



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

Idek Wolfowicz: un testimone ritrovato

Candidato: *Diletta Lelli*

Relatore: *Prof.ssa Marina Riccucci*

Correlatore: *Dott. Angelo Mario Del Grosso*

Anno Accademico 2022-2023

A Elena, il cui sorriso illumina ogni mia notte.
Ai miei genitori, Angela e Gabriele, per aver sempre creduto in me.

Indice

Introduzione	4
<i>1 Il testimone ritrovato: Idek Wolfowicz</i>	6
<i>1.1 La testimonianza e il testimone.....</i>	6
<i>1.2 Idek e la sua storia</i>	7
<i>2 Stefania Fedewicz.....</i>	17
<i>2.1 Biografia.....</i>	17
<i>2.2 Incontro tra le famiglie Wolfowicz e Fedewicz.....</i>	17
<i>3 la codifica della testimonianza.....</i>	22
<i>3.1 L'importanza della codifica</i>	22
<i>3.2 Passaggi preliminari</i>	25
<i>3.3 Linee guida per la rappresentazione dei fenomeni del parlato</i>	25
<i>4 L'applicazione</i>	28
<i>4.1 eXist-DB.....</i>	28
<i>4.2 Il linguaggio XQuery</i>	29
<i>4.3 Stadio iniziale</i>	31
<i>5 Standardizzazione della trascrizione del linguaggio parlato</i>	33
<i>5.1 Cenni preliminari.....</i>	33
<i>5.2 Gli strumenti di trascrizione</i>	33
<i>5.3 Convenzioni di trascrizione.....</i>	34
<i>5.4 Convenzioni utilizzate all'interno del progetto</i>	37
<i>6 La ricerca nel catalogo.....</i>	41
<i>6.1 La ricerca.....</i>	41
<i>6.2 Apache Lucene</i>	41
<i>6.3 teiHeader.....</i>	44
<i>6.4 Realizzazione della ricerca.....</i>	46
<i>7 Conclusioni e sviluppi futuri</i>	49
Bibliografia	51
Sitiografia.....	53

Introduzione

Il presente lavoro di tesi triennale nasce all'interno e nell'ambito del progetto di ricerca *Voci dall'Inferno* dell'Università di Pisa. Il progetto, coordinato dalla Professoressa Marina Riccucci, si pone come obiettivi principali la creazione di un *corpus* digitalizzato di testimonianze non letterarie e inedite di sopravvissuti ai Lager e lo studio del lessico dantesco presente in esse.

Il primo obiettivo che si pone *Voci dall'Inferno* è quello della creazione di un corpus digitalizzato contenente testimonianze inedite e non letterarie come, per esempio, diari, lettere e interviste, tanto per conservarle quanto per potervi svolgere attività di studio e di ricerca. Un altro obiettivo del progetto *Voci dall'Inferno* è quello di rintracciare la presenza del lessico dantesco all'interno delle testimonianze raccolte: si è notato che Dante Alighieri (1265-1321), attraverso la sua opera, ha fornito un vero e proprio vocabolario al quale poter attingere quando le parole per descrivere la realtà sembrano essere poco efficaci. Il legame tra il lessico dantesco e il lessico utilizzato dai testimoni nasce quindi dal bisogno di esprimere e dar nome a un orrore che mai prima era esistito. Ad oggi il progetto procede grazie al lavoro di molti laureandi che collaborano per arricchire il patrimonio archivistico e grazie al supporto del CISE (Centro Interdipartimentale di Studi Ebraici), del dipartimento di Filologia, Letteratura e Linguistica dell'Università di Pisa, del CDEC (Centro di Documentazione Ebraica Contemporanea) e dell'Istituto di Linguistica Computazionale "A. Zampolli" di Pisa. Alle origini del progetto, per la conservazione delle testimonianze raccolte è stato implementato dalla Dottoressa Frida Valecchi una banca dati digitale chiamata *Memoriarchivio*. Nella sua architettura generale, si tratta di un'applicazione web che permette di gestire e consultare le testimonianze, caricare ed elaborare testi codificati in XML-TEI e analizzare e confrontare i testi.

Io ho lavorato sulla testimonianza inedita di Idek Wolfowicz, raccolta durante un'intervista condotta da Anna Segre e Gloria Pavoncello nel 2006.

Ho codificato in formato digitale le parole del testimone.

