

UNIVERSITÀ DI PISA

Corso di Laurea Magistrale in Informatica Umanistica

**Le voci dei sopravvissuti ai Lager:
l'applicazione web per l'analisi, per
l'interrogazione e per la fruizione delle
testimonianze**

Relatore:

Prof. ssa Marina Riccucci

Secondo relatore:

**Prof. Angelo Mario Del
Grosso**

Candidato:

Elvira Mercatanti

Indice

Introduzione	6
1 Dante per dire il Lager	9
1.1 Il progetto <i>Voci dall'Inferno</i>	12
1.2 Definizione dell'archivio digitale	15
1.3 Il repertorio delle testimonianze del corpus	16
1.4 Definizione di una tassonomia	18
1.5 I numeri di <i>Voci dall'Inferno</i>	19
1.6 Dante nelle testimonianze di <i>Voci dall'Inferno</i>	20
2 La codifica XML delle testimonianze	24
2.1 La codifica informatica di un testo	24
2.2 Gli obiettivi della codifica	26
2.3 Codifica di fonti scritte	28
2.4 Codifica di fonti orali	32
3 Armonizzazione del corpus	36
3.1 Interventi di revisione e di uniformazione delle codifiche	37
3.2 Interventi per identificare i dati principali di una testimonianza	38
3.3 Marcatura delle entità nominate e dei riferimenti bibliografici	39
3.4 Codifica delle tessere dantesche	42
3.4.1 Codifica delle citazioni	42
3.4.2 Codifica delle allusioni	44
3.4.3 Codifica dei vocaboli danteschi	45
4 Creazione di un'applicazione web	46
4.1 eXist-db	47
4.2 L'applicazione web <i>Voci dall'Inferno</i>	49
4.2.1 <i>Automatic speech recognition (ASR)</i> e funzionalità di <i>search</i> all'interno del corpus	68
Conclusioni	71
Bibliografia	73

Elenco delle figure

1.1	Tassonomia di <i>Voci dall'Inferno</i>	19
2.1	Esempio struttura <timeline> relativa all'elemento <abstract>	34
3.1	Identificazione tipologia di testimonianza per le fonti scritte e per le fonti orali	38
4.1	Diagramma HTML Templating Framework	47
4.2	Homepage dell'applicazione web	50
4.3	Ricerca di un testimone mediante il form	51
4.4	Classificazione delle testimonianze all'interno dell'archivio digitale . .	51
4.5	Elenco dei testimoni che costituiscono l'archivio digitale	52
4.6	Scheda anagrafica del testimone ed elenco delle relative testimonianze consultabili	53
4.7	Informazioni sulla testimonianza consultata	53
4.8	Contenuto dei segmenti di sintesi per una testimonianza orale	54
4.9	Trascrizione di una testimonianza	55
4.10	Trascrizione di una testimonianza con legenda attiva	56
4.11	Trascrizione testimonianza con attivazione solo di alcuni fenomeni .	57
4.12	Grafico che mostra i fenomeni presenti in una testimonianza	57
4.13	Grafico che analizza la modalità di espressione del testimone	58
4.14	Grafico delle entità nominate presenti (testimonianza di Emanuele Artom)	59
4.15	Tabella relativa alle persone menzionate (testimonianza di Emanuele Artom)	59
4.16	Tabella relativa ai riferimenti bibliografici (testimonianza di Emanuele Artom)	60
4.17	Visualizzazione delle tessere dantesche (testimonianza di Emanuele Artom)	60
4.18	Totale tessere dantesche rilevate nel corpus	61
4.19	Ripartizione citazioni per cantiche	62
4.20	Ripartizione citazioni per canti dell'Inferno	62
4.21	Vocaboli danteschi nelle testimonianze di <i>Voci dall'Inferno</i>	63
4.22	Testimoni che attingono maggiormente dalla <i>Commedia</i>	63

4.23 Chi, in che modo e in che misura si serve del vocabolario dantesco	64
4.24 Tabelle relative alle citazioni	65
4.25 Testimoni che alludono alla <i>Commedia</i>	66
4.26 Tabella delle allusioni	66
4.27 Testimoni che utilizzano vocaboli danteschi	67
4.28 Tabella dei vocaboli danteschi	67
4.29 <i>Treegraph chart</i> che illustra la tassonomia completa dell'archivio	68
4.30 Tipologie di testimonianze del corpus	68
4.31 Provenienza delle fonti orali e scritte	69
4.32 Trascrizione automatica di un file audio	69

Listings

2.1	Struttura principale del documento XML	27
2.2	Schema di codifica di una fonte scritta	29
2.3	Esempio struttura <facsimile> nella codifica del diario di Emanuele Artom	30
2.4	Esempio struttura <text> codifica fonte scritta	31
2.5	Schema di codifica di una fonte orale	32
2.6	Esempio codifica della trascrizione di una fonte orale all'interno di <text>	35
3.1	Contenuto suddiviso in base al giorno (Diario Emanuele Artom)	36
3.2	Contenuto suddiviso in base al numero di pagina (Diario Nicola Ricci)	36
3.3	Contenuto suddiviso in base al numero di pagina e al giorno (Diario di Bruno Cimoli)	37
3.4	Definizione tipologia di fonte scritta	38
3.5	Definizione tipologia di fonte orale	38
3.6	Struttura <standOff>	39
3.7	Struttura <listPerson>	40
3.8	Struttura <listBibl>	41
3.9	Codifica di una citazione esplicita (dal Diario di Luigi Giuntini)	42
3.10	Codifica di una citazione implicita (dal Diario di Emanuele Artom)	43
3.11	Meccanismo di collegamento di una citazione con il relativo riferimento bibliografico in <listBibl>: codifica della citazione nel testo	43
3.12	Meccanismo di collegamento di una citazione con il relativo riferimento bibliografico in <listBibl>: codifica del riferimento bibliografico	43
3.13	Codifica di una allusione nella testimonianza di Goti Bauer	44
3.14	Codifica di un termine dantesco	45
3.15	Lista dei termini danteschi presenti nella testimonianza	45
4.1	Funzione XQuery per la creazione delle liste dei testimoni in base alla collezione	48

Introduzione

Il presente elaborato si colloca nell'ambito del progetto di ricerca *Voci dall'Inferno*, dal 2016 diretto e coordinato dalla professoressa Marina Riccucci del Dipartimento di Filologia, Letteratura e Linguistica dell'Università di Pisa, con il supporto del professor Angelo Mario Del Grossi dell'Istituto di Linguistica Computazionale "A.Zampolli" del CNR di Pisa. Il progetto nasce nell'anno accademico 2015-2016 in occasione del lavoro di tesi della dottoressa Sara Calderini, 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*, ovvero la tipologia testuale attraverso cui chi ha vissuto il campo di sterminio ne ha riferito in forme che si collocano nello spazio compreso tra il resoconto orale (l'intervista) e quello scritto (il diario, il racconto autobiografico/memoriale, la lettera).¹ L'indagine parte dal fatto che moltissimi testimoni sopravvissuti ai Lager, indipendentemente dal loro livello di formazione scolastica e culturale, hanno trovato nel lessico dantesco della *Divina Commedia*, e in particolar modo dell'*Inferno*, un vocabolario da cui attingere per riuscire a raccontare ciò che non sarebbe stato possibile dire in altro modo.

Il progetto si pone quindi due obiettivi, integrati e correlati:

1. la digitalizzazione e la codifica del primo corpus digitale di testimonianze *non letterarie* di sopravvissuti ai Lager;
2. l'individuazione, la quantificazione e la valutazione della presenza di lessico e immagini dantesche all'interno di quelle testimonianze.

La realizzazione del corpus digitale consente in primo luogo di raccogliere, conservare e tutelare la tipologia delle testimonianze *non letterarie*, importante patrimonio storico che a differenza delle testimonianze letterarie non è salvaguardato dalla pubblicazione e dalle ristampe, e in secondo luogo di avvalersi di strumenti e metodi informatici per poter svolgere attività di studio, analisi e ricerca su di esse.

¹Cfr. Sara Calderini, *L'ineffabilità della nefandezza* (elaborato finale: a. a. 2015-2016, Università di Pisa, Corso di Laurea in Letterature europee per l'editoria e la produzione culturale. Primo relatore: Prof. ssa Marina Riccucci; secondo relatore: Prof. ssa Serena Grazzini).

Il progetto procede grazie al lavoro di numerosi laureandi, principalmente dei Corsi di Laurea in Informatica Umanistica e in Italianistica, e al supporto del Dipartimento di Filologia, Letteratura e Linguistica dell’Università di Pisa, del Centro di Documentazione Ebraica Contemporanea (CDEC),² del Centro Interdipartimentale di Studi Ebraici (CISE),³ del laboratorio CoPhiLab,⁴ dell’Istituto di Linguistica Computazionale del CNR di Pisa e del centro di conoscenza CLARIN-IT DiPText-KC.⁵

Il 10 dicembre 2021, nell’Aula Magna Nuova della Sapienza, a Pisa, si è tenuto il primo convegno di presentazione del progetto di ricerca *Voci dall’Inferno*.⁶ Il 18 e 19 aprile 2024, durante il Convegno Internazionale *Voci dall’Inferno: Dante nelle testimonianze dei sopravvissuti ai Lager*, tenutosi presso l’Aula Magna del Polo Fibonacci di Pisa, sono stati presentati i risultati raggiunti nei primi anni di lavoro.⁷

Il lavoro di tesi che illustro nel presente elaborato, si è svolto in due fasi: la prima fase è consistita nella revisione e nell’armonizzazione dell’intero corpus di testimonianze codificate in formato XML-TEI⁸ dagli studenti che hanno collaborato al progetto nel corso degli anni, in modo tale da ottenere una struttura omogenea per tutti i dati che compongono l’archivio; una volta ottenuto un corpus uniforme, la seconda fase è consistita nella realizzazione dell’applicazione web *Voci dall’Inferno* che permette di raccogliere, di consultare, di analizzare e di interrogare tutte le testimonianze che costituiscono l’archivio. L’applicazione permette inoltre di censire la presenza di lessico dantesco all’interno del corpus e di analizzare le differenti tipologie nelle quali si presenta nelle testimonianze.

L’applicazione è stata sviluppata in ambiente eXist-db,⁹ database NoSQL open source progettato per la gestione e per l’interrogazione di collezioni di documenti XML, tramite funzioni implementate mediante il linguaggio di programmazione XQuery.¹⁰

Preciso che alcune funzionalità integrate nell’attuale applicazione sono state sviluppate in collaborazione con la dottoressa Chiara Baiolo in occasione del suo lavoro di tesi per il conseguimento della Laurea Triennale in Informatica Umanistica.¹¹

²Cfr. <https://www.cdec.it/>.

³Cfr. <https://www.cise.unipi.it/>.

⁴Cfr. <https://cophilab.ilc.cnr.it/>.

⁵Cfr. <https://diptext-kc.clarin-it.it/>.

⁶Cfr. <https://www.unipi.it/index.php/unipieventi/event/6081-voci-dall-inferno-le-testimonianze-dei-sopravvissuti-ai-lager>.

⁷Cfr. <https://www.unipi.it/index.php/unipieventi/event/7932-voci-dall-inferno-dante-nelle-testimonianze-dei-sopravvissuti-ai-lager>.

⁸Cfr. <https://tei-c.org/>.

⁹Cfr. <https://www.exist-db.org/exist/apps/homepage/index.html>.

¹⁰Cfr. <https://www.w3.org/TR/xquery-31>.

¹¹Cfr. Chiara Baiolo, *La codifica del Diario partigiano di Emanuele Artom: sulle tracce di una biblioteca da studiare* (elaborato finale: a.a. 2023-2024. Primo relatore: prof.ssa Marina Riccucci; correlatore: Angelo Mario Del Grosso).

L'elaborato è suddiviso in quattro capitoli. Nel capitolo 1 vengono illustrati i motivi per i quali molti sopravvissuti hanno fatto uso di Dante per dire il Lager e viene descritto il progetto *Voci dall'Inferno*, presentando le fonti primarie che costituiscono il corpus, le fasi di lavoro attraversate nel corso degli anni e le tecnologie che sono state sperimentate e utilizzate. Il capitolo 2 è dedicato alla rappresentazione digitale delle testimonianze mediante l'uso dello schema di codifica definito dalle linee guida TEI: si è prestata particolare attenzione alle differenze che presentano le macro-categorie che costituiscono il corpus e che hanno portato alla definizione di differenti modelli di codifica. Nel capitolo 3 sono descritti gli interventi che è stato necessario svolgere al fine di armonizzare l'intero corpus. Nel capitolo 4, infine, è presentata l'applicazione web *Voci dall'Inferno* e vengono descritte in dettaglio le funzionalità che essa offre, ripercorrendo sia il processo di lavoro seguito sia le scelte implementate per la sua attuale realizzazione.

Capitolo 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.¹² 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.¹³

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 rattraversare 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:

- (a) 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);
- (b) 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

¹²Cfr. <https://encyclopedia.ushmm.org/content/it/article/nazi-camps>.

¹³Primo Levi. *Conversazioni e interviste (1963-1987)*. A cura di Marco Belpoliti. Einaudi, 1997, p. 202.

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 [...].¹⁴

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.¹⁵ 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.¹⁶

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».¹⁷

¹⁴Marina Riccucci - Sara Calderini, *L'ineffabilità della nefandezza: Dante 'per dire' il Lager. Un sondaggio preliminare nelle testimonianze non letterarie*, in «Italianistica», 31 (2020), pp. 214-215.

¹⁵Federica Sustersic. *La dicibilità del male. La ricezione dantesca nelle testimonianze concentrazionarie*. Dante: Rivista internazionale di studi su Dante Alighieri, Vol. 13 (2016), pp. 57-78.

¹⁶Valeria M. M. Traversi «Per Dire L'orrore: Primo Levi E Dante». Dante: Rivista internazionale di studi su Dante Alighieri, Vol. 5 (2008), pp. 109-125.

¹⁷Federica Sustersic. *La dicibilità del male. La ricezione dantesca nelle testimonianze concentrazionarie*, p. 59.

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.¹⁸

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.¹⁹

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 dilettarmi 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.²⁰

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

¹⁸Martin Walser, *Unser Auschwitz: Auseinandersetzung mit der deutschen Schuld*. Rowohlt Verlag GmbH 2015.

¹⁹Taterka, Thomas. *Dante Deutsch. Studi sulla letteratura dei lager*. Viterbo, Sette Città 2002.

²⁰Ivi, p. 37.

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. È in questi testi che il ricorso a Dante si configura come un atto maggiormente istintivo.²¹ Il progetto nasce quindi nell'anno accademico 2015-2016, in occasione del lavoro di tesi della dottoressa Sara Calderini, 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*.²² 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, e non solo, all'interno delle testimonianze.

Come verificato da Federico Siragusa, 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.²³

²¹Marina Riccucci - Sara Calderini, *L'ineffabilità della nefandezza: Dante 'per dire' il Lager. Un sondaggio preliminare nelle testimonianze non letterarie*.

²²Cfr. Sara Calderini, *L'ineffabilità della nefandezza*.

²³Cfr. Federico Siragusa, *I luoghi danteschi della memoria dei sopravvissuti* (elaborato finale: a. a. 2023-2024, Università di Pisa, Corso di Laurea in Magistrale in Italianistica. Primo relatore: Prof. ssa Marina Riccucci; secondo relatore: Prof. ssa Leyla Livraghi).

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*.²⁴

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 sapientia e 'l primo amore.
Dinanzi a me non fuor cose create
se non etterne, e io eterno duro.
Lasciate ogne speranza, voi ch'intrate.
(*Inferno* III vv. 1-9)²⁵

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:

²⁴Cfr. Giannantonio Agosti da Romallo, *Nei Lager vinse la bontà. Memorie dell'Internamento nei campi di eliminazione tedeschi*, Milano, Missioni estere dei Cappuccini 1960.

²⁵Dante Alighieri, *La Divina Commedia. Inferno*. A cura di Umberto Bosco e Giovanni Reggio. Firenze: Le Monnier, 1980.

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*».²⁶

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?*».²⁷

Come messo in luce da Irene Petrilli, 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.²⁸

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.²⁹

²⁶Giannantonio Agosti da Romallo, *Nei lager vinse la bontà. Memorie dell'Internamento nei campi di eliminazione tedeschi*.

²⁷Luigi Giuntini, *I lunghi giorni della pena. Diario di prigionia (8 settembre 1943 – 15 aprile 1945)*, a cura di Gabriel Francesco Gabrielli, vol. II, p. 242, 2021.

²⁸Cfr. Irene Petrilli, *Dante fra gli Internati Militari Italiani: il caso 'Luigi Giuntini'. Codifica preliminare del suo diario "I lunghi giorni della pena"* (elaborato finale: a.a. 2021-2022. Primo relatore: prof.ssa Marina Riccucci; correlatore: Angelo Mario Del Grosso).

²⁹Marina Riccucci - Laura Ricotti, *Il dovere della parola. La Shoah nelle testimonianze di Liliana Segre e di Goti Herskovitz Bauer*, Pisa, Pacini Editore 2021.

