https://arxiv.org/abs/2406.01607v2>.
- [8] Giuntini, Luigi. 2014. *I lunghi giorni della pena. Diario di prigionia (8 settembre 1943 – 15 aprile 1945)*. A cura di Gabriel Francesco Gabrielli. Vol. II. Pisa University Press.
- [9] Giuntini, Luigi. 2021. *I lunghi giorni della pena: diario di prigionia (8 settembre 1943-15 aprile 1945)*. A cura di Gabriel Francesco Gabrielli. Vol. III. Edizioni dell'Assemblea Memorie. Firenze: Regione Toscana, Consiglio regionale.
- [10] «I campi di concentramento nazisti». s.d. Consultato 24 gennaio 2025. <https://encyclopedia.ushmm.org/content/it/article/nazi-camps>.
- [11] Lelli, Diletta. 2023. «Idek Wolfowicz: un testimone ritrovato». Tesi Triennale, Pisa: Università di Pisa.
- [12] Levi, Primo. 1997. *Conversazioni e interviste (1963-1987)*. A cura di Marco Belpoliti. Einaudi.
- [13] Mercatanti, Elvira. 2024. «Le voci dei sopravvissuti ai Lager: l'applicazione web per l'analisi, per l'interrogazione e per la fruizione delle testimonianze». Tesi Magistrale, Pisa: Università di Pisa.
- [14] Montera, Greta. 2022. «Stralcio sulla deportazione in Germania: la testimonianza di Bruno Cimoli. Codifica e analisi di una sezione del testo». Tesi Triennale, Pisa: Università di Pisa.
- [15] Petrilli, Irene. 2022. «Dante fra gli Internati Militari Italiani: il caso 'Luigi Giuntini'. Codifica preliminare del suo diario I lunghi giorni della pena». Tesi Triennale, Pisa: Università di Pisa.
- [16] Reimers, Nils, e Iryna Gurevych. 2019. «Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks». arXiv. <https://doi.org/10.48550/arXiv.1908.10084>.
- [17] Ricci, Pietro. 2022. «Mio nonno deportato ad Armensterin, il suo diario inedito». Tesi Triennale, Pisa: Università di Pisa.
- [18] Riccucci, Marina, Angelo Mario Del Grosso, Frida Valecchi, e Giulia Causarano. 2021. «Testimoniare il Lager: l'informatica al servizio della memoria». In *AIUCD*

- 2021 - *Book of Extended Abstracts*, 264. Pisa: Quaderni di Umanistica Digitale.
<https://doi.org/10.6092/unibo/amsacta/6712>.
- [19] Riccucci, Marina, e Laura Ricotti. 2021. Il dovere della parola. La Shoah nelle testimonianze di Liliana Segre e di Goti Herskovitz Bauer. Pisa: Pacini Editore.
- [20] Schmidt, Thomas. 2011. «A TEI-Based Approach to Standardising Spoken Language Transcription». *Journal of the Text Encoding Initiative*, fasc. Issue 1 (giugno).
<https://doi.org/10.4000/jtei.142>.
- [21] Segre, Anna, e Gloria Pavoncello, a c. di. 2009. *Judenrampe. Gli ultimi testimoni*. Antidoti. Roma: Elliot.
- [22] Seidita, Sara. 2024. «Goti Herskovitz Bauer testimone della Shoah». Tesi Triennale, Pisa: Università di Pisa.
- [23] Siragusa, Federico. 2024. «I luoghi danteschi della memoria dei sopravvissuti». Tesi Magistrale, Pisa: Università di Pisa.
- [24] Sustersic, Federica. 2016. «La Dicibilità Del Male. La Ricezione Dantesca Nelle Testimonianze Concentrazionarie». *Dante: Rivista internazionale di studi su Dante Alighieri* 13:57–78.
- [25] Taterka, Thomas. 2002. *Dante Deutsch. Studi sulla letteratura dei lager*. A cura di Enrico Paventi. Sette Città.
- [26] Traversi, Valeria M. M. 2008. «Per Dire L'orrore: Primo Levi E Dante». *Dante: Rivista internazionale di studi su Dante Alighieri* 5:109–25.
- [27] Walser, Martin. 2015. *Unser Auschwitz: Auseinandersetzung mit der deutschen Schuld*. Rowohlt Verlag GmbH.
- [28] Alighieri, Dante. 1994. *La Commedia secondo l'antica vulgata*. A cura di Giorgio Petrocchi. *La Commedia secondo l'antica vulgata*. 2. ristampa riveduta. Le opere di Dante Alighieri 7. Firenze: Le lettere.
- [29] Calamai, Silvia, Jeannine Beeken, Henk Van Den Heuvel, Max Broekhuizen, Arjan van Hessen, Christoph Draxler, e Stefania Scagliola. 2021. «Voices from Ravensbrück. Towards the Creation of an Oral and Multilingual Resource Family». In *Proceedings of CLARIN Annual Conference 2021*, 16-19. CLARIN ERIC.
- [30] Calamai, Silvia, Chiara Kolletzek, e Aleksei Kelli. 2018. «Towards a protocol for the curation and dissemination of vulnerable people archives». In *Selected papers from the CLARIN Annual Conference 2018, Pisa, Italy, October 8-10, 2018*, a cura di Inguna Skadina e Maria Eskevich, 159:28–38. Linköping Electronic Conference Proceedings. Linköping University Electronic Press.
- [31] Calamai, Silvia, Stefania Scagliola, Fabio Ardolino, Christoph Draxler, Arjan van Hessen, e Henk van den Heuvel. 2022. «Ravensbrück Interviews: How to Curate Legacy Data to Make it CLARIN Compliant». In *Selected Papers from the CLARIN Annual Conference 2021, virtual event, September 27-29, 2021*, a cura di Monica Monachini e Maria Eskevich, 189:1–9. Linköping Electronic Conference Proceedings. Linköping University Electronic Press.