La codifica digitale della testimonianza è stata realizzata mediante l'utilizzo dello schema e delle linee guida XML-TEI, ovvero uno standard di codifica del testo basato sul linguaggio a marcatori XML (eXtensible Markup Language) progettato per rappresentare e conservare dati complessi. La TEI, invece, è un'iniziativa

internazionale nata nel 1987, il cui obiettivo principale è quello di fornire linee guida condivise per la rappresentazione digitale dei testi d'interesse umanistico che consentano l'archiviazione, l'elaborazione e lo scambio di informazioni testuali mediante uno schema formalmente definito. XML-TEI è stato quindi utilizzato per la registrazione dei metadati e per l'esplicitazione e l'analisi dei fenomeni testuali. Inoltre, sono stati utilizzati i fogli di stile XSL per permettere l'elaborazione dei documenti XML.

Ho poi lavorato su una sezione di search avanzato che permette di fare ricerche nel catalogo mediante l'utilizzo di metadati presenti nel `teiHeader` dei file XML e sulla visualizzazione della trascrizione delle testimonianze con specifiche convenzioni di trascrizione.

Capitolo I

Il testimone ritrovato: Idek Wolfowicz

1.1 La testimonianza e il testimone

Nel 2006 Anna Segre e Gloria Pavoncello¹ intervistano, presso la sua abitazione a Roma, Idek Wolfowicz. In quel periodo Segre e Pavoncello stanno raccogliendo le testimonianze di alcuni sopravvissuti alla Shoah che vivono in Italia: questo lavoro esiterà nel volume *Judenrampe*.²

Segre e Pavoncello sono entrate in contatto con Idek grazie alla figlia, Miriam Wolfowicz, poiché questa conosceva personalmente Anna Segre. Idek parla e Segre e Pavoncello registrano. La testimonianza ‘consiste’ in due microcassette TDK MC-90: le cassette sono state oggetto del lavoro informatico che ho condotto e di cui rendo conto nel capitolo secondo. Segre e Pavoncello non sapevano quello che solo questo lavoro di tesi ha permesso di sapere: Idek era, a tutti gli effetti, un *testimone ritrovato*. Cercando informazioni su Idek sono entrata in contatto con Laura Brazzo³, vicedirettore della Fondazione CDEC⁴ (Centro di Documentazione Ebraica Contemporanea) dal 2022 e ho scoperto che non esistono informazioni su Idek all’interno della banca dati del CDEC. All’interno degli archivi Arolsen (Centro internazionale di Documentazione, Informazione e Ricerca sulla persecuzione nazista, il lavoro forzato e l’Olocausto nella Germania nazista e nelle regioni occupate)⁵ si trovano invece alcuni documenti su Idek, sulla sua vita e che testimoniano della sua prigionia nei lager nazisti⁶. Esiste inoltre una video-intervista rilasciata da Idek a Sira Fatucci, dell’Unione delle Comunità Ebraiche Italiane e membro dell’IHRA⁷

¹ Istituzione del fondo “Segre-Pavoncello”

² Tra le testimonianze raccolte da Anna Segre e Gloria Pavoncello ci sono quelle di ebrei catturati in Italia, a Fiume, a Rodi, in Grecia, in Ungheria, in Croazia, in Libia, oppositori politici e militari sconfitti, ci sono le loro voci in presa diretta che scavano nella memoria per restituirci non l’accurata ricostruzione storica di quegli eventi, ma gli istanti in cui persero la loro innocenza (<https://www.annasegre.it/judenrampe/>)

³ <https://www.cdec.it/team/laura-brazzo/>

⁴ <https://www.cdec.it/>

⁵ <https://arolsen-archives.org/en/>

⁶ <https://collections.arolsen-archives.org/en/archive/3-2-1-203020102-058-022>

⁷ L’IHRA unisce governi ed esperti per rafforzare e promuovere l’educazione, la ricerca e la memoria sull’Olocausto e per sostenere gli impegni assunti con la Dichiarazione di Stoccolma del 2000. La rete

(International Holocaust Remembrance Alliance). Ho potuto disporre di questa intervista perché me l'ha fornita il figlio di Idek, il sig. William (citerà stralci indicandola come *Intervista a Fatucci*). Purtroppo non è nota la data a cui questa intervista risale.

Racconterò e ripercorrerò la biografia di Idek utilizzando le informazioni che ho raccolto attraverso colloqui con i familiari di Idek, consultando i documenti sopracitati e le parole che lo stesso Idek ha consegnato sia a Segre e Pavoncello che a Fatucci.

1.2 Idek e la sua storia

Idek Wolfowicz nacque a Tarnów,⁸ in Polonia, il 3 aprile 1922.

Idek non prese il cognome dal padre, Wolf Schwarzbach, ma dalla madre, Miriam Wolfowicz. Come Idek spiega nella video intervista realizzata da Sira Fatucci, Wolf Schwarzbach era un sottoufficiale dell'esercito austriaco che aveva il compito di sorvegliare il campo profughi all'interno del quale si trovava Miriam Wolfowicz.