1.2 Definizione 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*,³⁰
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 del contenuto testuale della testimonianza. In aggiunta, l'ambiente web permetteva anche il confronto del lessico della testimonianza con il testo della *Commedia* dantesca, 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, i quali si sono occupati della trascrizione e della marcatura seguendo le indicazioni fornite dalle linee guida del consorzio TEI.³¹

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:

- (a) per il primo approccio, le applicazioni web sono state sviluppate sfruttando le funzionalità di una libreria client-side per l'elaborazione di documenti XML: SaxonJS;³²
- (b) per il secondo, quello attuale, le applicazioni web sono state sviluppate adottando le tecnologie messe a disposizione dall'ambiente eXist-db³³ mediante l'uso del modulo *HTML-templating* (server-side). EXist-db è un software open source progettato per database NoSQL basati su XML. La piattaforma integra un modulo dedicato alla generazione dinamica di pagine HTML, partendo da collezioni di documenti XML e da funzioni implementate mediante il linguaggio di

³⁰Cfr. Marina Riccucci, Angelo Mario Del Grosso, Frida Valecchi e Giulia Causarano, *Testimoniare il lager: l'informatica al servizio della memoria*, Quaderni di Umanistica Digitale, Pisa 2021.

³¹Cfr. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

³²Cfr. <https://www.saxonica.com/html/welcome/welcome.html>.

³³Cfr. <https://exist-db.org/exist/apps/homepage/index.html>.

interrogazione XQuery. Il funzionamento prevede l'uso di template in HTML, in cui si aggiungono chiamate a funzioni XQuery che, andando a navigare i nodi XML, permettono l'estrazione dell'informazione relativa a quel nodo e la generazione di frammenti HTML che vanno a completare la pagina.

Fino ad oggi sono state implementate e proposte alcune versioni di applicazione web da diversi studenti durante i loro lavori di tesi triennale. Le applicazioni realizzate consentivano molteplici funzionalità ma erano state sviluppate prendendo in analisi uno o pochi testimoni del corpus. L'applicazione web *Voci dall'Inferno*, presentata e descritta in questo elaborato, è stata realizzata aprendo il proprio bacino di utenza all'intero archivio digitale. Essa consente infatti la consultazione, l'analisi e la fruizione di tutte le tipologie di testimonianze che costituiscono il corpus. L'applicazione verrà ampiamente descritta in dettaglio all'interno del capitolo 4.

1.3 Il repertorio delle testimonianze del corpus

L'archivio digitale *Voci dall'Inferno* è attualmente costituito da 23 testimonianze appartenenti a 18 testimoni. Le testimonianze sono state rese da sopravvissuti con storie e provenienze diverse e attraverso differenti modalità, il che ha contribuito a rendere il corpus molto eterogeneo, offrendo allo stesso tempo una visione molto ampia 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.

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 di ridisegnamento 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).³⁴

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*.³⁵ 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 triennale: è stata questa ricerca a rivelare l'assenza di Idek all'interno dell'archivio della Fondazione CDEC.³⁶

Degna di menzione è anche la storia del diario di prigione di Nicola Ricci, un internato militare italiano. Ricci fu catturato pochi chilometri dopo il confine franco-italiano dall'esercito nazista e fu inizialmente deportato come I.M.I. in un campo di smistamento per prigionieri nel sud della Germania e, successivamente, ad Armentstein, 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. Il diario inedito è stato oggetto della tesi di laurea da parte del nipote di Nicola: Pietro Ricci.³⁷

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

³⁴Cfr. Greta Montera, *Stralcio sulla deportazione in Germania: la testimonianza di Bruno Cimoli. Codifica e analisi di una sezione del testo* (elaborato finale: a.a. 2021-2022. Primo relatore: prof. ssa Marina Riccucci; secondo relatore: prof. Angelo Mario Del Grosso).

³⁵Cfr. Anna Segre and Gloria Pavoncello. *Judenrampe. Gli ultimi testimoni*. Antidoti. Elliot, 2009.

³⁶Cfr. Diletta Lelli, *Idek Wolfowicz: un testimone ritrovato* (elaborato finale: a.a. 2022-2023. Primo relatore: prof.ssa Marina Riccucci; correlatore: prof. Angelo Mario Del Grosso).

³⁷Cfr. Pietro Ricci: *Mio nonno deportato d'Armenstein: il suo diario inedito* (elaborato finale: a.a. 2021-2022. Primo relatore: prof.ssa Marina Riccucci; secondo relatore: prof. Angelo Mario Del Grosso).

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.³⁸

Infine, arricchiscono l’archivio anche le testimonianze di due prigionieri che furono obbligati a collaborare con le SS³⁹ svolgendo il ruolo di *Sonderkommando*⁴⁰, 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 quello 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*, 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, come verrà spiegato nel capitolo 4.

³⁸Cfr. Chiara Baiolo, *La codifica del Diario partigiano di Emanuele Artom: sulle tracce di una biblioteca da studiare* (elaborato finale: a.a. 2023-2024. Primo relatore: prof.ssa Marina Riccucci; correlatore: prof. Angelo Mario Del Grosso).

³⁹Cfr. <https://www.treccani.it/enciclopedia/ss/>.

⁴⁰Cfr. <https://www.treccani.it/vocabolario/sonderkommando/>.

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, Samuele 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.

In figura 1.1 è riportata graficamente la tassonomia dei testimoni.

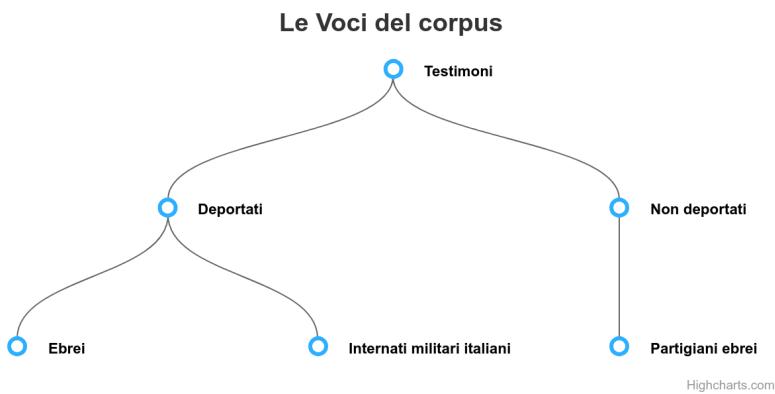


Figura 1.1: Tassonomia di *Voci dall'Inferno*

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 dottesse Anna Segre e Gloria Pavon-

cello negli anni 2005-2006.⁴¹ 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,⁴² 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.⁴³ La testimonianza di Herskovitz Bauer è stata realizzata il 7 marzo 2020 da Marina Riccucci e da Laura Ricotti.⁴⁴ La testimonianza di Wachsberger appartiene alla Fondazione CDEC ed è stata realizzata a Milano da Liliana Picciotto nell'ambito del progetto "Ricerca sulla deportazione".

Le tre testimonianze orali in formato video provengono tutte da YouTube: si tratta delle interviste di Enrico Vanzini, Samuele Modiano e Arminio Wachsberger.⁴⁵

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 prigione. Delle rimanenti 3 testimonianze scritte, le testimonianze di Emanuele Artom e di Romana Feld appartengono dall'archivio digitale del Centro di Documentazione Ebraica Contemporanea di Milano⁴⁶, mentre la testimonianza di Bruno Cimoli proviene dalla sede centrale dell'ANPI di Massa.⁴⁷

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. Come verrà spiegato in dettaglio nella sezione 3.4, durante la fase di codifica si sono stabilite diverse modalità di marcatura che permettessero di esplicitare e di distinguere op-

⁴¹Cfr. Anna Segre and Gloria Pavoncello. *Judenrampe. Gli ultimi testimoni*. Antidot. Elliot, 2009

⁴²Di Liliana Segre il corpus dispone di due interviste realizzate da Segre e Pavoncello: una del 2006 e una del 2007.

⁴³Di Wachsberger sono state digitalizzate due testimonianze: una del 1987 e un del 1998. L'intervista in questione è quella datata 1998.

⁴⁴Cfr. Marina Riccucci - Laura Ricotti, *Il dovere della parola. La Shoah nelle testimonianze di Liliana Segre e di Goti Herskovitz Bauer*.

⁴⁵L'intervista in questione è quella datata 1987.

⁴⁶Cfr. <https://digital-library.cdec.it/cdec-web/>.

⁴⁷Cfr. <https://www.anpi.it/sedi?titolo=massa&area=>.

portunamente 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,⁴⁸ 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.

Tra i testimoni che ricorrono maggiormente a citazioni dantesche, c'è Luigi Giuntini, che nel suo diario⁴⁹ 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.⁵⁰

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. 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 pieta*», «*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

⁴⁸Cfr. Federico Siragusa, *I luoghi danteschi della memoria dei sopravvissuti*.

⁴⁹Cfr. Gabriel Francesco Gabrielli, *I lunghi giorni della pena. Il diario di prigonia di Luigi Giuntini (settembre 1943 – aprile 1945)*. Pisa University Press, Pisa, 2014.

⁵⁰Cfr. Irene Petrilli, *Dante fra gli Internati Militari Italiani: il caso ‘Luigi Giuntini’*. Codifica preliminare del suo diario “I lunghi giorni della pena”.

*a lui “qual tu sii, od ombra od omo certo?”» [...]*⁵¹.

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.⁵²

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.⁵³

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 [...].⁵⁴

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.⁵⁵

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

⁵¹Gabriel Francesco Gabrielli, *I lunghi giorni della pena. Il diario di prigionia di Luigi Giuntini (settembre 1943 - aprile 1945)*, cit., voll. II., p. 242.

⁵²Luigi Giuntini, *I lunghi giorni della pena*, cit., voll. II., p. 437.

⁵³Luigi Giuntini, *I lunghi giorni della pena*, cit., vol. III., p. 365.

⁵⁴G. Schwarz, Emanuele Artom, *Diario di un partigiano ebreo, gennaio 1940-febbraio 1944*, Bollati Boringhieri, Torino, 2008, p. 95.

⁵⁵Marina Riccucci - Laura Ricotti, *Il dovere della parola. La Shoah nelle testimonianze di Liliana Segre e di Goti Herskovitz Bauer*.

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.⁵⁶

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.

⁵⁶Pietro Ricci: *Mio nonno deportato d'Armenstein: il suo diario inedito*.

Capitolo 2

La codifica *XML* delle testimonianze

In questo capitolo verrà descritta l'attività di codifica che è stata svolta per la digitalizzazione delle testimonianze del corpus. In particolare, verrà brevemente introdotto il linguaggio di marcatura XML-TEI e successivamente verranno spiegate le differenze semantiche e strutturali che una fonte scritta presenta rispetto ad una fonte orale con la conseguente codifica che ne rappresenta il modello.

2.1 La codifica informatica di un testo

Gli studi umanistici hanno come oggetto privilegiato di studio i testi, intorno ai quali ciascuna disciplina ha sviluppato nel corso dei secoli un insieme di procedimenti di rappresentazione e di analisi. La diffusione delle tecnologie informatiche ha portato ad una rivoluzione degli strumenti per la produzione e trasmissione del sapere. Il testo, trasferito su un supporto elettronico, diventa testo elettronico o testo digitale, portando, di conseguenza, all'incontro tra ambiti disciplinari che in passato erano ritenuti incompatibili: il mondo della ricerca umanistica e quello della ricerca tecnologica.

Un settore nel quale questa interazione si è resa particolarmente evidente è lo studio dei problemi connessi alla codifica informatica dei testi. Per codifica informatica dei testi intendiamo la rappresentazione formale di un testo, ad un qualche livello descrittivo, su di un supporto digitale, in un formato utilizzabile da un elaboratore (*Machine Readable Form*), mediante un opportuno linguaggio informatico.⁵⁷ Si tratta quindi di un processo di rappresentazione che implica una serie di operazioni di selezione e di classificazione degli elementi rilevanti in funzione di un determinato punto di vista.

Sin dagli anni '80 nel campo dell'informatica umanistica si è avvertita l'esigenza di definire uno standard per la rappresentazione di testi su supporto digitale, definendo ed implementando un linguaggio che deve essere trattato da un elaboratore e allo stesso tempo deve anche essere sufficientemente espressivo per rappresentare la

⁵⁷Cfr. Fabio Ciotti. *Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari*. Aracne, 2007.

complessità di un testo. Una struttura di questo tipo è definita *schema di codifica*. Nel 1988 è stato quindi avviato un progetto internazionale chiamato *Text Encoding Initiative* (TEI) con l'obiettivo di definire e sviluppare uno schema di codifica il più possibile condiviso dalla comunità degli studiosi e sufficientemente espressivo per soddisfare le numerose e diversificate necessità di rappresentazione per molteplici tipologie di testualità. Per conseguire questo obiettivo fu inizialmente individuato come linguaggio di base un linguaggio di marcatura standard di tipo dichiarativo, chiamato *Standard Generalized Markup Language* (SGML).⁵⁸

Un linguaggio di marcatura (*Markup Language*) può essere definito come un insieme di convenzioni attraverso le quali vengono descritti i meccanismi di rappresentazione di una risorsa. I linguaggi di marcatura possono essere divisi in due tipologie:

- Linguaggi procedurali (*specific markup languages*), che specificano le operazioni che un certo sistema deve compiere sul testo per ottenere un determinato output, e sono quindi orientati alla creazione di un documento che ha una certa impaginazione e impostazione grafica
- Linguaggi dichiarativi (*generic markup languages*), che sono invece orientati alla rappresentazione delle caratteristiche o degli elementi che costituiscono un testo

A partire dal 1989 la TEI ha sviluppato diverse versioni del suo schema di codifica, le cui specifiche provvisorie sono state pubblicate per la prima volta nel 1991 con il titolo *Guidelines for Electronic Text Encoding and Interchange, TEI P1*. Gli obiettivi della TEI erano due: stabilire quali caratteristiche testuali potevano essere codificate in un testo elettronico e in che modo. Negli anni successivi le versioni sono state riviste portando alla pubblicazione di manuali aggiornati. Nel 2000 i membri della TEI hanno deciso di rendere permanente il progetto, costituendo il TEI *Consortium*, un'organizzazione internazionale senza scopo di lucro fondata con lo scopo di sostenere e sviluppare lo schema di codifica TEI.

Sebbene SGML fosse un linguaggio molto complesso, fu adottato da numerose comunità per sviluppare linguaggi di markup. Proprio da SGML deriva il noto *Hyper Text Markup Language* (HTML)⁵⁹, linguaggio con cui codificare i documenti da pubblicare nel Web, sviluppato da Tim Berners Lee e rilasciato nella sua prima versione nel 1993. A seguito della forte diffusione di tale linguaggio grazie ai primi utilizzi commerciali del web, ci fu l'esigenza di continuare a sviluppare il formato HTML. Per fare questo, nel 1994 fu fondato il *World Wide Web Consortium* (W3C),⁶⁰ un'organizzazione non governativa internazionale la cui principale attività consiste nello stabilire standard tecnici per il World Wide Web inerenti sia i linguaggi di markup che i protocolli di comunicazione.

⁵⁸Cfr. https://it.wikipedia.org/wiki/Standard_Generalized_Markup_Language.

⁵⁹Cfr. <https://it.wikipedia.org/wiki/HTML>.

⁶⁰Cfr. <https://www.w3.org/>.

La necessità di potenziare la capacità di gestione dei documenti per il Web, ha portato, nel 1996, alla costituzione di un gruppo di lavoro denominato *XML Working Group* nell'ambito del W3C. Ebbe quindi inizio lo sviluppo di un nuovo linguaggio di markup, chiamato *Extensible Markup Language* (XML), di cui nel 1998 sono state pubblicate le specifiche ufficiali.⁶¹

Le *Guidelines TEI* definiscono un linguaggio per descrivere la struttura, la semantica e la funzionalità di un testo, e propongono delle categorie nominate per descrivere e classificare i suoi componenti. Questo linguaggio è espresso nella sintassi XML ed è definito da grammatiche formali, quali la *Document Type Definition* (DTD), oppure dai più attuali schemi *RELAX NG*⁶² oppure *schematron*.⁶³

2.2 Gli obiettivi della codifica

All'interno dell'archivio digitale le testimonianze si dividono in due macro-classi, le quali 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)⁶⁴ 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.

Un documento XML presenta una struttura gerarchica, nella quale gli elementi sono organizzati secondo una relazione di tipo padre-figlio. Un documento può essere rappresentato come un albero etichettato e ordinato, con un elemento radice (*root*) e altri elementi annidati secondo una gerarchia. Ogni nodo dell'albero può contenere quindi sotto-elementi, testo o attributi, riflettendo così le relazioni logiche e le dipendenze tra i diversi componenti del documento.

La radice del documento è <TEI>, che contiene un documento TEI-conforme, comprendente un'intestazione e una rappresentazione del testo, sia esso isolato o parte

⁶¹Cfr. Ciotti, Fabio. *Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari*.

⁶²Cfr. <https://relaxng.org/compact-tutorial-20030326.html>.

⁶³Cfr. <https://schematron.com/>.

⁶⁴I file ODD sono documenti XML-TEI utilizzati per esprimere la *customization* di qualsiasi schema TEI.

di un elemento.⁶⁵

La radice <TEI> è padre dei seguenti elementi:

- <**teiHeader**>: fornisce metadati descrittivi e dichiarativi associati ad una risorsa digitale o ad un insieme di risorse.
- <**standOff**>: contiene i dati collegati, le annotazioni e le informazioni aggiuntive incorporate nel documento.
- <**text**>: contiene un testo di qualsiasi tipo, unitario o composito.

Nel caso delle fonti scritte, come verrà spiegato in dettaglio nell'apposita sezione 2.3, <TEI> sarà genitore anche dell'elemento <**facsimile**>, che contiene la rappresentazione della fonte scritta sotto forma di un insieme di immagini piuttosto che come testo trascritto o codificato.

La struttura principale del documento XML, sia per le fonti scritte che per le fonti orali, è riportata nel listato 2.1.