A quel tempo Wolf era sposato.

Dopo aver conosciuto Miriam, che aveva 20 anni meno di lui, ed essersi innamorato di lei, Wolf divorziò, lasciando la prima moglie – dalla quale aveva avuto tre figli -, e si sposò con Miriam.

Mia mamma era la seconda moglie di mio papà. Il mio papà ha divorziato, si era innamorato di mia mamma perché nella prima guerra c'era un raccolto dei... Li chiamavano profughi, e mia mamma era una di quelli e mio papà era nell'esercito austriaco. Era sottoufficiale dell'esercito, perché aveva 20 anni in più della mamma, e aveva la moglie e tre figli.⁹

Nell'intervista condotta da Segre e Pavoncello, Idek racconta che il matrimonio dei genitori fu celebrato solo dal rabbino, motivo per cui non fu riconosciuto dalle autorità polacche. Per questo motivo e per il risentimento che Idek provava nei confronti del

di esperti fidati dell'IHRA condivide le proprie conoscenze sui primi segnali di allarme del genocidio odierno e sull'informazione sull'Olocausto. (<https://www.holocaustremembrance.com/>)

⁸ <https://www.tarnow.pl/>

⁹ *Intervista a Fatucci*.

padre per aver abbandonato la prima famiglia, soprattutto dopo averla conosciuta, Idek preferì mantenere il cognome della madre:

io con mio papà ho avuto certi discorsi, leggeri, perchè non potevo criticarlo però non mi andava bene, non mi è andato ancora di più bene quando ho conosciuto mio fratello, poi le mie sorelle e la prima moglie che mi ha accolto [...] più che un figlio.¹⁰

A completare il quadro familiare della famiglia Wolfowicz abbiamo anche due sorelle: Sara, la maggiore, e Rivka, la minore.

Gli unici due sopravvissuti alle atrocità naziste saranno Idek e Sara.



Figura 1: famiglia Wolfowicz

Prima dell'invasione tedesca della Polonia, Idek frequentava due scuole: la scuola statale, la mattina, e la scuola ebraica, il pomeriggio.

Come racconta egli stesso nella video intervista realizzata da Sira Fatucci: «io la mattina andavo alla scuola statale, il pomeriggio andavo a [. . .] la scuola ebraica che studiavo ebraico perchè da noi era questa usanza».

Nell'ottobre 1939, all'età di diciassette anni, Idek fu costretto a fuggire dalla città natale, Tarnów, per dirigersi dalla prima famiglia del padre, verso il confine orientale.

¹⁰ *Intervista a Fatucci.*

In un primo momento Idek non voleva lasciare la propria casa: fu convinto dalla madre, che lo spinse a scappare per salvarsi.

Miriam, la madre, non poteva immaginare ciò che sarebbe successo in seguito: a causa del patto stretto tra Germania e Unione Sovietica,¹¹ Idek si ritrovò circondato.

Wolf, il padre, fu fucilato nella piazza di Tarnów, nel 1939. Idek racconta nell'intervista rilasciata a Segre e Pavoncello di non aver assistito alla sua morte poiché era già fuggito: «mio padre è stato fucilato nella piazza della mia città ma non l'ho visto perchè io non c'ero».

Idek non fuggì da solo: era insieme alla propria fidanzata.

Come racconta egli stesso nella video intervista realizzata da Sira Fatucci:

chi pensava che poi avevano un patto [...] che poi da questa parte avanzavano i tedeschi da questa invece è successo che i russi andavano avanti però non si sapeva ancora, così io ho lasciato la casa però l'ho lasciata con una ragazza.

Ma Idek e la sua ragazza non furono gli unici a fuggire verso il confine orientale: come loro fuggirono centinaia di ebrei. Si trattava prevalentemente di giovani, i quali venivano mandati dai genitori verso il confine. Ma, a differenza di quanto si possa pensare, non formavano un gruppo unito: Idek racconta che ognuno stava per conto suo e cercava di salvarsi:

Per conto loro ognuno cercava di scappare perché, specialmente i genitori dalle mie parti, cercavano di mandare via i giovani perchè dicevano che i tedeschi li prendevano e facevano tante cose eccetera. Allora facevano andare via quando uno c'aveva parenti o cose andavano. Ma non è che eravamo legati insieme ognuno era per conto suo cercava di salvarsi io mi sono salvato più volte per conto mio altri sono morti.

Con l'avanzata dell'Unione Sovietica, la ragazza di Idek, per paura, decise di tornare verso casa. Idek si rifiutò di tornare verso Tarnów e decise di continuare la fuga e di arrivare dalla famiglia del padre. Una volta arrivato dalla famiglia del padre, Idek

¹¹ Il patto tedesco-sovietico fu firmato nell'agosto del 1939 e aprì la strada all'invasione e all'occupazione congiunta della Polonia da parte della Germania nazista e dell'Unione Sovietica, nel settembre dello stesso anno. (<https://encyclopedia.ushmm.org/content/it/article/german-soviet-pact>)

conobbe la prima moglie e le due figlie: il figlio lo aveva già conosciuto poiché egli aveva vissuto a Tarnów con la famiglia di Idek per un certo periodo di tempo.

Idek racconta di essere stato trattato come un figlio dalla prima moglie del padre: «“tu sei per me il secondo figlio”. Io c’avevo la commozione, ecco, perchè c’avevo un senso di dispiacere da morire contro il papà perchè era una persona bellissima, una ragazza, donna con tre figli».¹²

Idek rimase con loro per alcune settimane poiché suo cugino, che era un membro della polizia ebraica, lo avvertì che era stato stilato un elenco di tutti gli ebrei che sarebbero stati rastrellati nella nottata e che questi, una volta catturati, sarebbero stati portati in Siberia. Idek, temendo la Siberia, decise di tornare verso casa.

Come racconta nella video intervista:

Avevo un cugino che era dentro agenti di polizia. È venuto e ha detto: “senti abbiamo fatto un elenco, stanotte veniamo a rastrellare tutti e facciamo un treno e vi mandiamo in Siberia, perciò, tu o vai in Siberia o scappa vai via” [...] e io ho fatto marcia indietro per tornare a casa.¹³

Nel giugno 1940, Idek riprese il cammino verso la propria città natale, Tarnów. Durante il tragitto si fermò in una cittadina chiamata Przemyśl:¹⁴ «Ho proseguito per il ritorno a casa e sono arrivato in una città che si chiama Przemyśl».¹⁵

In questa città conobbe Stefania Fedewicz (della quale mi occuperò tra poco, in questo stesso capitolo, § 3).

Idek conobbe Stefania aiutandola a portare delle buste. Dopo essersi confidato con lei ed averle raccontato di essere ebreo, Stefania gli propose di nascondersi all’interno del giardino della propria abitazione e di occuparsi della propria mucca.

Il giardino di Fedewicz confinava con il ghetto della città di Przemyśl: per questo motivo e perchè altri ebrei vi si nascondevano, vi venivano fatte molto spesso retate.

In entrambe le interviste rilasciate, Idek racconta che Stefania Fedewicz era la moglie del vicesindaco di Przemyśl e che il loro figlio faceva parte delle SS e che per questo motivo in quel periodo non si trovava a casa. Dai colloqui con William Wolfowicz, figlio di Idek, è emerso che si tratta di un errore: Stefania non era la moglie del

¹² *Intervista a Fatucci.*

¹³ *Intervista a Fatucci.*

¹⁴ <https://przemysl.pl/70302/miasto-przemysl-strona-oficjalna.html>

¹⁵ *Intervista a Fatucci.*

xslvicesindaco, bensì un'insegnante che viveva vicino al ghetto di Przemyśl. Suo marito era a sua volta un insegnante: inoltre il figlio non faceva parte delle SS.

Quello con Stefania Fedewicz non fu il solo incontro fortunato fatto da Idek durante la sua permanenza a Przemyśl: occupandosi della mucca di Fedewicz e portandola ogni mattina al fiume San, Idek incontrò un pescatore.

Idek, che era molto agile e che amava arrampicarsi sugli alberi, dopo aver visto che l'uomo stava riscontrando difficoltà nel pescare, si offrì di aiutarlo catturando alcuni insetti da utilizzare come esche.

Idek non poteva saperlo, ma l'uomo che aveva appena aiutato era un comandante delle SS. Come racconta Idek nella video intervista realizzata da Sira Fatucci:

Al fiume ci portavo la mucca, io intanto mi arrampicavo sugli alberi così ho incontrato uno che stava pescando, un tipo alto eccetera in borghese e che non riusciva a pescare nessun pesce. Io gli ho detto: "guarda vedi questo albero vedi questi [...] ai pesci piacciono". Io ho raccolto 10 15 di questi qui, lui li attaccava e con questi qui pescava.

E nell'intervista rilasciata a Segre e Pavoncello:

Io ogni mattina mungevo la mucca, andavo vicino a qualche centinaio di metri c'era il fiume San [...] e là ho incontrato un signore in borghese che pescava ma non riusciva a pescare [. . .] acchiappavo questi qui e come esca pescava sempre, allora mi ha preso molto a benvolere [. . .] però mi ha salvato la vita.