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="it">
2   <teiHeader></teiHeader>
3   <standOff></standOff>
4   <text></text>
5 </TEI>
```

Listing 2.1: Struttura principale del documento XML

In entrambi gli schemi di codifica, all'interno dell'elemento <**teiHeader**>, figlio della radice <TEI>, sono state fornite tutte le informazioni descrittive e dichiarative che costituiscono un frontespizio elettronico che precede qualsiasi testo TEI-conforme.

Gli elementi figli di <**teiHeader**> sono:

- <**fileDesc**>, contiene la descrizione bibliografica completa del documento elettronico. Fornisce informazioni sull'edizione digitale (<**editionStmt**>), sul titolo (<**titleStmt**>) e sulla pubblicazione della stessa (<**publicationStmt**>), e sulla fonte primaria (<**sourceDesc**>). I figli di <**sourceDesc**> differiscono in base alla tipologia della fonte, come verrà descritto più in dettaglio nelle relative sezioni 2.3 e 2.4.
- <**encodingDesc**>, documenta la relazione fra un documento elettronico e la fonte o le fonti da cui è derivato.

⁶⁵Cfr. <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-TEI.html>.

- <profileDesc>, fornisce una descrizione dettagliata degli aspetti non bibliografici di un testo, in modo particolare le lingue e i linguaggi secondari utilizzati, la situazione in cui è stato prodotto il testo, i partecipanti e la relativa ambientazione.

L’altro figlio di <TEI>, fratello di <teiHeader>, è <standOff>, che contiene le liste delle entità nominate e di tutti i riferimenti storico-letterari presenti nella testimonianza che durante la fase di codifica sono stati identificati e marcati.

Infine, il terzo figlio di <TEI>, l’elemento <text>, rappresenta il contenuto testuale del documento. In base alla tipologia della testimonianza e alla struttura della fonte primaria, <text> può contenere tre elementi:

- <front>: contiene qualsiasi materiale preliminare (intestazioni, frontespizio, prefazioni, dediche e altro ancora) trovato all’inizio del documento, prima del corpo principale.
- <body>: contiene l’intero corpo del testo, ovvero la trascrizione vera e propria della testimonianza. L’articolazione interna della struttura dell’elemento <body> differisce in base alla tipologia della testimonianza, come descritto nelle sezioni 2.3 e 2.4.
- <back>: contiene qualsiasi appendice che segua il testo vero e proprio.

Nelle sezioni successive 2.3 e 2.4 sono descritte le principali caratteristiche che distinguono una fonte scritta da una fonte orale, i conseguenti schemi di codifica realizzati, e i relativi differenti fenomeni che in fase di codifica sono stati individuati e marcati, ricorrendo all’utilizzo di appositi elementi definiti nei moduli delle linee guida TEI.

2.3 Codifica di fonti scritte

Le testimonianze scritte seguono uno schema di edizione *image-based* di tipo diplomatico-interpretativo con approccio *parallel-transcription*⁶⁶ alla rappresentazione del testo-documento. A tale scopo sono stati utilizzati gli elementi definiti nel modulo 12 (*Representation of Primary Sources*)⁶⁷, per la trascrizione della fonte primaria; modulo 14 (*Names, Dates, People, and Places*)⁶⁸, per la rappresentazione delle entità nominate; modulo 17 (*Linking, Segmentation, and Alignment*)⁶⁹, per le analisi di particolari strutture del testo; modulo 18 (*Simple Analytic Mechanisms*)⁷⁰, per l’annotazione semantica e linguistico-lessicale delle unità testuali. Per quanto riguarda la descrizione della fonte primaria, sono stati impiegati, come norma, gli elementi

⁶⁶Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/it/html/PH.html#PH-bov>.

⁶⁷Cfr. <https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.

⁶⁸Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>.

⁶⁹Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html>.

⁷⁰Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html>.

definiti nel modulo 11 (*Manuscript Description*)⁷¹ delle linee guida TEI.

Nel listato 2.2 è riportato un esempio dello schema di codifica per le fonti scritte.

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="it">
2   <teiHeader>
3     <fileDesc>
4       <titleStmt></titleStmt>
5       <editionStmt></editionStmt>
6       <publicationStmt></publicationStmt>
7       <sourceDesc>
8         <msDesc></msDesc>
9       </sourceDesc>
10      </fileDesc>
11      <encodingDesc></encodingDesc>
12      <profileDesc>
13        <creation></creation>
14        <langUsage></langUsage>
15        <textDesc></textDesc>
16      </profileDesc>
17    </teiHeader>
18    <facsimile>
19      <surface>
20        <graphic/>
21        <zone></zone>
22      </surface>
23    </facsimile>
24    <text>
25      <front></front>
26      <body></body>
27      <back></back>
28    </text>
29    <standOff>
30      <listPerson></listPerson>
31      <listPlace></listPlace>
32      <listOrg></listOrg>
33      <listBibl></listBibl>
34    </standOff>
35 </TEI>
```

Listing 2.2: 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>`. Come mostrato nell'esempio nel listato 2.3, ciascuna pagina del documento è rappresentata da un elemento `<surface>` ed è stata

⁷¹Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>.

mappata riga per riga tramite lo strumento TEI Zoner⁷², che ha permesso di generare dei figli `<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.

```

1 <facsimile>
2   <surface xml:id="pag1" n="1"> <!-- PAG1 -->
3     <graphic url="fotodiario/pag1.jpeg" width="1024px" height="
781px"/>
4     <zone n="0" ulx="20" uly="14" lrx="1009" lry="762"
rendition="HotSpot" xml:id="pag1_10"/>
5     <zone n="1" ulx="259" uly="258" lrx="677" lry="290"
rendition="Line" xml:id="pag1_11"/>
6     <zone n="2" ulx="261" uly="297" lrx="630" lry="322"
rendition="Line" xml:id="pag1_12"/>
7   </surface>
8
9   <surface xml:id="pag2" n="2"> <!-- PAG2 -->
10    <graphic url="fotodiario/pag2.jpeg" width="1024px" height="
697px"/>
11    <zone n="0" ulx="12" uly="16" lrx="1009" lry="682"
rendition="HotSpot" xml:id="pag2_10"/>
12    <zone n="1" ulx="147" uly="113" lrx="767" lry="151"
rendition="Line" xml:id="pag2_11"/>
13    <zone n="2" ulx="143" uly="165" lrx="917" lry="190"
rendition="Line" xml:id="pag2_12"/>
14    <!-- [...] -->
15  </surface>
16  <!-- [...] -->
17 </surface>

```

Listing 2.3: Esempio struttura `<facsimile>` nella codifica del 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

⁷²Cfr. <http://teicat.huma-num.fr/zoner.php>.

beginning), i quali indicano rispettivamente l'inizio di una nuova pagina e di una nuova riga nel testo originale.

Nel listato 2.4 è riportato un esempio di una possibile articolazione interna dell'elemento `<text>`.

```

1 <text xml:lang="it">
2   <front>
3     <pb xml:id ="pre1" facs ="#pag1"/>
4     <p></p>
5     <pb xml:id ="pre2" facs ="#pag2"/>
6     <p></p>
7   </front>
8   <body >
9     <div type ="giorno" xml:id ="nov_43"></div>
10    <div type ="giorno" xml:id ="g23nov_43"></div>
11    <div type ="giorno" xml:id ="g24nov_43"></div>
12    <div type ="giorno" xml:id ="g25nov_43"></div>
13    <div type ="giorno" xml:id ="g27nov_43"></div>
14    <div type ="giorno" xml:id ="g28nov_43"></div>
15    <div type ="giorno" xml:id ="g29nov_43"></div>
16    <div type ="giorno" xml:id ="g30nov_43"></div>
17    <! -- [...] -->
18  </body>
19 </text>

```

Listing 2.4: Esempio struttura `<text>` codifica fonte scritta

Durante la fase di codifica della testimonianza, sono stati individuati e marcati tutti gli elementi che permettono una ricostruzione semanticamente più corretta possibile. 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.

2.4 Codifica di fonti orali

Gli elementi XML/TEI più significativi utilizzati per la descrizione delle fonti orali possono essere così sintetizzati: le informazioni relative al supporto sono state registrate mediante l'elemento `<recordingStmt>`, contenuto a sua volta nell'elemento `<sourceDesc>`, appartenente al modulo 8 delle linee guida TEI (*Transcriptions of Speech*)⁷³. 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 16 (*Language Corpora*)⁷⁴, in particolare l'elemento `<particDesc>` offre una descrizione accurata delle persone che hanno preso parte alla conversazione.

Il listato 2.5 riporta un esempio dello schema di codifica realizzato per le fonti orali.

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="it">
2   <teiHeader>
3     <fileDesc>
4       <titleStmt></titleStmt>
5       <editionStmt></editionStmt>
6       <publicationStmt></publicationStmt>
7       <sourceDesc>
8         <recordingStmt>
9           <recording></recording>
10          </recording>
11        </sourceDesc>
12      </fileDesc>
13      <encodingDesc></encodingDesc>
14      <profileDesc>
15        <particDesc></particDesc>
16        <langUsage></langUsage>
17        <abstract></abstract>
18      </profileDesc>
19    </teiHeader>
20    <standOff>
21      <timeline></timeline>
22    </standOff>
23    <standOff>
24      <listPerson></listPerson>
25      <listPlace></listPlace>
26      <listOrg></listOrg>
27      <listBibl></listBibl>

```

⁷³Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html>.

⁷⁴Cfr. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CC.html>.

```

28      </standOff>
29      <text>
30          <body></body>
31      </text>
32  </TEI>
```

Listing 2.5: 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>`.

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

- 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 `@sync` allo scopo di collegare il minutaggio specificato dal relativo tag `<when/>`, come mostrato in figura 2.1.
- 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>`.
- La terza timeline è stata realizzata per raggruppare gli istanti in cui avviene un cambio di parlante
- 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>`.

```

1 <standOff>
2 <!--Timeline regesto-->
3 <timeline xml:id="TL1" source=
  "#reg_1A" unit="s">
4 <when xml:id="TR1" absolute="
  00:00:00"/>
5 <when xml:id="TR2" absolute="
  00:05:20"/>
6 <when xml:id="TR3" absolute="
  00:09.05"/>
7 <when xml:id="TR4" absolute="
  00:11:46"/>
8 <when xml:id="TR5" absolute="
  00:13:57"/>
9 <when xml:id="TR6" absolute="
  00:22:32"/>
10 </timeline>
11 </standOff>
1 <abstract> <!--regesto-->
2 <ab>
3 <list>
4 <item synch="#TR1"></item>
5 <item synch="#TR2"></item>
6 <item synch="#TR3"><persName
  ref="#LS"><forename>Liliana
</forename><surname>Segre</
  surname></persName> parla
  dell'umiliazione nel Campo ,
  totale, sia dal punto di
  vista psicologico che
  fisico. [...]
7 </item>
8 <!--[...].-->
9 </list>
10 </ab>
11 </abstract>

```

Figura 2.1: Esempio struttura `<timeline>` relativa all’elemento `<abstract>`

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>`.

Il listato 2.6 mostra un esempio di codifica della trascrizione di una testimonianza

con la marcatura di alcuni fenomeni descritti pocanzi.

```
1 <u who="#MARCHERIA" xml:id="m223" synch="#tlp457">
2     In questo <supplied resp="ED">caos</supplied>, sì, perché
     arrivavano i russi <pause type="short"/> e c'era il caos. Siamo
     <del type="repetition">siamo</del> arrivati in uno spiazzo, c'
     erano sempre gli ufficiali. <del type="reformulation">ci hanno</
     del> <del type="repetition">in uno spiazzo</del> In questo
     spiazzo, era fai conto <anchor synch="#tls282"/>così, <anchor
     synch="#tls283"/> <kinesic> <desc>si intuisce che Ida stia
     mostrando la grandezza della piazza</desc> </kinesic> e qui c'
     era un tronco d'albero abbattuto e io mi sono seduta su questo
     tronco e ho detto: <q type="spoken" who="#MARCHERIA" toWhom="#
     Stella"> <emph> "Qui può venire giù Padreterno, Cristo, ma io di
     qui non mi muovo più!"</emph> </q> Avevo tutti i piedi
     insanguinati, non c'avevo più scarpe, zoccoli,<pause type="short
     "/> non c'era più nulla, <emph>cinque giorni!</emph>
3     [...]
4 </u>
```

Listing 2.6: Esempio codifica della trascrizione di una fonte orale all'interno di `<text>`

Capitolo 3

Armonizzazione del corpus

Il corpus di *Voci dall'Inferno* è quindi molto disomogeneo, per due motivi principali: innanzitutto, l'attività di codifica è stata svolta da studenti diversi durante le attività di tirocinio e tesi; inoltre, le fonti che costituiscono l'archivio digitale si presentano attraverso diverse tipologie.

Come descritto nel capitolo precedente, le testimonianze che costituiscono l'archivio sono divisibili in due macro-categorie: scritto e orale. A loro volta, però, ciascuna categoria è costituita da tipologie diverse di fonti primarie: le testimonianze orali possono presentarsi in formato audio o in formato video, mentre le testimonianze scritte principalmente in forma di diario, lettera o cartolina. Ciascuna tipologia presenta delle differenze strutturali, funzionali e semantiche che in fase di codifica devono essere identificate e opportunamente rappresentate. Non tutte le stesse tipologie di fonti primarie, però, sono articolate allo stesso modo. Infatti, nel caso dei diari, ad esempio, l'autore può aver scelto di articolare il contenuto suddividendolo in base al giorno (listato 3.1), in base al numero di pagina (listato 3.2), oppure in base ad entrambi (listato 3.3).

```
1 <text>
2     <body>
3         <div type="giorno" xml:id="nov_43">
4             <pb xml:id="p1" facs="#pag46"/>
5             <opener>
6                 <dateline></dateline>
7                 </opener>
8                 <p></p>
9             </div>
10            <!-- [...] -->
11        </body>
12    </text>
```

Listing 3.1: Contenuto suddiviso in base al giorno (Diario Emanuele Artom)

```

1 <text>
2   <body>
3     <div type="pagina" xml:id="p1">
4       <pb n="1"/>
5       <ab></ab>
6       <!-- [...] -->
7     </div>
8     <!-- [...] -->
9   </body>
10 </text>
```

Listing 3.2: Contenuto suddiviso in base al numero di pagina (Diario Nicola Ricci)

```

1 <text>
2   <body>
3     <div type="pagina" n="1">
4       <pb n="1" facs="#pagina1"/>
5       <div type="giorno" n="1">
6         <lb facs="#PPriga1" n="1"/>
7         <opener>
8           <dateline></dateline>
9         </opener>
10        <ab></ab>
11      </div>
12      <!-- [...] -->
13    </div>
14  </body>
15 </text>
```

Listing 3.3: Contenuto suddiviso in base al numero di pagina e al giorno (Diario di Bruno Cimoli)

Per poter quindi realizzare un'applicazione web che consentisse una corretta gestione, presentazione e analisi delle informazioni contenute nei dati espressi in formato XML, è stato necessario e propedeutico procedere dapprima con l'armonizzazione dell'intero corpus.

Nel presente capitolo verranno illustrati i principali interventi effettuati.

3.1 Interventi di revisione e di uniformazione delle codifiche

Come prima cosa sono state revisionate tutte le codifiche del corpus per intervenire su eventuali errori presenti e sull'uniformazione degli stessi dati espressi però in modo differente.

Un esempio di ciò, è l'attività effettuata al fine di uniformare il formato per indicare la durata temporale di una fonte orale. L'elemento `<recording>` fornisce i dettagli di una registrazione audio o video e può essere accompagnato dall'attributo `@dur` che permette di indicare la durata di una registrazione, esprimendola mediante il formato ISO 8601. Poiché non tutte le testimonianze presentavano il tempo espresso nel formato corretto, si è intervenuto sull'intero corpus di fonti orali uniformando il dato, così da poter successivamente non solo estrarre l'informazione ma anche effettuare delle operazioni come il calcolo delle ore complessive di registrazione che sono state trascritte e codificate.

3.2 Interventi per identificare i dati principali di una testimonianza

Altri interventi hanno riguardato delle specifiche integrazioni di elementi e di informazioni che permettessero di visualizzare e di consultare correttamente ciascuna tipologia di testimonianza all'interno dell'applicazione web.

Tra questi c'è stata la definizione di un criterio per individuare, censire e distinguere tra loro le varie tipologie di testimonianza, come mostrato in figura 3.1. Per quanto riguarda le fonti scritte, è stato inserito all'interno dell'elemento `<objectDesc>` l'attributo `@form`, attribuendogli un valore che permettesse di identificare il tipo di testimonianza in questione: *manuscript diary* per i diari manoscritti, *letters* per le lettere, *dossier* per i fascicoli e *manuscript sheet* per i fogli manoscritti. Per quanto riguarda le fonti orali, invece, l'identificazione della tipologia avviene sull'elemento `<recording>` mediante l'attributo `@type` con valore *video* o *audio*.