Tra i due si venne a creare un rapporto di amicizia: Idek racconta che il comandante, di cui non ricorda il nome, mangiava sempre un sandwich e che glielo offriva.

Una notte, a causa di una retata fatta all'interno del giardino di Stefania, Idek fu catturato e portato davanti al comandante delle SS: per la prima volta Idek venne a conoscenza della vera identità dell'uomo che aiutava ogni giorno a pescare. Il comandante, dopo aver riconosciuto Idek, lo liberò, salvandogli la vita.

Lui mangiava sandwich e mi dava anche da mangiare sandwich. Quando è successo una notte hanno fatto la retata e hanno fatto anche retata nel giardino eccetera, quando mi hanno preso mi hanno portato davanti al comandante e chi

era il comandante? Era lui, era in divisa da ufficiale [...] e a me ha detto:” ma tu piccolo cosa ci fai qua?” Eh, si vede che ero piccolino in confronto e mi ha mandato via e quella volta mi sono salvato.

Idek si salvò altre volte prima di essere catturato definitivamente e caricato su un camion diretto a Plaszow. Il camion fece però tappa a Tarnów. Miriam, la madre di Idek, da otto giorni si era messa sulla piazza, convinta che avrebbe visto suo figlio. Così racconta Idek: «Mia mamma si era messa sulla piazza, proprio quella piazza, seduta, da 8 giorni non voleva andare a dormire a casa diceva: “mio figlio io lo voglio vedere e lo vedrò”. Sembrano cose impossibili e mi ha trovato che sono arrivato»¹⁶. Grazie all’aiuto del genero che riuscì a corrompere le guardie che sorvegliavano la cantina in cui era stato rinchiuso Idek offrendo loro una bottiglia di vodka, Miriam riuscì a rivedere il figlio per qualche ora. Come racconta Idek a Segre e a Pavoncello: «ho visto la mamma l’ultima volta poi. Un domani ho proseguito il viaggio e ci hanno portato a Cracovia, a Plaszow».

Al tempo dell’intervista realizzata da Sira Fatucci Idek racconta di non avere idea di che fine avesse fatto la madre: «questa era l’ultima volta che ho visto la mamma mia, non l’ho più vista e non so fino ad adesso che fine ha fatto».

La mattina dopo Idek fu portato al campo di Plaszow.¹⁷

Idek racconta di essere stato fortunato perchè non fu rinchiuso subito nel campo di Plaszow, ma fu mandato in un altro piccolo campo che si trovava a due km di distanza. Questo era il campo in cui si trovava anche un uomo di nome Schindler di cui Idek parla, ma solo a Fatucci:

Ero fortunato nella disgrazia [...] nel senso che quando siamo arrivati c’era Plaszow proprio campo di concentramento di Plaszow che era un vecchio cimitero ebraico che poi sono stato anche là e poi c’era a 2 km di distanza un altro piccolo campo di concentramento recintato [...] era molto prezioso stare qua perché, come ammazzavano subito molti prigionieri a Plaszow qui ci salvavamo perchè si lavorava si facevano le pentole e c’era Schindler che

¹⁶ *Intervista a Fatucci.*

¹⁷ Il campo di concentramento di Krakow-Plaszow si trovava presso il sobborgo di Plaszow, nella parte meridionale della città di Cracovia, in Polonia. In un primo momento si trattava di un campo di lavoro forzato nazista che forniva manodopera a diverse fabbriche di armamenti e ad una cava di pietra che si trovava nelle vicinanze, divenne un campo di concentramento soltanto nel gennaio del 1944.

protegeva, però quelli davano i soldi anche là i connazionali, quelli di Cracovia, che venivano poi presi davano i soldi così cosa succedeva? Levava da qua e mandava a Plaszow e quelli li metteva dentro e si salvavano invece noi no. Qui tutti i giorni morivano.¹⁸

Dopo alcune settimane in questo campo, Idek fu trasferito a Plaszow, Era il 1940. Idek racconta a Segre e Pavoncello che il campo di Plaszow si trovava su un vecchio cimitero ebraico dal quale veniva prelevato il marmo delle lapidi che veniva poi utilizzato per costruire le strade: «Plaszow era un ex cimitero che si rompeva [. . .] le lapidi e si facevano le strade [...] con il marmo delle lapidi si facevano le strade». All'interno del campo di Plaszow si trovava un altro piccolo campo centrale, recintato, riservato ai patrioti polacchi, i quali venivano imprigionati e giustiziati: Idek era stato preso nella squadra che si occupava di seppellire i patrioti giustiziati e di disinfettare le fosse comuni.

Dentro di lì nel campo di Plaszow c'era un piccolo campo recintato dove mettevano dentro non ebrei ma i famosi combattenti polacchi che erano molto forti, non gli importava niente che quando li portavano sulla montagnola [. . .] e li fucilavano e io ero preso per fortuna mia tra quelli a seppellirli a buttare il come si chiama? Lo chiamavano disinfettante. E questi polacchi andavano a morire e cantavano, gli ebrei no, non cantavano mai.¹⁹

Idek racconta a Sira Fatucci, che, per salvarsi, ognuno doveva arrangiarsi in qualche modo: Idek riusciva sempre a trovarsi qualche lavoro. Tra i vari lavori fatti racconta di aver lavorato all'interno di una struttura simile a un ospedale, dove dice di avere visto morire molti ebrei che, invece di essere curati, ricevevano iniezioni di acqua:

Là dentro nel campo di Plaszow [...] ognuno doveva arrangiarsi in qualche maniera: io riuscivo sempre a infilarmi a lavorare, per esempio poi ho avuto la fortuna, da una parte disgrazia a vedere certe cose così, in una specie di ospedale lo chiamavano dove facevano morire tanta gente ebrea. Sai cosa

¹⁸ *Intervista a Fatucci.*

¹⁹ *Intervista a Fatucci.*

facevano? gli facevano iniezione di acqua e questa non poteva vivere io stavo dentro a pulire.²⁰

Grazie a questo lavoro, però, Idek ebbe modo di conoscere molto bene la struttura del campo. Ciò lo aiutò quando fu emanato l'ordine di rastrellare il campo: Idek riuscì a scappare e a nascondersi all'interno di un palazzo vuoto con delle grandi travi nel sottotetto. Idek, grazie alla sua agilità, riuscì ad arrampicarsi ed a nascondersi su queste travi, scampando a morte certa:

Un bel giorno è venuto un ordine di rastrellare e mi sono trovato in mezzo io e, conoscendo tutte indicazioni perchè andavo a pulire, c'era un palazzo vuoto però c'era un tetto e sotto al tetto c'erano i travi larghi [. . .] e io sono riuscito, come una scimmia, cosa vuol dire la forza per salvezza, ad arrampicarmi e mi sono messo su una trave che era larga così. Loro sono entrati, bestemmiavano, gli aguzzini dentro, le SS, eccetera hanno trovato due tre eccetera, però me no eppure guardavano ma io stavo proprio su questa trave per fortuna mia anche con lampade sopra non mi vedevano. Mi sono salvato la vita anche quella volta.²¹

Altre volte Idek è riuscito a salvarsi, finché, nell'ottobre 1943, il campo di Plaszow venne evacuato: i prigionieri dovevano essere trasferiti ad Auschwitz,²² come racconta Idek a Segre e Pavoncello: «in un certo momento evacuavano Plaszow [...] e ci portavano con il treno verso Auschwitz [...] ci hanno caricato sui treni e ci hanno portato ad Auschwitz, siamo stati tre giorni e tre notti all'ingresso di Auschwitz però non c'era posto».

Nell'intervista rilasciata a Fatucci Idek racconta che, all'interno del proprio vagone, ricorse a un *escamotage* per evitare liti per l'acqua tra i prigionieri: utilizzava un cucchiaino per dare la stessa porzione di acqua ad ogni prigioniero.

²⁰ *Intervista a Fatucci.*

²¹ *Intervista a Fatucci.*

²² Auschwitz era un vasto complesso di campi di concentramento e di sterminio situato nella cittadina polacca di Oswiecim. Il complesso divenne il più grande mai realizzato.

Io nel mio vagone sono diventato il comandante, cosa vuol dire? Che ho escogitato io un cucchiaino e ho detto no dopo di che mi volevano tutti bene [. . .] prendevo l'acqua, per non dare uno tanto come succedevano altri vagoni che poi si menavano e la rovesciavano l'acqua e rimanevano senza acqua e morivano più io davo ognuno 2-3 cucchiaini di acqua.²³

Auschwitz era, come racconta Idek, pieno, per questo motivo i prigionieri attesero diversi giorni all'interno dei vagoni, fuori dal campo, per poi essere trasferiti al campo di Mauthausen.²⁴

Durante l'intervista a Segre e a Pavoncello Idek racconta che tutto ciò è avvenuto in estate: mentre nei documenti presenti all'interno degli archivi Arolsen il trasferimento a Mauthausen viene datato 'ottobre 1943'. Non si ha quindi la certezza di quando sia avvenuto di preciso questo fatto.