```

1 <physDesc>
2   <objectDesc form="
3     manuscript_diary">
4     <!-- [...] -->
5   </objectDesc>
6   <!-- [...] -->
7 </physDesc>

```

Listing 3.4: Definizione tipologia di fonte scritta

```

1 <sourceDesc>
2   <recordingStmt>
3     <recording type="audio"
4       dur="PT1H34M6S">
5       <!-- [...] -->
6     </recording>
7   </recordingStmt>
8 </sourceDesc>

```

Listing 3.5: Definizione tipologia di fonte orale

Figura 3.1: Identificazione tipologia di testimonianza per le fonti scritte e per le fonti orali

Tra gli ulteriori interventi volti all'estrazione delle informazioni principali riguardo la testimonianza stessa, c'è stata la definizione di un criterio per identificare ed estrapolare dal documento XML il contenuto dell'elemento che corrisponde al titolo

della testimonianza. Si è proceduto quindi con l'aggiunta dell'attributo `@xml:id` all'interno dell'elemento `<title>`, figlio di `<titleStmt>`, in tutte le testimonianze del corpus.

Un intervento simile è stato effettuato all'interno degli elementi `<person>`. Mediante l'aggiunta dell'attributo `@role` è stato possibile specificare il ruolo della persona in questione: in particolare, mediante il valore *testimone* si è potuto individuare ed estrarre le informazioni relative al testimone all'interno della codifica. La stessa operazione si è fatta, nel caso delle fonti orali, per identificare l'intervistatore.

Per poter invece specificare ed estrarre le informazioni circa la provenienza di una testimonianza, così da poter studiare a quale fondo appartengono le fonti primarie, nelle testimonianze scritte si è aggiunto l'elemento `<repository>`, nelle testimonianze orali il tag `<broadcast>`.

3.3 Marcatura delle entità nominate e dei riferimenti bibliografici

Come anticipato nella sezione 2.2, durante la fase di codifica è stata dedicata particolare attenzione all'identificazione e alla marcatura delle *named entities* e dei riferimenti storico-letterari presenti nella testimonianza.

Ulteriore passo nella fase di armonizzazione è stato quindi quello di verificare che in tutte le codifiche realizzate fossero stati marcati, all'interno della trascrizione, tutti i nomi di persona, di luogo, di organizzazione e tutti i riferimenti bibliografici, intervenendo laddove in alcune codifiche ci fossero delle lacune. Sono quindi state create quattro liste distinte, una per ciascuna *named entities* e una per la bibliografia: `<listPerson>`, `<listPlace>`, `<listOrg>` e `<listBibl>`. In ciascuna lista è riportato il relativo elenco delle entità menzionate nella testimonianza. Ad ogni voce di ciascuna lista è stato associato un identificatore univoco tramite l'attributo `@xml:id`. Questo permette di fare riferimento a questi elementi all'interno del testo utilizzando gli attributi `@ref` e `@target` con il relativo ID. Le liste create sono state tutte inserite all'interno dell'elemento `<standOff>`, riprendendo ed estendendo a tutte le codifiche il modello definito dalla dottoressa Chiara Baiolo.⁷⁵

```

1 <standOff>
2   <listPerson></listPerson>
3   <listPlace></listPlace>
4   <listOrg></listOrg>
5   <listBibl></listBibl>
6 </ standOff>

```

Listing 3.6: Struttura `<standOff>`

⁷⁵Cfr. Chiara Baiolo, *La codifica del Diario partigiano di Emanuele Artom: sulle tracce di una biblioteca da studiare*.

Nello specifico, le liste sono state create e organizzate come segue:

- <listPerson>: è la lista contenente le informazioni sulle persone menzionate nel testo. Ciascuna voce dell'elenco, annotata con <persName>, è stata realizzata per raccogliere tutte le informazioni biografiche su ciascuna persona: è stato indicato il nome (<forename>), il cognome (<surname>), eventuali note (<note>), il sesso (<sex>), la data e il luogo di nascita (<birth>), la data e il luogo di morte (<death>). Grazie all'attributo @source è possibile anche aggiungere il link ad una fonte esterna al documento riferita alla persona, come riportato nel listato 3.7.

```

1 <listPerson>
2     <person xml:id="EA" role="testimone" source="http://digital
-library.cdec.it/cdec-web/persone/detail/person-it-cdec-
eaccpf0001-000015/artom-emanuele.html?persone=%22Artom%2C+
Emanuele%22">
3         <persName>
4             <forename>Emanuele</forename>
5             <surname>Artom</surname>
6             <addName type="alias">Eugenio Ansaldi</addName>
7         </persName>
8         <note><p>Nome di battaglia: Eugenio Ansaldi</p></note>
9         <sex> M </sex>
10        <birth>
11            <date when="1915-06-23">23 giugno 1915</date>
12            <placeName>
13                <settlement type="municipality">Aosta</
settlement>
14                <country key="IT">Italia</country>
15            </placeName>
16        </birth>
17        <death>
18            <date when="1944-04-07">7 aprile 1944</date>
19            <placeName>
20                <settlement type="municipality">Torino</
settlement>
21                <country key="IT">Italia</country>
22            </placeName>
23        </death>
24        <note>Nato ad Aosta il 23 giugno 1915, figlio di Emilio
ed Amalia Segre. Nome di Battaglia: Eugenio Ansaldi.
Catturato dai fascisti nel marzo 1944, consegnato ai
tedeschi e rinchiuso nelle Carceri Nuove di Torino. Morto il
7 aprile 1944 a causa delle torture inflittegli.</note>
25    </person>
26    <!-- [...] -->
27 </listPerson>
```

Listing 3.7: Struttura <listPerson>

- <listPlace>: è la lista contenente le informazioni sui luoghi menzionati nel testo. Per ciascun elemento della lista, annotato con <placeName>, sono stati inseriti i tag <settlement>, che specifica il nome di un comune/municipio usato talvolta anche per indicare un quartiere di una città (con l'attributo @type) e <country>, che specifica il Paese in cui si trova il luogo.
- <listOrg>: è la lista contenente le informazioni sulle organizzazioni menzionate nel testo. Per ciascun elemento della lista, annotato con <orgName>, è stata specificata la sede (<placeName>) ed è stata inserita una breve descrizione (<desc>).
- <listBibl>: è la lista contenente la bibliografia, ovvero tutti i riferimenti storico-letterari citati all'interno del testo, ciascuno dei quali è codificato nella lista con <bibl> o <biblStruct>. Come mostrato nell'esempio all'interno del listato 3.8 riportato di seguito, il canto dal quale sono stati estratti i versi è codificato a livello analitico; la cantica cui esso appartiene, a livello monografico; l'intera opera – la Divina Commedia – a livello di serie.

```

1 <listBibl>
2   <head> La biblioteca di Emanuele Artom </head>
3   <biblStruct xml:id="Dante6"></biblStruct>
4   <biblStruct xml:id="Dante21"></biblStruct>
5   <biblStruct xml:id="Dante13">
6     <analytic>
7       <title level="a"> Canto XIII </title>
8     </analytic>
9     <monogr>
10       <author> Dante Alighieri </author>
11       <title level="m"> Inferno </title>
12       <imprint><date when="1314"> 1314 </date></imprint>
13     </monogr>
14     <series>
15       <title level="s"> Divina Commedia </title>
16     </series>
17   </biblStruct>
18   <bibl xml:id="FDL">
19     <author ref="#AT"> Adriano Tilgher </author>
20     <title> La filosofia di Leopardi </title>
21     <pubPlace> Roma </pubPlace>
22     <publisher> Edizioni di Religio </publisher>
23     <date when="1940"> 1940 </date>
24   </bibl>
25   <!-- [...] -->
26 </listBibl>
```

Listing 3.8: Struttura <listBibl>

3.4 Codifica delle tessere dantesche

Durante la fase di codifica si sono incontrate numerose tessere dantesche, che si sono però presentate in forma diversa. Si sono infatti identificate le citazioni, sia in forma implicita che in forma esplicita, di specifici versi della *Commedia*, le allusioni o le immagini che rimandano più genericamente all'opera e, infine, i termini danteschi. È stato quindi necessario stabilire una diversa modalità di marcatura che permettesse di esplicitare e di distinguere opportunamente le differenti tipologie. Nelle seguenti sezioni verrà approfondita la codifica implementata per ciascuna tipologia.

3.4.1 Codifica delle citazioni

Le citazioni dantesche possono presentarsi in forma esplicita o in forma implicita: nel primo caso si intendono le citazioni nelle quali vengono riportati fedelmente i versi di un determinato canto della *Commedia*; nel secondo si intendono i passaggi non riportati fedelmente ma dei quali si intuiscono comunque i versi di riferimento. Le citazioni sono state marcate con il tag `<cit>`, distinguendo tra citazioni esplicite e implicite mediante il valore dell'attributo `@type`: per le prime, `@type="explicit"`, per le seconde `@type="implicit"`.

Nei listati 3.9 e 3.10 sono riportati due esempi di codifica di citazioni esplicite e implicite.

```

1 <cit type="explicit">
2   <quote>
3     <lg>
4       <l>
5         <said>"Miserere di me" gridai a lui </said>
6       </l>
7       <l>
8         <said>"qual tu sii, od ombra od omo certo?"</said>
9       </l>
10      </lg>
11    </quote>
12    <ptr target="#n43"/>
13    <ref target="#infI">
14      <bibl>
15        <author ref="#dante">
16          <forename> Dante </forename>
17          <surname> Alighieri </surname>
18        </author>
19        <title level="s" ref="#DC"> La Divina Commedia</title>
20        <title level="m" ref="#dcInf"> Inferno </title>
21        <title level="a" ref="#infI"> Canto I </title>
22        <citedRange>vv.65-66</citedRange>
23      </bibl>

```

```

24      </ref>
25 </cit>
```

Listing 3.9: Codifica di una citazione esplicita (dal Diario di Luigi Giuntini)

```

1 <cit type="implicit">
2   <quote>l'altro sonava
3     <lb xml:id="p128_17" n="7" facs="#pag131_17"/>
4     di continuo la tromba di Barbariccia<pc>.</pc>
5   </quote>
6   <ref target="#Dante21"><bibl>XXI canto Inferno, v. 139</bibl></ref>
7 </cit>
```

Listing 3.10: Codifica di una citazione implicita (dal Diario di Emanuele Artom)

L’elemento `<cit>` presenta al suo interno due figli: `<quote>` e `<ref>`, padre a sua volta di `<bibl>`. L’elemento `<quote>` marca il passaggio in cui viene riportata la citazione, mentre `<ref>`, mediante l’attributo `@target`, serve per collegare la citazione al relativo riferimento bibliografico riportato all’interno dell’elenco definito in `<listBibl>`. L’elemento `<bibl>`, invece, contiene i componenti della citazione bibliografica.

Nei listati 3.11 e 3.12 è mostrato un esempio del meccanismo di collegamento tra la citazione codificata nella testimonianza e la relativa fonte bibliografica inserita in `<listBibl>`.

```

1 <cit type="explicit">
2   <quote>
3     <lg>
4       <l>Grandine grossa<pc>, </pc> acqua tinta e neve</l>
5       <lb xml:id="p68_13" n="3" facs="#pag70_13"/>
6       <l>per l'<choice><sic>aer</sic><corr>aere</corr></choice>
7         tenebroso si riversa<pc>; </pc></l>
8     </lg>
9   </quote>
10  <ref target="#Dante6">
11    <bibl>VI canto Inferno, vv. 10-11</bibl>
12  </ref>
13 </cit>
```

Listing 3.11: Meccanismo di collegamento di una citazione con il relativo riferimento bibliografico in `<listBibl>`: codifica della citazione nel testo

```

1 <listBibl>
```

```

2   <biblStruct xml:id="Dante6">
3     <analytic>
4       <title level="a"> Canto VI </title>
5     </analytic>
6     <monogr>
7       <author> Dante Alighieri </author>
8       <title level="m"> Inferno </title>
9       <imprint><date when="1314"> 1314 </date></imprint>
10    </monogr>
11    <series>
12      <title level="s"> Divina Commedia </title>
13    </series>
14  </biblStruct>
15 </listBibl>
```

Listing 3.12: Meccanismo di collegamento di una citazione con il relativo riferimento bibliografico in `<listBibl>`: codifica del riferimento bibliografico

3.4.2 Codifica delle allusioni

All'interno delle testimonianze sono stati registrati anche casi in cui sono presenti allusioni, rimandi o riferimenti alla *Commedia*. Non essendo citazioni vere e proprie, in questo caso non è stato utilizzato l'elemento `<cit>` ma solamente l'elemento `<ref>`⁷⁶ con l'aggiunta dell'attributo `@type` con valore `"allusion"`. È presente, inoltre, l'attributo `@target` che permette di collegare l'allusione al relativo elemento definito in `<listBibl>`. In questo caso, non trattandosi di citazioni ma di immagini che rimandano alla *Commedia* in modo più generico, si è scelto di indicare solamente la cantica a cui si fa riferimento, senza specificare il canto.

Nel listato 3.13 è mostrato un esempio di codifica di un'allusione.

```

1 <ref type="allusion" target="#Dante">Liliana ha detto
2   "Per me Goti è come <term type="dante" ref="#virgilio">Virgilio
3   </term> per <term type="dante" ref="#dante">Dante</term>""
4   <bibl>Inferno , Divina Commedia</bibl>
5 </ref>
```

Listing 3.13: Codifica di una allusione nella testimonianza di Goti Bauer

⁷⁶ Attualmente la codifica delle allusioni avviene mediante l'elemento `<ref>`, ma la questione non è propriamente definita in quanto potrebbero essere utilizzanti anche altri tag, quali `<rs>` o `<seg>`.

3.4.3 Codifica dei vocaboli danteschi

I termini danteschi presi in prestito dalla *Commedia* sono stati individuati e marcati all'interno della trascrizione della testimonianza mediante l'elemento <term>, come mostrato nel listato 3.14.

```
1   <term type="dante" ref="#bolgia">bolgia</term>
```

Listing 3.14: Codifica di un termine dantesco

In ciascuna testimonianza è stata inoltre creata, all'interno dell'elemento <standOff>, una lista di <item>, ciascuno corrispondente ad un vocabolo dantesco presente nella testimonianza. All'interno quindi degli elementi <term> marcati nella trascrizione, è stato aggiunto l'attributo @type con valore "dante", per specificare che si tratta di un vocabolo di tipo dantesco, e l'attributo @ref, per collegare l'elemento al relativo <item> definito in <list>, come visibile nell'esempio 3.15.

```
1  <list type="termini_danteschi" xml:id="listaTerminiDante">
2    <!-- Bolgia -->
3    <item xml:id="bolgia">
4      <term>Bolgia </term>
5      <ptr target="https://www.treccani.it/enciclopedia/bolgia_
%28Enciclopedia-Dantesca%29/">
6    </item>
7    <-- [...] -->
8 </list>
```

Listing 3.15: Lista dei termini danteschi presenti nella testimonianza

Capitolo 4

Creazione di un'applicazione web

Una volta terminata la fase di armonizzazione delle testimonianze, si è passati alla creazione di un'applicazione web che permetesse la presentazione, la consultazione e l'analisi dei dati prodotti durante la fase di codifica.

Come anticipato nel capitolo 1, nel corso del progetto sono state sperimentate due differenti strategie di restituzione dei dati, facendo uso di due diversi approcci architettonici:

- (a) per il primo approccio, le applicazioni web sono state sviluppate sfruttando le funzionalità di una libreria client-side per l'elaborazione di documenti XML: SaxonJS⁷⁷
- (b) per il secondo, quello attuale, le applicazioni web sono state sviluppate adottando le tecnologie messe a disposizione dall'ambiente eXist-db⁷⁸ mediante l'uso del modulo *HTML-templating* che segue invece un approccio server-side.

Fino ad oggi sono state implementate e proposte alcune versioni di applicazione web da parte di alcuni studenti del progetto in occasione del loro lavoro di tesi. Le applicazioni realizzate consentivano diverse funzionalità ma erano state sviluppate prendendo in analisi una o poche testimonianze del corpus. L'applicazione web *Voci dall'Inferno*, descritta e presentata in questo capitolo, 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. In questo capitolo ne verranno spiegate in dettaglio le funzionalità e verranno illustrati i risultati emersi da una prima indagine svolta considerando tutti i dati a disposizione dell'archivio digitale.

⁷⁷Cfr. <https://www.saxonica.com/html/welcome/welcome.html>.

⁷⁸Cfr. <https://exist-db.org/exist/apps/homepage/index.html>.

4.1 eXist-db

L'applicazione web *Voci dall'Inferno* è stata realizzata mediante le tecnologie messe a disposizione dall'ambiente eXist-db, un software open source scritto in Java e progettato per database NoSQL basati su XML. 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 templates in HTML, in cui si aggiungono opportune chiamate a funzioni XQuery che navigando la collezione di dati in formato XML, favoriscono l'estrazione dell'informazione relativa ai nodi XML di interesse e la generazione di frammenti HTML che permettono di completare la pagina visualizzata, infine, dal browser. Il principale vantaggio dell'uso dei templates è quello di mantenere separato il codice di presentazione da quello di elaborazione dei dati.

In figura 4.1 è riportato un diagramma che illustra il funzionamento di HTML Templating Framework.

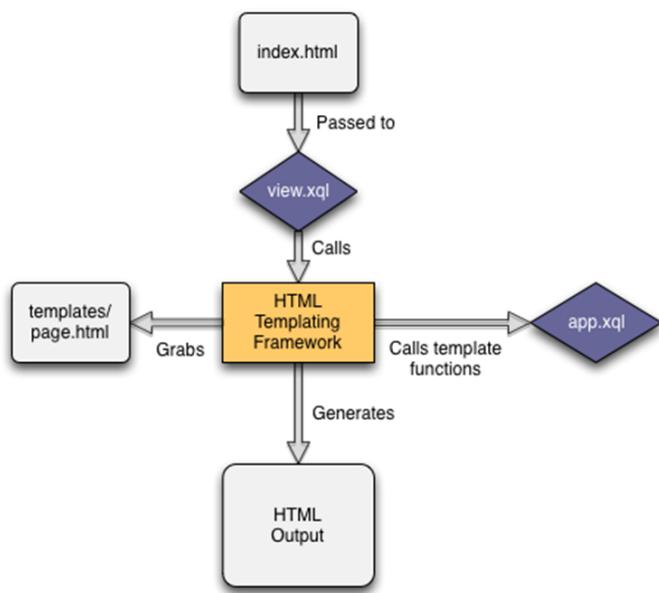


Figura 4.1: Diagramma HTML Templating Framework

Una caratteristica rilevante della tecnologia eXist-db è la possibilità di avvalersi della libreria *Apache Lucene*⁷⁹ per l'indicizzazione dei dati testuali e per la conseguente interrogazione degli stessi. Lucene permette a eXist di offrire funzionalità di ricerca avanzate, come la ricerca di parole vicine tra loro o parole simili ad altre, utilizzando operatori di confronto testuale booleani e molto altro.

⁷⁹Cfr. <https://lucene.apache.org/>.

Il linguaggio XQuery (*XML Query Language*) è un linguaggio di interrogazione progettato per selezionare e per manipolare dati in formato XML. È uno standard definito e manutenuto dal W3C (*World Wide Web Consortium*) ed è basato su espressioni XPath⁸⁰ per navigare all'interno di documenti XML e per selezionare nodi specifici.

XQuery is a query language designed by the W3C to address these needs. It allows you to select the XML data elements of interest, reorganize and possibly transform them, and return the results in a structure of your choosing.⁸¹

Nonostante ciò, XQuery presenta una serie di funzionalità che non sono incluse in XPath, come l'espressione *FLWOR* (For, Let, Where, Order By, Return), che consente di iterare, di filtrare, di ordinare e di restituire dati XML.

XQuery and XPath overlap to a very large degree. They have the same data model and the same set of built-in functions and operators. XPath is essentially a subset of XQuery. XQuery has a number of features that are not included in XPath, such as FLWORs, user-defined functions, and XML constructors. This is because these features are not relevant to selecting, but instead have to do with structuring, sorting query results or more complex programming.⁸²

Nel listato 4.1 è riportato un esempio di una funzione XQuery attraverso la quale viene creato l'elenco dei testimoni che appartengono ad una specifica collezione.

```

1 declare function app:ElencoTestimoniPerCollezione($node as node() ,
2   $model as map(*)){
3   let $nome_collezione :=request:get-parameter("collezione", "")
4   let $collezione:=
5     if ($nome_collezione="deportati") then
6       let $path:="/db/apps/Voci_Dall_Inferno/xml/Deportati"
7       return $path
8     else
9       let $path:="/db/apps/Voci_Dall_Inferno/xml/NonDeportati
"
10
11      return $path
12
13      let $testimoni:=
14        for $xml in collection($collezione)
15          let $find_testimone := $xml//tei:person[@role="testimone"]
16          let $nome_testimone := fn:normalize-space($find_testimone/
17            tei:persName/tei:forename)
```

⁸⁰Cfr. <https://www.w3.org/TR/xpath-31/>.

⁸¹Priscilla Walmsley, *XQuery*, O'Reilly Media, University of Michigan, 2016, p. 1.

⁸²Priscilla Walmsley, *XQuery*, p. 16.

```

15      let $cognome_testimone := fn:normalize-space(string-join(
16          $find_testimone/tei:persName/tei:surname, " "))
17      let $testimone := concat($cognome_testimone, " ",
18          $nome_testimone)
19
20
21      order by $testimone
22      return $testimone
23
24
25      for $testimone in distinct-values($testimonii)
26      return
27          <div class="elenco_testimonii">
28              <section class="testimone">
29                  <p>
30                      <a class="nome_testimone" onclick='
31                          mostra_testimonianze("{$testimone}")'>{($testimone)}</a>
32                  </p>
33              </section>
34          </div>
35      };

```

Listing 4.1: Funzione XQuery per la creazione delle liste dei testimoni in base alla collezione

4.2 L'applicazione web *Voci dall'Inferno*

Mediante l'ambiente eXist-db è stata quindi realizzata l'applicazione web *Voci dall'Inferno*, che consente la presentazione, la fruizione e l'analisi delle informazioni contenute ed estratte di tutti i dati in formato XML/TEI che costituiscono il corpus.

Tutti i grafici presenti nell'applicazione web sono stati realizzati mediante *Highcharts*,⁸³ una libreria JavaScript per la *data visualization*. Highcharts funziona su tutti i tipi di browser senza bisogno di plugin aggiuntivi e offre la possibilità di creare un'ampia varietà di grafici interattivi che possono essere esplorati, migliorando la comprensione delle informazioni fornite e rendendo l'analisi più intuitiva. Tra le azioni possibili c'è l'approfondimento mediante zoom su un'area specifica del grafico, utile per esaminare meglio dettagli su grandi dataset. Inoltre, quando l'utente passa il mouse su un punto o su una serie di dati in un grafico, appare un tooltip che mostra informazioni dettagliate su quel punto. I tooltip sono altamente personalizzabili e possono includere valori numerici, testo, formattazioni personalizzate e immagini. È possibile interagire con la legenda del grafico per nascondere o mostrare diverse serie di dati, funzionalità utile per confrontare diverse categorie o per concentrarsi solo su alcune parti del dataset.

⁸³Cfr. <https://www.highcharts.com/>.

Highcharts include anche animazioni che migliorano l'interattività e l'aspetto visivo del grafico. Per esempio, quando un grafico viene caricato, le serie possono apparire gradualmente o crescere visivamente. La funzionalità *drilldown* consente di cliccare su una categoria o su un punto approfondire i dettagli e visualizzare un grafico più specifico, operazione utile per esplorare ulteriormente dati gerarchici. Un'altra caratteristica interessante di Highcharts è la dinamicità, che consente di caricare in modo dinamico dati provenienti da fonti esterne aggiornandone la visualizzazione. Tale operazione può essere fatta, ad esempio, passando i dati in formato JSON. Highcharts offre inoltre all'utente la possibilità di esportare i grafici in diversi formati (PNG, JPG, PDF o SVG) tramite un pulsante posizionato nell'angolo del grafico. Infine, i grafici realizzati mediante questa libreria sono responsive, funzionalità che consente ai grafici di adattarsi a diverse dimensioni di schermo e di dispositivi, garantendone una visualizzazione ottimale.

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*.

Di seguito si procede all'illustrazione e alla descrizione di ciascuna sezione.

Come mostrato in figura 4.2 la homepage presenta il logo del progetto *Voci dall'Inferno*,⁸⁴ il titolo e il sottotitolo.



Figura 4.2: Homepage dell'applicazione web

La sezione *Il progetto* descrive i motivi per i quali i sopravvissuti si sono serviti del vocabolario offertogli da Dante per riuscire a 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

⁸⁴Il logo del progetto *Voci dall'Inferno* è stato realizzato dalla dottoressa Martina Felicina Rossi.

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 il progetto e 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, come mostrato in figura 4.3. 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:	
Cognome:	
Anno di nascita:	
Luogo di nascita:	
Anno di morte:	
Luogo di morte:	

CERCA

Figura 4.3: 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.

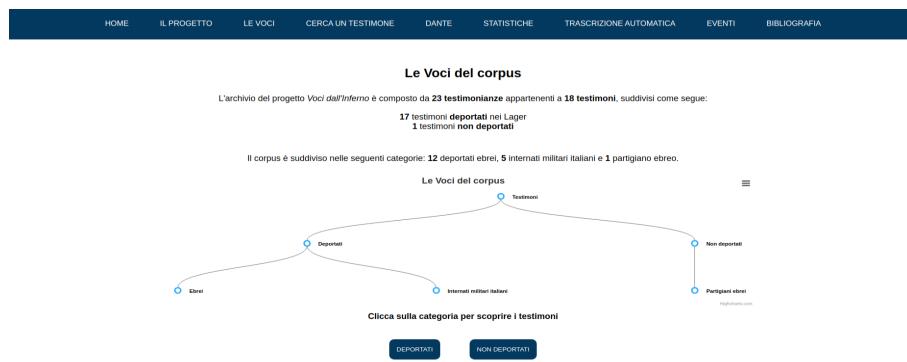


Figura 4.4: Classificazione delle testimonianze all'interno dell'archivio digitale

Come visibile in figura 4.4, 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 *treemap chart* che illustra la struttura gerarchica dei testimoni all'interno della tassonomia attualmente definita. Come

anticipato nel capitolo 1, il corpus dispone, infatti, anche della testimonianza di chi non visse il Lager. È il caso di Emanuele Artom, simbolo della Resistenza ebraica italiana, il cui diario partigiano rappresenta un'importante testimonianza storico-letteraria. Nonostante 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. In quanto non deportato, per poter integrare la sua testimonianza all'interno dell'applicazione è stata introdotta una nuova categoria, attualmente chiamata *Non deportati*.⁸⁵ È ancora in fase di studio una più precisa e puntuale tassonomia dei testimoni, con l'obiettivo futuro di ampliare ulteriormente la classificazione all'interno del corpus.

Attualmente l'archivio del progetto è quindi costituito da 18 testimoni, suddivisi come segue: 17 deportati nei Lager, di cui 12 ebrei e 5 internati militari italiani, e 1 non deportato, ovvero il partigiano ebreo Emanuele Artom.

Cliccando su uno dei due pulsanti disponibili, *Deportati* o *Non deportati*, si accede all'elenco dei testimoni che costituiscono la categoria scelta, come visibile in figura 4.5.

Figura 4.5: 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 vasto. 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 fruibili dall'utente. In figura 4.6 è mostrata la pagina in questione.

Una volta scelta e cliccata la testimonianza che si desidera esaminare, si accederà ad una pagina in cui sono inizialmente riportate alcune informazioni riguardo la te-

⁸⁵Cfr. Chiara Baiolo, *La codifica del Diario partigiano di Emanuele Artom: sulle tracce di una biblioteca da studiare*.

Liliana Segre

Liliana Segre
Nata a Milano il 10 dicembre 1930, figlia di Alberto e Lucia Folgini. Perseguitata nazista. Arrestata a Settimo di Stura (VA) il 16 gennaio 1943 dai fascisti. Deposta da Milano ad Auschwitz il 30 gennaio 1944 allora di 13 anni. Matricola 75190. Liberata nel circondario di Ravensbrück il 30 aprile 1945. È sopravvissuta alla Shoah.

NUMERO DI TESTIMONIANZE DISPONIBILI: 2

Testimonianza di Liliana Segre, 3 dicembre 2006
Testimonianza di Liliana Segre, 2007

FONTE

Figura 4.6: Scheda anagrafica del testimone ed elenco delle relative testimonianze consultabili

stimonianza, come il tipo di fonte (se si tratta di una testimonianza scritta o orale), le lingue presenti all'interno della testimonianza, la data di creazione della risorsa e da chi è stata realizzata la trascrizione e la codifica. Sulla base della tipologia di fonte che si sta consultando, avverrà poi una prima distinzione. Mediante opportune funzioni XQuery viene navigato il documento XML e, in base al criterio di marcatura stabilito per distinguere la tipologia illustrato nella sezione 3.2, verranno estratti dati differenti a seconda della testimonianza consultata.

Nel caso di interviste orali saranno visualizzate specifiche informazioni riguardo il numero di cassette di cui si compone la registrazione, la durata e da chi è stata realizzata l'intervista (figura 4.7a). Se si tratta di una testimonianza scritta, saranno estratte altre informazioni utili come il materiale di supporto, che fornisce una descrizione fisica del manoscritto, il riferimento alla fonte primaria e il codice identificativo della risorsa (figura 4.7b).

Enrico Vanzini
Testimonianza di Enrico Vanzini, l'ultimo Sonderkommando italiano

Fonte: orale
Tipologia: video
Link: <https://youtu.be/S50qkVEpXjw>
Numero cassette: 1
Durata: 0 ore 36 minuti 50 secondi
Realizzata da: TelePace
Lingue: Italiano
Note: /
Trascrizione e codifica a cura di: Elvira Mercatani

Emanuele Artom
Diario di Emanuele Artom (novembre 1943-febbraio 1944)

Fonte: scritta
Tipologia: diario
Data: novembre 1943 - febbraio 1944
Lingue: Italiano Piemontese Latino Francese Inglese
Materiale: Fogli di carta
Source: <http://digital.library.cedc.it/cdec-web/tbrcodes/rst/IT-CDEC-ST003.000006/dan.html>
Codice identificativo: busta 1, fascicolo 9
Trascrizione e codifica a cura di: Chiara Basile

LEGGI LA TESTIMONIANZA CONSULTA LE STATISTICHE IL DANTE DI ENRICO VANZINI

LEGGI LA TESTIMONIANZA CONSULTA LE STATISTICHE IL DANTE DI EMANUELE ARTOM

(a) Informazioni estratte per una testimonianza orale

(b) Informazioni estratte per una testimonianza scritta

Figura 4.7: Informazioni sulla testimonianza consultata

Mediante tre pulsanti, è possibile accedere direttamente a tre diverse sezioni della pagina: il primo pulsante permette di consultare la trascrizione della testimonianza,

il secondo di approfondirne l'analisi mediante alcuni grafici, il terzo di visualizzare le tessere dantesche presenti all'interno della testimonianza e di confrontarle con il testo della *Divina Commedia*.

Come descritto nella sezione 2.4, 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 la fase di codifica, ciò ha portato alla definizione di specifici elementi <timeline> 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, chiamata *regesto*, che contiene un riasunto del contenuto della trascrizione. Tale sezione è stata rappresentata mediante l'elemento <abstract> che è figlio di <profileDesc> e a sua volta padre dell'elemento <ab>. All'interno di <ab> è presente una lista di segmenti <item>, ciascuno dei quali contiene un'unità testuale che mediante l'attributo @synch consente il collegamento al relativo elemento della timeline.

Grazie a tale meccanismo, nel caso della consultazione di una registrazione orale, è stato possibile mostrare in fase di visualizzazione il contenuto delle porzioni di testo contenute nei singoli segmenti, come mostrato in figura 4.8. Avendo allineato il minutaggio definito nell'elemento <timeline> con i segmenti, è anche possibile ascoltare la determinata parte della registrazione in cui viene trattato l'argomento riassunto all'interno di ciascun <item>.

Figura 4.8: Contenuto dei segmenti di sintesi per una testimonianza orale

Anche nel caso della trascrizione della testimonianza, essa presenterà una struttura differente dal punto di vista della visualizzazione in base alla tipologia della testimonianza consultata. Nel caso di una fonte orale, come nell'esempio in figura 4.9a, saranno presenti i nomi dei parlanti e i relativi enunciati che compongono la conversazione. Nel caso, invece, di una fonte scritta, la trascrizione cambierà ulteriormente la sua presentazione in base al tipo di fonte, il cui contenuto differisce strutturalmente se si tratta di un diario oppure di una lettera. Nell'esempio in figura 4.9b è

mostrata una porzione di trascrizione del diario di Luigi Giuntini. In futuro verrà definita una modalità di allineamento e visualizzazione della trascrizione con la fonte primaria.

(a) Trascrizione di una testimonianza orale

(b) Trascrizione di una testimonianza scritta

Figura 4.9: Trascrizione di una testimonianza

La trascrizione è affiancata da una legenda che riporta l'elenco di alcuni fenomeni esplicitati durante la codifica. Cliccando sul tasto *Mostra tutti i fenomeni* è possibile mettere in evidenza nel testo gli elementi marcati, come mostrato in figura 4.10. La legenda cambierà la sua struttura in base al tipo di fonte consultata, presentando elementi differenti opportunamente identificati e marcati durante la fase di codifica. Una conversazione orale presenta una serie di fenomeni che possono essere considerati come una componente del testo parlato. Durante la produzione di un enunciato è infatti possibile incontrare ripetizioni, false partenze, parole troncate, pause vocali, suoni non lessicali ed eventi prossemici, come descritto precedentemente nella sezione 2.4. Nel caso, quindi, della consultazione di una testimonianza orale, le voci della legenda corrisponderanno a tali elementi.