Idek racconta a Segre e a Pavoncello che, una volta arrivato a Mauthausen, vide che lì i bambini austriaci gettavano alimenti contenenti vetri ai prigionieri del campo affamati in quali, mangiandoli, andavano incontro alla morte: «cosa facevano ci buttavano le patate ma dentro alle patate cosa c'erano? Rompevano le bottiglie e c'erano dentro i vetri».

Dopo tre settimane passate nel campo di Mauthausen, Idek fu trasferito in un sotto-campo denominato Gusen 1: qui conobbe un gruppo di testimoni di Geova, ai quali si affiancò e si unì, fingendosi uno di loro. Ai testimoni di Geova venivano affidate mansioni meno pesanti e privilegiate rispetto a quelle che venivano imposte agli ebrei: ottenne così un lavoro all'interno di una cava.

Con l'arrivo degli americani nel 1945, le SS evacuarono anche il sotto-campo di Gusen 1: Idek fu trasferito nuovamente, questa volta nel vicino campo di Wels, un campo di concentramento destinato agli ebrei ungheresi. Fino alla liberazione, avvenuta per mano degli Americani, Idek resterà prigioniero in questo campo: al momento della liberazione peserà 37 kg.

Il 5 maggio fu fatto nuovamente prigioniero e rinchiuso in un campo vicino Linz, dal quale fu poi liberato una volta riconosciuto. Definitivamente libero, Idek si diresse

²³ *Intervista a Fatucci.*

²⁴ Mauthausen era un campo di concentramento nazista costruito nel 1938, noto per la sua alta mortalità. Inizialmente usato per recludere oppositori politici, divenne un luogo di sfruttamento brutale della manodopera, soprattutto nell'industria bellica.

verso il confine dove salì a bordo di un camion proveniente dalla Palestina e diretto a Bologna. Una volta giunto a Bologna, proseguì fino a Roma, dove si stabilirà e dove rimarrà fino alla morte, avvenuta il 12 ottobre 2017.

L'unica altra superstite della famiglia Wolfowicz fu Sara, la sorella maggiore di Idek, la quale si trasferì a Rochester New York, dove divenne la direttrice di un hotel che Idek ricorda come luogo nel quale si trovava il bagno rituale (particolare, questo, riferito a Segre e a Pavoncello: in questa intervista Idek dice che la sorella è morta, ma non fornisce la data).

Capitolo II

Stefania Fedewicz

2.1 Biografia

Stefania Fedewicz nacque il 16 agosto 1890 a Przemyśl, in Polonia. All'epoca la città faceva parte dell'impero Astro-ungarico. Stefania si immatricolò nel 1908 iscrivendosi alla facoltà di Medicina dell'Università di Leopoli, in Ucraina: fu una delle prime tre donne ammesse a studiare medicina. Poco dopo abbandonò gli studi medici e si iscrisse alla Facoltà di Lettere Classiche: conseguita la laurea, andò a insegnare presso l'Istituto per ragazze ucraine di Przemyśl, dove lavorò per decenni.

Nel 1919 Stefania sposò il suo insegnante di latino, Mykhailo Fedewicz: anche lui insegnante, ma nella sezione in lingua ucraina del liceo maschile di Przemyśl. Il 3 marzo 1921 nasce il loro unico figlio, Roman.

Alla morte del marito, nel 1959, Stefania si trasferì in Australia, dove Roman era andato a vivere: lì rimase fino alla morte, avvenuta nel 1976.

2.2 Incontro tra le famiglie Wolfowicz e Fedewicz

La nipote di Stefania, la figlia di Roman, Daria Fedewytsch-Dickinson, conosce molto bene la storia di Idek perchè la nonna ha sempre parlato del ragazzo a cui aveva salvato la vita facendolo nascondere nel giardino della propria abitazione a Przemyśl. Daria lo racconta in una mail indirizzata al figlio di Idek, William Wolfowicz, alla quale allegò un documento contenente la storia della nonna Stefania. Daria aveva messo per iscritto la storia della nonna per fare richiesta a Yad Vashem,²⁵ per inserire il nome della nonna nell'elenco dei *Giusti*.

Daria sapeva che tanti anni prima la nonna aveva soccorso, nascosto e salvato da morte sicura un ragazzo ebreo, ma di questo ragazzo non aveva mai saputo il nome e di lui non aveva avuto più notizie.