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

Fenomeni marcati

- Bucco nella registrazione: GAP XXX
- Parole non chiare: UNCLEAR
- Pause: PAUSE (...)
- Esclamazione: OH
- Rumore: NOISE
- Movimento: KINESIC
- Frasi o parole riformulata/petute: DEL
- Parola errata: SIC
- Parola corretta: CORR
- Forma dialettale: ORIG
- Forma regolarizzata: REG
- Abbreviazione: ABBR
- Forma estesa: EXPAN
- Parola enfatizzata: EMPH
- Parola in lingua straniera: FOREIGN
- Antroponimo: PERSONNAME
- Luogo: PLACENAME
- Organizzazione: ORGANNAME

MOSTRA TUTTI I FENOMENI

Figura 4.10: Trascrizione di una testimonianza con legenda attiva

dialettologia e l'analisi delle conversazioni e del discorso. Negli ultimi anni sono state sviluppate molteplici convenzioni di trascrizione, ciascuna con il proprio modello di dati e formati. Non esiste attualmente uno standard comune a tutti gli ambiti e necessità. Questa funzionalità è attualmente in corso di sviluppo. Per il momento, riprendendo il lavoro svolto dalla dottessa Diletta Lelli⁸⁶, per la trascrizione delle testimonianze orali si sono utilizzate le convenzioni offerte da CHAT⁸⁷, adottando attualmente 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 (...).

Per quanto riguarda le fonti scritte, la legenda presenterà specifici elementi relativi a questo tipo di risorsa, quali ad esempio le lacune presenti nel testo, le aggiunte, gli errori di scrittura, le parole che il testimone scrive in forma abbreviata e le parole arcaiche.

Per entrambe le macro-categorie, è possibile inoltre visualizzare nel testo le *entità nominate*: saranno quindi evidenziati tutti i nomi di persona, di luogo e di organizzazione.

Se all'interno della trascrizione si vogliono mettere in evidenza soltanto determinati fenomeni, è possibile cliccare nella legenda direttamente sulla voce relativa al fenomeno di interesse. In figura 4.11 è riportato l'esempio della trascrizione della testimonianza orale di Ida Marcheria: in questo caso sono stati messi in evidenza le lacune presenti nella registrazione e le pause vocali.

Come precedentemente anticipato, cliccando sul secondo pulsante, ovvero *Consulta le statistiche*, si accede alla sezione che presenta alcune analisi statistiche riguardo le testimonianze, ottenute studiando i fenomeni marcati in fase di codifica. I dati relativi a tali fenomeni sono stati estratti e utilizzati per la creazione di tre grafici,

⁸⁶Cfr Diletta Lelli, *Idek Wolfowicz: un testimone ritrovato*.

⁸⁷Cfr. <https://talkbank.org/manuals/CHAT.pdf>.

The screenshot shows a web page with a dark header containing navigation links: HOME, IL PROGETTO, LE VOCI, CERCA UN TESTIMONE, DANTE, STATISTICHE, TRASCRIZIONE AUTOMATICA, EVENTI, and BIBLIOGRAFIA. Below the header is a large text area containing a transcription of a witness statement. On the right side, there is a sidebar titled "Fenomeni marcati" (Marked phenomena) which lists various types of phenomena such as "Buchi nella registrazione: GAP XXX", "Parola non chiara: UNCLEAR", etc. A button at the bottom of the sidebar says "MOSTRA TUTTI I FENOMENI".

Figura 4.11: Trascrizione testimonianza con attivazione solo di alcuni fenomeni

ciascuno riguardante un determinato aspetto. Anche in questo caso, i fenomeni illustrati nei tre grafici differiscono in base al tipo di testimonianza consultata.

Il primo grafico, un *pie chart*, mostra i fenomeni che caratterizzano la fonte in questione: nel caso delle testimonianze orali, come in figura 4.12a, sono visualizzati elementi quali le lacune presenti all'interno della registrazione, i passaggi poco chiari, le esclamazioni, i rumori accidentali e gli altri fenomeni descritti precedentemente; nel caso delle testimonianze scritte, come in figura 4.12b, sono messe in evidenza alcune pratiche editoriali, come le lacune, le cancellazioni e le aggiunte presenti nella testimonianza. Poiché i grafici sono stati realizzati mediante la libreria interattiva *Highcharts* precedentemente illustrata, è possibile effettuare un *hover* con il cursore sulla porzione di grafico relativa al fenomeno di interesse per visualizzarne il numero di occorrenze all'interno della testimonianza.

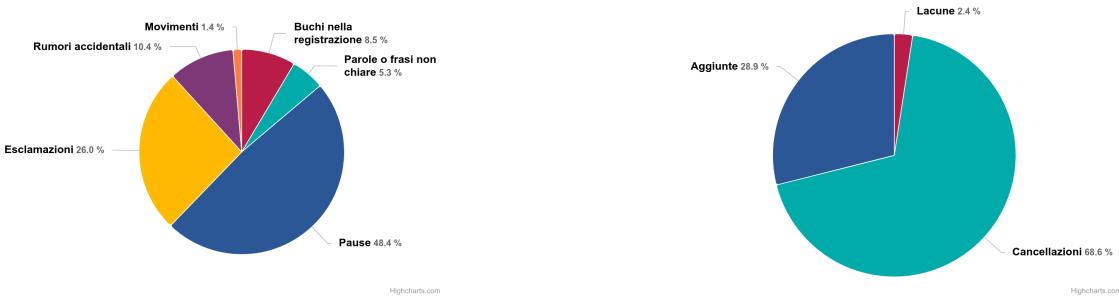


Figura 4.12: Grafico che mostra i fenomeni presenti in una testimonianza

Il secondo grafico, un *bar chart*, pone invece l'attenzione sul modo di esprimersi del testimone, mostrando il dato espresso sia in forma percentuale che in forma assoluta. Nel caso di fonti orali, come visibile in figura 4.13a, sono stati estratti i dati

riguardanti le parole o le frasi riformulate, ripetute, corrette o interrotte durante l'enunciato, le parole errate, abbreviate, espresse in forma dialettale, enfatizzate e quelle in lingua diversa rispetto alla lingua principale della testimonianza. Nel caso, invece, di fonti scritte, come in figura 4.13b, nel grafico sono riportati gli errori di scrittura del testimone, le parole normalizzate, quelle scritte in forma abbreviata, quelle enfatizzate, cioè sottolineate per effetto linguistico o retorico, le parole in lingua diversa rispetto alla lingua principale del testo e le parole arcaiche.

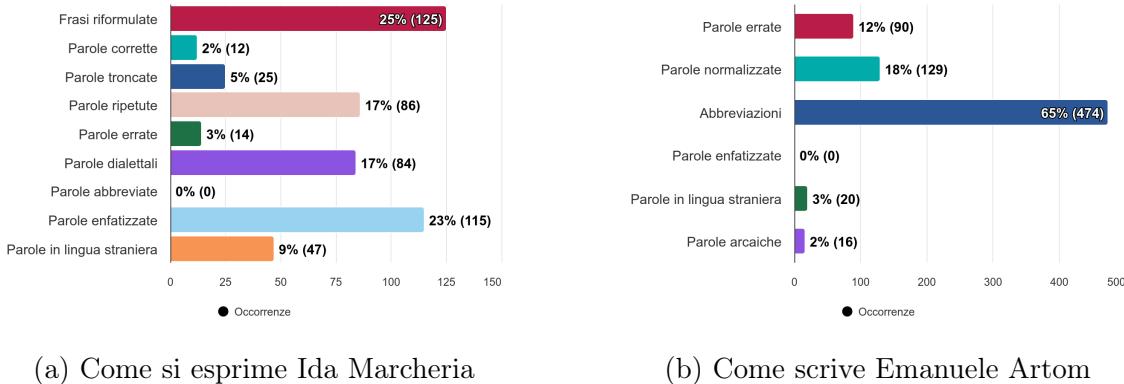


Figura 4.13: Grafico che analizza la modalità di espressione del testimone

Durante la fase di codifica è stata dedicata particolare attenzione all'identificazione e alla marcatura delle *named entities*, secondo la modalità descritta nella sezione 3.3. Il terzo grafico prende in analisi proprio le entità nominate. Nello specifico, per ciascuna di esse, sono stati messi a confronto, mediante un *grouped bar chart*, i dati ottenuti contando il numero di elementi menzionati all'interno della testimonianza (ovvero gli elementi contenuti all'interno di ciascuna lista definita nell'elemento `<standOff>`) e quante volte questi elementi occorrono nel testo (ovvero all'interno del blocco della trascrizione racchiuso all'interno di `<text>`).

Come mostrato in figura 4.14, nel diario di Emanuele Artom, ad esempio, le persone citate sono 144 mentre i nomi di persona che occorrono nel testo sono 581.

Per un maggiore approfondimento delle *named entities* è stata creata, per ciascuna entità nominata, una tabella ordinata in maniera decrescente per numero di occorrenze, che mostra quali sono le persone, i luoghi e le organizzazioni menzionate e quante volte ricorrono nel testo. In figura 4.15 è riportata l'esempio della tabella relativa alle persone menzionate nella testimonianza di Emanuele Artom, per le quali, oltre al nome, sono state estratte altre informazioni quali, ad esempio, i dati anagrafici, eventuali note e un link che contenente un approfondimento esterno.

Oltre alle entità nominate, durante la fase di codifica sono stati marcati anche tutti i riferimenti bibliografici presenti nelle testimonianze. Anche i dati relativi ad essi sono stati estratti e mostrati mediante un'apposita tabella, come visibile in figura 4.16.

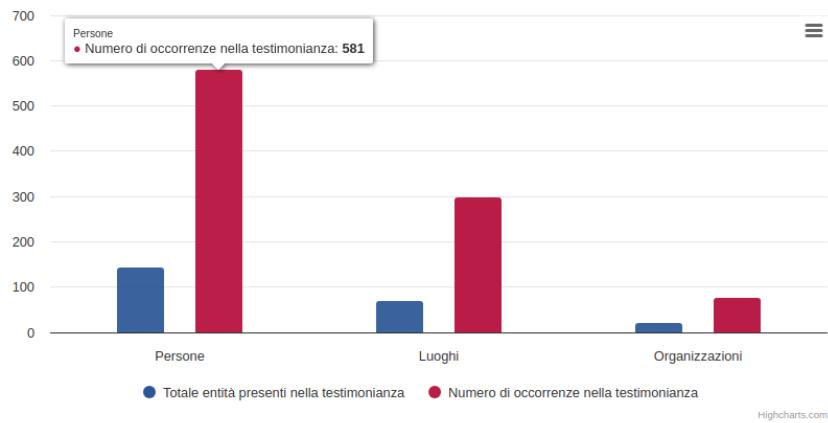


Figura 4.14: Grafico delle entità nominate presenti (testimonianza di Emanuele Artom)

Quali persone e quante volte vengono citate nella testimonianza?							
ID	Occorrenze	Nome	Note	Sesso	Nascita	Morte	Link
GS	75	Giorgio Segre	/	M	20 febbraio 1920 (Torino Piemonte Italia)	13 agosto 1996	Scopri di più
Z	29	Eduardo Zapata Granya	Nome di battaglia: Zama	M	Ecuador	/	Scopri di più
RM	29	Roberto Malan	/	M	24 marzo 1920 (Catania Sicilia Italia)	2006	Scopri di più
GINA	24	Pasqualina Rossi	Nome di battaglia: Gina	F	1908 (Valmacca Alessandria Piemonte Italia)	/	Scopri di più
BARB	22	Pompeo Colajanni	Nome di battaglia: Nicola Barbatto	M	4 gennaio 1906 (Caltanissetta Sicilia Italia)	8 dicembre 1987 (Palermo Sicilia Italia)	Scopri di più
GEO	20	Ruggero Levi	Nome di battaglia: Geo	M	12 marzo 1927 (Torino Piemonte Italia)	/	Scopri di più
		Giudalbeno			7 ottobre 1801 (Torino Piemonte)		Scopri

Figura 4.15: Tabella relativa alle persone menzionate (testimonianza di 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, come descritto in sezione 3.4, le differenti tipologie di tessere dantesche: le citazioni implicite, le citazioni esplicite, le allusioni e i vocaboli danteschi. Cliccando sulla citazione o allusione riportata nella parte sinistra della schermata, come visibile in figura 4.17, sarà possibile mettere in evidenza i versi o la porzione di testo della *Commedia* alla quale si fa riferimento.

Attualmente questa funzionalità di collegamento tra le citazioni presenti nella testimonianza e l'opera dantesca è disponibile solamente per la testimonianza di Emanuele Artom, per la quale la marcatura dei versi danteschi per i quali è presente un riferimento è stata realizzata manualmente. Inoltre, attualmente, anche l'identificazione delle tessere dantesche presenti nelle testimonianze avviene manualmente. Tra gli obiettivi di ricerca futuri del progetto, infatti, è previsto lo sviluppo di tecniche avanzate per il riconoscimento automatico dei riferimenti a Dante tramite l'uso di metodi per l'estrazione automatica di informazioni da risorse testuali, quali ad

La biblioteca di Emanuele Artom						
ID	Occorrenze	Autore	Titolo	Luogo di pubblicazione	Editore	Data
OF	3	Ludovico Ariosto	Orlando Furioso	Ferrara	/	1516
Dante6	2	Dante Alighieri	Canto VI Inferno Divina Commedia	/	/	1314
M1	2	Alessandro Manzoni	Marzo 1821	/	/	1848
Dante21	1	Dante Alighieri	Canto XXI Inferno Divina Commedia	/	/	1314
Dante13	1	Dante Alighieri	Canto XIII Inferno Divina Commedia	/	/	1314
FDL	1	Adriano Tilgher	La filosofia di Leopardi	Roma	Edizioni di Religio	1940
DVA	1	Amerigo Ruggiero	L'America al bivio	Torino	Einaudi	1934
FED	1	Platone	Pedone	/	/	IV secolo a.C.
MCC	1	Angelo Brofferio	Ij bonbon 'd sor Cont	/	/	1853

Figura 4.16: Tabella relativa ai riferimenti bibliografici (testimonianza di Emanuele Artom)

The screenshot shows the 'BIBLIOGRAFIA' section of the application. At the top, there is a navigation bar with links to HOME, IL PROGETTO, LE VOCI, CERCA UN TESTIMONE, DANTE, STATISTICHE, TRASCRIZIONE AUTOMATICA, EVENTI, and BIBLIOGRAFIA. Below the navigation bar, the title 'Il Dante di Emanuele Artom' is displayed. A note states: 'Nella testimonianza di Emanuele Artom sono presenti 4 tessere dantesche, di cui:' followed by a list of counts: Citazioni esplicite: 1, Citazioni implicite: 3, Allusioni / riferimenti: 0, Termini danteschi: 0. The main content area is divided into sections for 'Citazioni esplicite' and 'Citazioni implicite'. Under 'Citazioni esplicite', there is one entry: 'Grande grossa, acqua tinta e neve per l'aer tenebroso si riversa; (VI canto Inferno, vv. 10-11)'. Under 'Citazioni implicite', there are three entries: 'Almeno mi pareva di essere come Clacco stasera , pochi minuti fa, quando tornavo in albergo dopo aver camminato per ore al buio, perdon = domi e disperato per la campagna, sotto una pioggia continua, affondando nel fango e nella fanghiglia' (VI canto Inferno, vv. 52-54); 'stare sempre agitato come selvaggina che può essere colta di sorpresa.' (XIII canto Inferno, vv. 109-114); and 'l'altro sonava di continuo la tromba di Barbariccia. (XXI canto Inferno, vv. 139)'.

Figura 4.17: Visualizzazione delle tessere dantesche (testimonianza di Emanuele Artom)

esempio modelli *Sentence Transformers*.⁸⁸

Un'altra interessante e importante sezione dell'applicazione web è quella chiamata *Dante*. Cliccando su tale voce nel menù si accede ad una pagina in cui sono riportati i risultati emersi dalla prima analisi svolta sull'intero corpus circa la presenza delle tessere dantesche nelle testimonianze.

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 ben 11 testimoni su 18 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:

- 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

⁸⁸Cfr. <https://sbert.net/>.

estraendo e analizzando in dettaglio i dati relativi alle diverse forme mediante le quali le tessere dantesche si sono presentate nelle testimonianze.

- 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.

Di seguito vengono presentati i risultati emersi dalle due analisi.

Nelle testimonianze che costituiscono l'archivio sono state rinvenute 68 tessere dantesche, di cui 4 citazioni implicite, 15 citazioni esplicite, 10 allusioni e 39 vocaboli danteschi. In figura 4.18 è riportata la ripartizione delle tessere dantesche.

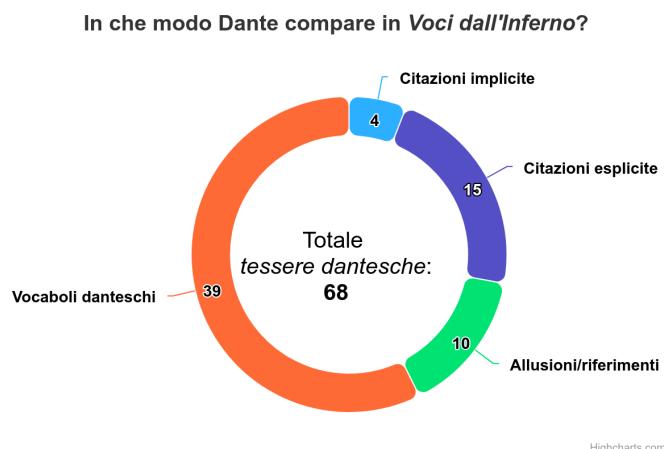


Figura 4.18: Totale tessere dantesche rilevate nel corpus

Ponendo l'attenzione esclusivamente sulle citazioni, i dati che sono emersi dallo studio delle testimonianze dell'archivio digitale, come descritto nel capitolo 1, 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.

All'interno dell'applicazione web è stato inserito un *bar chart* che mostra come sono ripartite le citazioni rispetto a Inferno, Purgatorio e Paradiso (figura 4.19).

Per approfondire l'analisi relativa invece ai canti più citati di ciascuna cantica, è stata aggiunta al grafico la funzionalità *drilldown* che permette di esplorare i dati cliccando sulla colonna di interesse. Come riportato in figura 4.20, ad esempio, cliccando sulla colonna *Inferno* il grafico cambierà la sua struttura mostrando quali sono i canti citati e quante volte occorrono, come visibile in figura 4.20. 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*

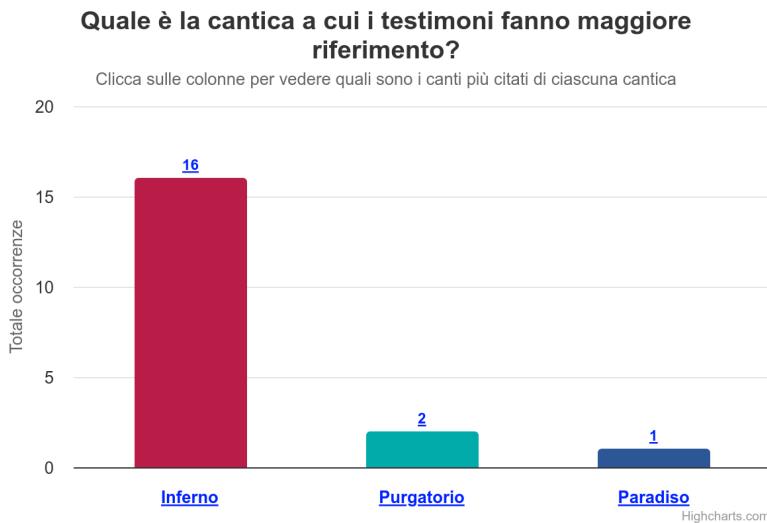


Figura 4.19: Ripartizione citazioni per cantiche

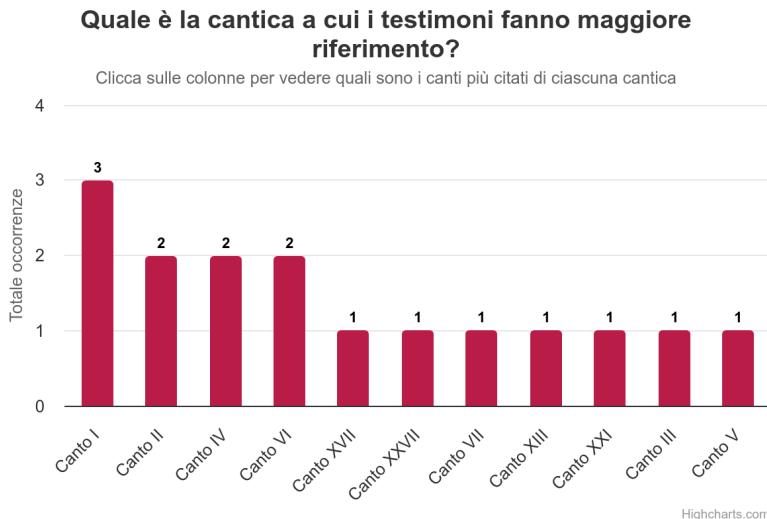


Figura 4.20: Ripartizione citazioni per canti dell'Inferno

XVII, *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.

I vocaboli danteschi che in fase di codifica sono stati individuati e marcati sono 39. All'interno dell'applicazione è stato realizzato il grafico di tipo *donut chart*, riportato in figura 4.21 per mostrare quali e quanti termini occorrono. 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.

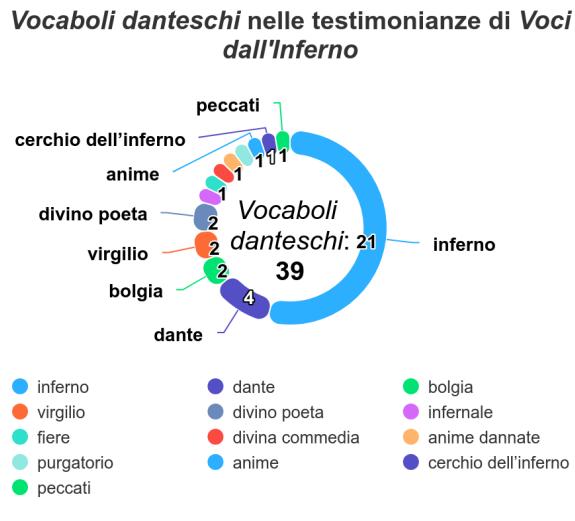


Figura 4.21: Vocaboli danteschi nelle testimonianze di *Voci dall'Inferno*

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 11 sono presenti tessere dantesche, come visibile dal grafico in figura 4.22. Luigi Giuntini è in assoluto colui che si serve maggiormente delle parole di Dante, nel cui diario si sono registrate 19 tessere dantesche. Numerosi sono i riferimenti rinvenuti anche nel diario di Nicola Ricci: 17. Gli altri testimoni che attingono maggiormente dalla *Commedia* sono Romana Feld e Goti Bauer, nelle cui testimonianze sono state individuati rispettivamente 7 riferimenti all'opera dantesca.

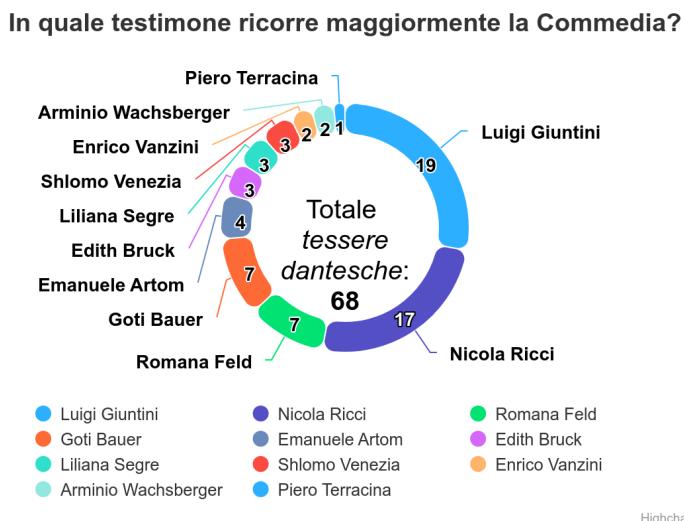


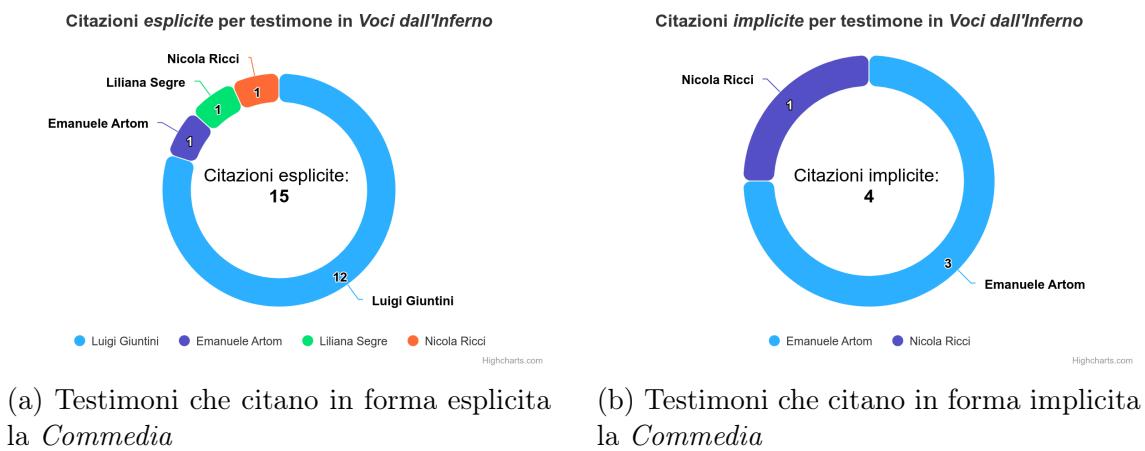
Figura 4.22: Testimoni che attingono maggiormente dalla *Commedia*

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, i cui risultati verranno illustrati di seguito, è stato creato un *donut chart*, per mostrare chi e quanto attinge dalla *Commedia*, ed è stata realizzata una tabella che riporta le citazioni, le allusioni e i termini identificati e registrati in fase di codifica.

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 2 citazioni, della terza, 1. Come visibile in figura 4.23a, Luigi Giuntini risulta colui che cita maggiormente in modo esplicito. Seguono, con 1 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, come visibile in figura 4.23b



(a) Testimoni che citano in forma esplicita la *Commedia*

(b) Testimoni che citano in forma implicita la *Commedia*

Figura 4.23: Chi, in che modo e in che misura si serve del vocabolario dantesco

Come accennato precedentemente, per approfondire l'analisi e consultare quali sono le citazioni, sia esplicite che implicite, registrate nelle testimonianze, sono state create delle apposite tabelle, come visibile in figura 4.24. Per ciascuna citazione viene riportato il relativo testimone, il titolo della testimonianza in cui è stata registrata, la cantica a cui si fa riferimento e il canto.

Le allusioni registrate nelle testimonianze del corpus sono 10 (Figura 4.25): 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.

Come mostrato in figura 4.26, per ciascuna allusione in tabella è riportato il relativo testimone, il titolo della testimonianza in cui è stata registrata e la cantica a cui si fa riferimento.

Cantica	Canto	Testimone	Titolo testimonianza	Citazione esplicita
Inferno	Canto XVII	Luigi Giuntini	I lunghi giorni della pena di Luigi Giuntini	« Le ripe eran grommate d'una muffa per l'alito di giù che vi s'appasta che con gli occhi e col naso facea zuffa »
Inferno	Canto XXVII	Luigi Giuntini	I lunghi giorni della pena di Luigi Giuntini	" contraddizion che noi consente "
Inferno	Canto I	Luigi Giuntini	I lunghi giorni della pena di Luigi Giuntini	« La notte ch'io passai con tanta pieta »
Inferno	Canto I	Luigi Giuntini	I lunghi giorni della pena di Luigi Giuntini	« 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 lunghi giorni della pena di Luigi Giuntini	« "Miserere di me" gridai a lui "qual tu sii, od ombra od omo certo?" »
Inferno	Canto II	Luigi Giuntini	I lunghi giorni della pena di Luigi Giuntini	« Lo giorno se n'andava, e l'aere bruno toglieva gli animali che sono in terra dalle fatiche loro »
Inferno	Canto IV	Luigi Giuntini	I lunghi giorni della pena di Luigi Giuntini	"la bella erba "

(a) Tabella delle citazioni esplicite

Cantica	Canto	Testimone	Titolo testimonianza	Citazione implicita
Inferno	Canto VI	Emanuele Artom	Diario di Emanuele Artom (novembre 1943-febbraio 1944)	Almeno mi pareva di essere come Ciacco stasera , pochi minuti fa, quando torna = vo in albergo dopo aver cam = minato per ore al buio, perden = domi e disperdoni per la campagna, sotto una pioggia continua, affondando nel fango e nella fanghiglia
Inferno	Canto XIII	Emanuele Artom	Diario di Emanuele Artom (novembre 1943-febbraio 1944)	stare sempre agitato come selvaggina che può essere colta di sorpresa.
Inferno	Canto XXI	Emanuele Artom	Diario di Emanuele Artom (novembre 1943-febbraio 1944)	l'altro sonava di continuo la tromba di Barbariccia.
Inferno	Canto III	Nicola Ricci	Diario di Nicola Ricci	Si va verso la fame, si va verso il freddo, si va verso l'inferno

(b) Tabella delle citazioni implicite

Figura 4.24: Tabelle relative alle citazioni

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. In figura 4.27 è mostrato il grafico in questione.

La tabella realizzata per i vocaboli danteschi presenta l'elenco dei testimoni che ricorrono al lessico di Dante, il titolo della testimonianza e l'elenco dei termini presenti in tale testimonianza (figura 4.28).

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. Come descritto nel capitolo 1, l'archivio dispone sia del racconto di chi visse l'esperienza del Lager sia di chi non fu mai deportato. Inoltre, tra coloro che hanno conosciuto il campo di concentramento sono presenti testimonianze rese da deportati ebrei e da internati militari italiani, mentre, di chi non visse il Lager, è attualmente presente solo la testimonianza del partigiano ebreo Emanuele Artom. Infine, le testimonianze sono state rese attraverso differenti modalità. L'eterogeneità, quindi, dei dati, sia dal punto di vista delle storie dei testimoni, sia in termini di tipologia e di provenienza delle fonti, ha portato allo svolgimento di un'analisi sull'intero corpus, i cui risultati sono stati riportati nella sezione *Statistiche*. Anche in questo caso sono stati creati degli appositi grafici per migliorare la fruizione dei dati.

Allusioni/riferimenti per testimone in Voci dall'Inferno

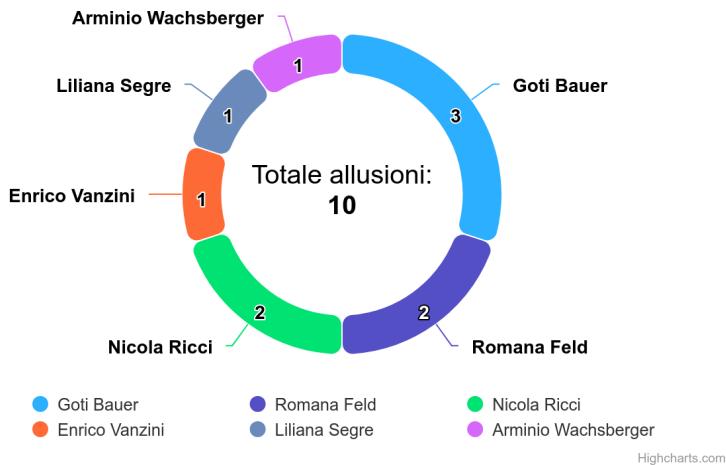


Figura 4.25: Testimoni che alludono alla *Commedia*

Cantica	Testimone	Titolo testimonianza	Allusione
Inferno	Enrico Vanzini	Testimonianza di Enrico Vanzini, l'ultimo Sonderkommando italiano	gli faccio come il segno della croce prima di buttarlo in questo inferno
Inferno	Romana Feld	Edizione digitale della testimonianza di Romana Feld	I quadri dell'inferno con i cortei delle anime dannate , diventavano realtà
Inferno	Romana Feld	Edizione digitale della testimonianza di Romana Feld	E' impossibile con le povere parole ridare quadro esatto dell'inferno vissuto
Inferno	Goti Bauer	Testimonianza di Goti Bauer	Liliana ha detto "Per me Goti è come Virgilio per Dante"
Inferno	Goti Bauer	Testimonianza di Goti Bauer	Liliana ci ha detto che l'ha conosciuta, che ha conosciuto lei, cui riconosce e ha ripetuto molte volte il titolo di Virgilio
Inferno	Goti Bauer	Testimonianza di Goti Bauer	Era pieno il campo e noi siamo arrivati in una bolgia

Figura 4.26: Tabella delle allusioni

Per prima cosa ci si è concentrati su chi sono i testimoni che attualmente popolano l'archivio: è stato quindi realizzato un *treemap chart* 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. In figura 4.29 è mostrato il grafico che è stato realizzato.

Successivamente è stato analizzato il tipo di testimonianze che costituiscono il corpus dal punto di vista della modalità attraverso le quali sono state restituite. Come descritto in precedenza, le fonti che costituiscono il corpus si dividono in testimonianze scritte e testimonianze orali. A sua volta, ciascuna macro-classe è suddivisa in diverse tipologie attraverso le quali è stata resa la testimonianza. È stato quindi realizzato un *donut chart* che mostra la suddivisione delle testimonianze nelle varie tipologie, come mostrato in figura 4.30.

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

Vocaboli danteschi per testimone in *Voci dall'Inferno*

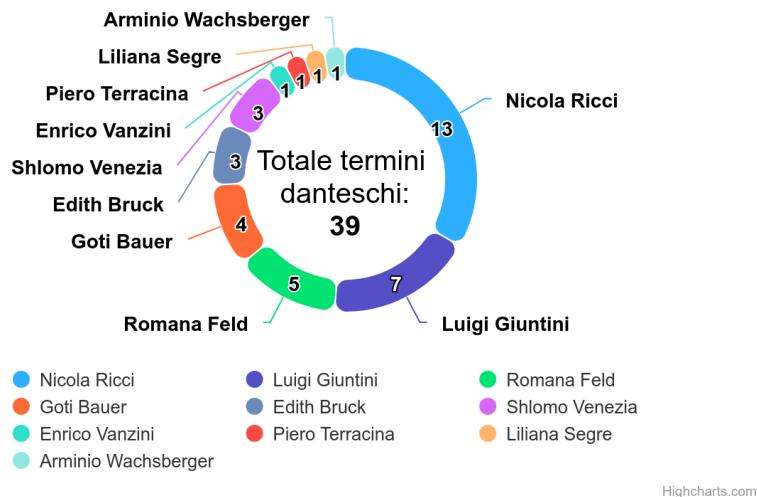


Figura 4.27: Testimoni che utilizzano vocaboli danteschi

Testimone	Titolo testimonianza	Vocaboli danteschi
Nicola Ricci	<i>Diario di Nicola Ricci</i>	inferno, Dante, inferno, Dante, divino poeta, inferno, inferno, inferno, divino poeta, anime, cerchio dell'inferno, peccati, inferno
Luigi Giuntini	<i>I lunghi giorni della pena di Luigi Giuntini</i>	inferno, inferno, bolgia, infernale, fiere, inferno , inferno
Romana Feld	<i>Edizione digitale della testimonianza di Romana Feld</i>	Divina Commedia, Dante, inferno, anime dannate, inferno
Goti Bauer	<i>Testimonianza di Goti Bauer</i>	Virgilio, Dante, Virgilio, bolgia
Edith Bruck	<i>Testimonianza di Edith Bruck</i>	inferno, inferno, inferno
Shlomo Venezia	<i>Shlomo Venezia: Testimonianza di un Sonderkommando</i>	inferno, inferno, purgatorio
Enrico Vanzini	<i>Testimonianza di Enrico Vanzini, l'ultimo Sonderkommando italiano</i>	inferno

Figura 4.28: Tabella dei vocaboli danteschi

testimonianze orali, 8 provengono da questo fondo. Delle restanti, le 3 in formato video provengono da YouTube,⁸⁹ 1 è stata realizzata da Marina Riccucci e 1 appartiene all'archivio della Fondazione CDEC. In figura 4.31a è mostrato il relativo grafico.

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, del quale sono stati digitalizzati 4 fogli manoscritti, di Nicola Ricci e Luigi Giuntini, dei quali sono stati codificati i 2 diari, e 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 Romana Feld appartengono dall'archivio digitale del Centro di Documentazione Ebraica Contemporanea di Milano, mentre la testimonianza di Bruno Cimoli proviene dalla sede centrale dell'ANPI di Massa. In figura 4.31b si può osservare la provenienza delle fonti scritte.

⁸⁹Sono le interviste rilasciate a Ivan Scalfarotto, Telepace e Associazione Figli della Shoah.

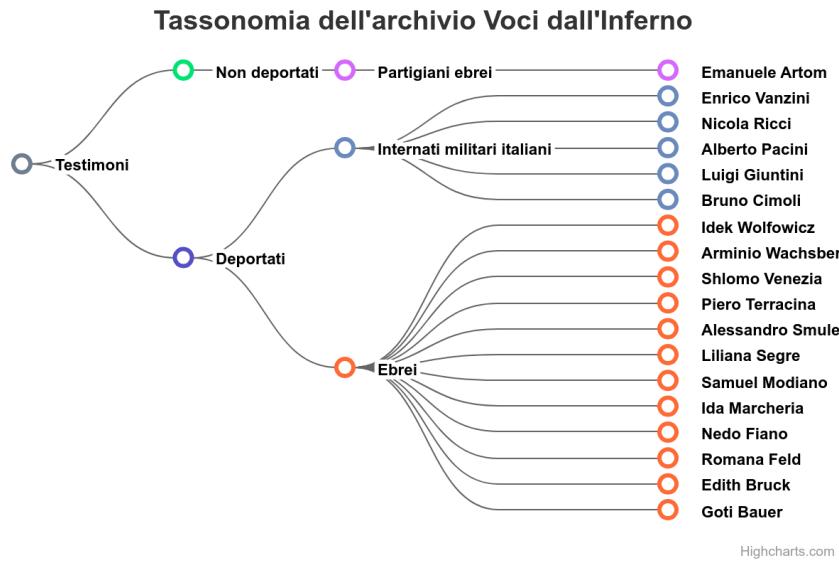


Figura 4.29: *Treegraph chart* che illustra la tassonomia completa dell'archivio

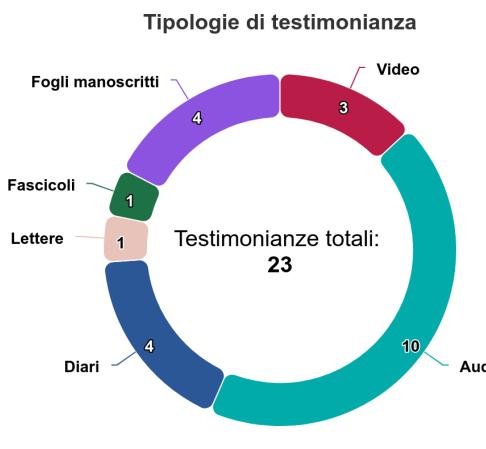


Figura 4.30: Tipologie di testimonianze del corpus

4.2.1 *Automatic speech recognition (ASR) e funzionalità di search all'interno del corpus*

È attualmente presente, ma ancora in fase di sviluppo, all'interno della sezione *Trascrizione automatica*, una prima implementazione di una funzionalità per effettuare l'*automatic speech recognition*. Come visibile in figura 4.32, cliccando sul tasto *Upload file* è possibile caricare un file audio del quale verrà restituita in automatico la trascrizione nell'area sottostante, che è possibile copiare mediante l'apposito tasto *Copia testo*. Nell'esempio mostrato è stata trascritta una frase della testimonianza di Sami Modiano.

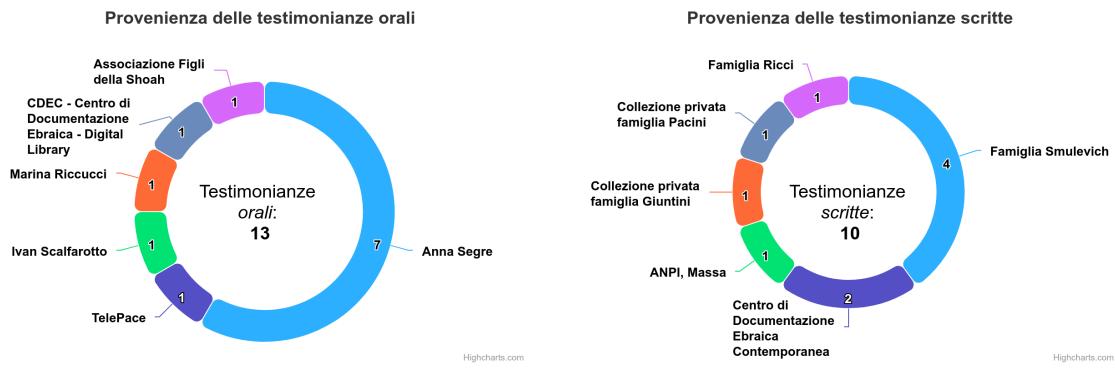


Figura 4.31: Provenienza delle fonti orali e scritte



Figura 4.32: Trascrizione automatica di un file audio

Per la realizzazione di tale funzionalità si è utilizzato *Whisper*,⁹⁰ un modello di intelligenza artificiale sviluppato da OpenAI⁹¹ 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*,⁹² addestrata su un numero di epoche 2,5 volte superiore rispetto alla prima versione. A novembre 2023, il modello *Whisper Large v3*,⁹³ 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, rendendolo più capace di riconoscere accenti meno comuni e

⁹⁰Cfr. <https://openai.com/index/whisper/>

⁹¹Cfr. <https://openai.com/>.

⁹²Cfr. <https://github.com/openai/whisper/discussions/661>.

⁹³Cfr. <https://github.com/openai/whisper/discussions/1762>.

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 HuggingFace⁹⁴.

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 3, 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 altri 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 riuscire ad 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'altra funzionalità in corso di sviluppo è quella che riguarda l'implementazione di funzionalità di *search* all'interno delle testimonianze del corpus. Riprendendo il lavoro svolto dalla collega Erika Deboni⁹⁵ sulla testimonianza di Ida Marcheria, verranno implementati ed estesi all'intero corpus di testimonianze almeno due tipi di ricerca:

- la ricerca di parole all'interno di tutte le testimonianze dell'archivio. L'utente potrà scegliere se cercare una parola esatta o se fare una ricerca per parole somiglianti. Verrà restituita a video la porzione di testo contenente quella determinata parola messa in evidenza.
- la ricerca di persone che vengono menzionate all'interno delle testimonianze del corpus. Come nel caso precedente, l'utente effettuerà una ricerca e verranno restituiti a video la porzione di testo nella quale viene citata la persona in questione.

⁹⁴Cfr. <https://huggingface.co/openai/whisper-large-v3>.

⁹⁵Cfr. Erika Deboni, *Una voce dall'inferno: la testimonianza inedita di Ida Marcheria* (elaborato finale: a.a. 2020-2021. Primo relatore: prof.ssa Marina Riccucci; correlatore: prof. Angelo Mario Del Grosso).

Conclusioni

Il lavoro di tesi illustrato e descritto nel presente elaborato, si è quindi svolto in due fasi distinte. La prima fase è consistita nella revisione e nell'armonizzazione dell'intero corpus di testimonianze codificate in formato XML-TEI dagli studenti che hanno collaborato al progetto nel corso degli anni, in modo tale da ottenere una struttura omogenea per tutti i dati che compongono l'archivio. La seconda fase del lavoro è consistita nella realizzazione dell'applicazione web *Voci dall'Inferno* che permette di raccogliere, di consultare, di analizzare e di interrogare tutte le testimonianze che costituiscono l'archivio. L'applicazione permette inoltre di censire la presenza di tesse dantesche all'interno del corpus e di analizzare le differenti tipologie attraverso le quali Dante si presenta nelle testimonianze, a seguito di una opportuna marcatura secondo i criteri di codifica stabiliti. Infine, è stato implementato e sperimentato un primo prototipo di funzionalità per il riconoscimento automatico del parlato.

Dalle prime domande che si posero Sara Calderini e Marina Riccucci nell'anno accademico 2015-2016, circa la presenza e la misura con cui Dante avesse fornito le parole per dire il Lager anche nelle testimonianze definite come *non letterarie*, il progetto *Voci dall'Inferno* ha raggiunto negli anni notevoli risultati. Mediante il supporto di strumenti informatici e la collaborazione in primis del professor Angelo Mario Del Grosso e poi di un vasto numero 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 web *Voci dall'Inferno* nella sua attuale implementazione è un punto di partenza per lo sviluppo di ulteriori numerose funzionalità. Le più significative, alcune delle quali in corso di sviluppo, riguardano:

- l'implementazione, per ciascun testimone, di un grafo etichettato e orientato che permetta di mettere in evidenza le relazioni tra le persone presenti all'interno di ciascuna testimonianza

- l'integrazione di una mappa che mostri i luoghi e gli spostamenti effettuati da ciascun testimone
- l'implementazione della funzionalità di *search* sull'intero corpus digitalizzato
- la creazione di un modello in grado di effettuare la trascrizione automatica del parlato e la marcatura di tutti i fenomeni che caratterizzano una fonte orale
- lo sviluppo di tecniche avanzate per il riconoscimento automatico dei riferimenti a Dante tramite l'uso di metodi per l'estrazione automatica di informazioni da risorse testuali
- l'allineamento nella visualizzazione della trascrizione di una testimonianza scritta con la relativa fonte primaria

Gli sviluppi futuri includono, inoltre, l'allineamento delle ricerche del progetto *Voci dall'Inferno* ai principi *FAIR*, che rappresentano uno standard di eccellenza per la gestione dei dati scientifici garantendo che essi siano rintracciabili (Findable), accessibili (Accessible), interoperabili (Interoperable) e riutilizzabili (Reusable). Questi principi guideranno lo sviluppo tecnologico del progetto, assicurando che i risultati siano accessibili e utili a lungo termine per un'ampia gamma di utenti e applicazioni.

Infine, ma non per importanza, il progetto *Voci dall'Inferno* contribuisce alla tutela e alla salvaguardia di un prezioso patrimonio storico che non deve essere dimenticato. Con il passare degli anni, il numero di coloro che possono e vogliono ancora rendere testimonianza di ciò che hanno vissuto durante il nazifascismo, sia dentro che fuori dal Lager, si riduce notevolmente. Inoltre, qualsiasi fonte primaria, sia scritta che orale, risente dell'usura del tempo e dell'avvento di nuove tecnologie. Un testo scritto può deteriorarsi e diventare difficilmente leggibile, allo stesso tempo anche una registrazione rischia lo stesso destino: il continuo mutare dei supporti e dei formati di registrazione, dovuti all'innovazione tecnologica, rendono obsolete le registrazioni, imponendo la continua necessità di conversioni in file audio compatibili con gli strumenti contemporanei.

È necessario quindi contribuire il più possibile all'arricchimento del *corpus* tramite la codifica di qualsiasi tipo di fonte, con lo scopo di rendere le testimonianze accessibili e interrogabili da chiunque, e di verificare la presenza di tessere dantesche al loro interno ma, soprattutto, di salvaguardare un patrimonio storico dal valore inestimabile.

Bibliografia

Agosti, G. *Nei lager vinse la bontà: memorie dell'internamento nei campi di eliminazione tedeschi*. Nostre missioni. Missioni estere dei Cappuccini, 1960.

Alighieri, Dante. *La Divina Commedia. Inferno*. A cura di Umberto Bosco e Giovanni Reggio. Firenze: Le Monnier, 1980.

Artom, Emanuele. *Diari di un partigiano ebreo, gennaio 1940-febbraio 1944*. A cura di Guri Schwarz. Torino: Bollati Boringhieri, 2008.

Baiolo, Chiara. «*La codifica del Diario partigiano di Emanuele Artom: sulle tracce di una biblioteca da studiare*». Università di Pisa, 2023.

Burnard, Lou. *Che cos'è l'iniziativa di codifica del testo?* . Marsiglia: OpenEdition Press, 2014. <https://doi.org/10.4000/books.oep.426>.

Calderini, Sara. «*L'ineffabilità della nefandezza*». Università di Pisa, 2015.

Calderini, Sara, e Marina Riccucci. «*L'ineffabilità della nefandezza: Dante "per dire" il Lager: un sondaggio preliminare nelle testimonianze non letterarie*». Italianistica: rivista di letteratura italiana: XLIX, 1, 2020, 2020. <https://doi.org/10.19272/202001301011>.

Caleffi, Pietro. *Un mondo fuori dal mondo: indagine DOXA fra i reduci dai campi nazisti*. Firenze: La Nuova Italia, 1971.

Ciotti, Fabio. *Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari*. Roma: Aracne, 2007.

Clifford B., Anderson, e Joseph Charles Wicentowski. *XQuery for Humanists*. First Edition. College Station, Texas, USA: Texas A&M University Press, 2020.

Deboni, Erika. «*Una voce dall'inferno: la testimonianza inedita di Ida Marcheria.*» Università di Pisa, 2020.

Del Grosso, Angelo Mario, Marina Riccucci, e Elvira Mercatanti. «*The Impact of Digital Editing on the Study of Holocaust Survivors' Testimonies in the context of Voci dall'Inferno Project*». In Proceedings of the First Workshop on Holocaust Testimonies as Language Resources (HTRes) @ LREC-COLING 2024, a cura di Isuri Anuradha, Martin Wynne, Francesca Frontini, e Alistair Plum, 1–9. Torino, Italia: ELRA and ICCL, 2024. <https://aclanthology.org/2024.htres-1.1>.

Giuntini, Luigi. *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, 2014.

Lelli, Diletta. «*Idek Wolfowicz: un testimone ritrovato*». Università di Pisa, 2022.

Levi, Primo. *Conversazioni e interviste (1963-1987)*. A cura di Marco Belpoliti. Einaudi, 1997.

Levi, Primo. *Opere complete*. A cura di Marco Belpoliti. Vol. 3. Einaudi, 2017.

Lévy-Hass, Hanna. *Diario Di Bergen-Belsen 1944-1945*. Milano: Jaca Book, 2018.

Montera, Greta. «*Stralcio sulla deportazione in Germania: la testimonianza di Bruno Cimoli. Codifica e analisi di una sezione del testo.*» Università di Pisa, 2021.

Petrilli, Irene. «*Dante fra gli Internati Militari Italiani: il caso ‘Luigi Giuntini’. Codifica preliminare del suo diario “I lunghi giorni della pena”*». Università di Pisa, 2021.

Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, e Ilya Sutskever. «*Robust Speech Recognition via Large-Scale Weak Supervision*».

Reimers, Nils, e Iryna Gurevych. «*Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*». In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), a cura di Kentaro Inui, Jing Jiang, Vincent Ng, e Xiaojun Wan, 3982–92. Hong Kong, China: Association for Computational Linguistics, 2019. <https://doi.org/10.18653/v1/D19-1410>.

Ricci, Pietro. «*Mio nonno deportato di Armenstein: il suo diario inedito.*». Università di Pisa, 2021.

Riccucci, Marina, e Laura Ricotti. *Il Dovere della parola. La Shoah nelle testimonianze di Liliana Segre e di Goti Herskovitz Bauer*. Pisa: Pacini Editore, 2021.

Segre, Anna, e Gloria Pavoncello, a c. di. *Judenrampe. Gli ultimi testimoni*. Antidot. Roma: Elliot, 2009.

Siegel, Erik, e Adam Retter. *Exist*. First. 1005 Gravenstein Highway North, Sebastopol, CA 95472.: O'Reilly Media, Inc, 2014.

Siragusa, Federico. «*I luoghi danteschi della memoria dei sopravvissuti*». Università di Pisa, 2023.

Sustersic, Federica. «*La Dicibilità Del Male. La Ricezione Dantesca Nelle Testimonianze Concentrazionarie*». Dante: Rivista internazionale di studi su Dante Alighieri 13 (2016): 57–78.

Taterka, Thomas. *Dante Deutsch. Studi sulla letteratura dei lager*. A cura di Enrico Paventi. Sette Città, 2002. Traversi, Valeria M. M. «Per Dire L'orrore: Primo Levi E Dante». Dante: Rivista internazionale di studi su Dante Alighieri 5 (2008): 109–25.

Walmsley, Priscilla. *XQuery*. Second Edition. 1005 Gravenstein Highway North, Sebastopol, CA 95472: O'Reilly Media, Inc., 2015.

Sitografia

«large-v3 release · openai/whisper · Discussion #1762».
<https://github.com/openai/whisper/discussions/1762>

«8 Transcriptions of Speech The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html>

«11 Manuscript Description The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>

«12 Representation of Primary Sources The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/it/html/PH.html#PH-bov>

«14 Names, Dates, People, and Places The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>

«16 Language Corpora The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CC.html>

«17 Linking, Segmentation, and Alignment The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html>

«18 Simple Analytic Mechanisms The TEI Guidelines».
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html>

«CDEC - Centro di Documentazione Ebraica - Digital Library»
<https://digital-library.cdec.it/cdec-web/>

«eXist-db Documentation».
<https://exist-db.org/exist/apps/doc/documentation>

GitHub. «Announcing the Large-v2 Model · Openai/Whisper · Discussion #661».

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

GitHub. «Whisper/Model-Card.Md at Main · Openai/Whisper».
<https://github.com/openai/whisper/blob/main/model-card.md>.

«Guidelines».

<https://tei-c.org/guidelines/>.

«Highcharts - Interactive Charting Library for Developers».
<https://www.highcharts.com/blog/homepage21may/>

«Hugging Face – The AI community building the future.»
<https://huggingface.co/>

«Introducing Whisper».

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

«Let Them Speak».

<https://lts.fortunoff.library.yale.edu/>

«OpenAI Platform».

<https://platform.openai.com>

«openai/whisper-large-v2 · Hugging Face»
<https://huggingface.co/openai/whisper-large-v2>.

«openai/whisper-large-v3 · Hugging Face»

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

«Saxon documentation».

<https://www.saxonica.com/html/documentation12/about/whatis.html>

«TEI: Text Encoding Initiative».

[https://tei-c.org/](https://tei-c.org)

«The TEI Guidelines».

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

United States Holocaust Memorial Museum. «*I campi di concentramento nazisti*».
[https://encyclopedia.ushmm.org/content\(it/article/nazi-camps](https://encyclopedia.ushmm.org/content(it/article/nazi-camps)

«Welcome to Apache Lucene».

<https://lucene.apache.org/index.html>

«XML Path Language (XPath) 3.1».

<https://www.w3.org/TR/xpath-31/>

«XQuery 3.1: An XML Query Language».

<https://www.w3.org/TR/xquery-31/>