²⁵ Yad Vashem, l'Ente nazionale per la Memoria della Shoah, è stato istituito nel 1953 con un atto del Parlamento israeliano. Ha il compito di documentare e tramandare la storia del popolo ebraico durante la Shoah, preservando la memoria di ognuna delle sei milioni di vittime per mezzo dei suoi archivi, della biblioteca, della Scuola e dei musei. Ha inoltre il compito di ricordare i Giusti fra le Nazioni, che rischiarono le loro vite per aiutare gli ebrei durante la Shoah (<https://www.yadvashem.org/>)

Nel documento presente nella mail, la quale mi è stata inoltrata da William, Daria racconta che era ancora piccola quando la nonna arrivò nella loro casa, ma che era grande abbastanza da capire la storia di cui parlava. Purtroppo, la nonna, racconta Daria, ha avuto un ictus e a causa di questo la sua memoria è stata danneggiata in modo significativo. In ogni caso, continua Daria, i genitori conoscevano molto bene la storia del ragazzo salvato da Stefania e gliel'hanno raccontata una volta cresciuta.

Daria racconta che al tempo sapeva che la nonna aveva nascosto un ragazzo ebreo di cui non conosceva il nome nel proprio giardino durante la guerra; le guardie del ghetto o le SS erano andate a cercare il ragazzo nella proprietà della nonna, ma non lo avevano trovato. Più tardi Daria ha scoperto che il ragazzo aveva scritto una lettera alla nonna nella quale la informava di essere sopravvissuto e la ringraziava per avergli salvato la vita e le forniva il proprio indirizzo per contatti futuri: la lettera, però, non era arrivata a Stefania poiché il suo indirizzo a Przemyśl non era più lo stesso che aveva avuto durante la guerra. La lettera era arrivata nelle mani di un suo lontano parente alla lontana, il dr. Jan Holowka.

Riporto le parole esatte di Daria presenti nel documento allegato alla mail:

I was still an infant when she arrived. By the time I was old enough to understand history, my grand-mother had suffered a stroke and her memory and speech were significantly affected. However, my parents knew some of the story from my grand-mother (neither of them was in Przemyśl at the time) which I learned as I grew up. Basically, I knew that my grandmother had:

- a) hidden a Jewish boy (no name) in her garden at some stage in the war
- b) ghetto guards or SS had come to search the property but had not found him
- c) he had written a letter to her after the war which informed her of his survival; thanked her for her rescue and provided his address for further contact
- d) the letter had not reached her as she was no longer at the Przemyśl address she had during the war. The letter ended up with a Dr Jan Holowka – her distant relative.

Appena ne ha avute le possibilità, Daria ha iniziato la sua ricerca per trovare notizie del ragazzo ebreo che la nonna aveva salvato, ma fino al 1989 non è riuscita a trovare alcuna informazione.

Lo dichiara ella stessa, sempre nel documento allegato alla mail indirizzata a William Wolfowicz:

once I began to understand the Holocaust on a personal level (after reading the 'Diary of Anne Frank' when I was 13 years old) I was determined to try to find the Jewish boy who was hidden with my grand-parents. But there were no sources of information available. Until the fall of Communism in Eastern Europe in 1989.

Un giorno, verso la fine degli anni Novanta, Daria ricevette una lettera: gliela mandava un parente alla lontana, figlio del dr. Jan Holowka. In questa lettera la informava che tra le carte del padre, deceduto, aveva trovato un estratto di una lettera che un certo Idek Wolfowicz aveva scritto a Stefania nel 1947, da Roma, per ringraziarla di avergli salvato la vita e informarla che era riuscito a sopravvivere ai lager.

Allegato alla lettera indirizzata a Daria c'era un articolo uscito sul *Pogranicze*, un giornale locale di Przemyśl: quell'articolo parlava delle persone di Przemyśl che durante la guerra avevano aiutato gli ebrei e riportava inoltre un estratto, l'unico ritrovato, della lettera originale.



Figura 2: articolo pubblicato sul *Pogranicze*

Quello è stato il primo momento in cui Daria è venuta a conoscenza del nome del ragazzo salvato dalla nonna: Idek Wolfowicz.

Quella lettera non arrivò mai a Stefania, non solo poiché aveva cambiato il proprio domicilio di Przemyśl rispetto a quello che aveva durante la guerra, ma anche perché nel frattempo si era trasferita dal figlio Roman, in Australia. Ma il caso aveva voluto che quella lettera di Idek non fosse andata perduta e arrivasse nelle mani del dr. Jan Holowka.

Nel 2011 Daria si rivolse agli archivi Arolsen cercando ulteriori informazioni su Idek ma, sfortunatamente, questi non ne avevano alcuna.

Riporto le parole esatte di Daria:

Once I knew his name and that he was born in Tarnow, I enquired about Idek at the Arolsen Archives in 2011. They informed me that they had not had any success in locating him after passing the case on to the Italian Red Cross; and officially ‘closed the case’ in 2014. Unfortunately, and strangely, I have just now found out, from the Wolfowicz family, that the Italian Red Cross had the address in which Idek was living at the time (Rome, 2014) but couldn’t make contact and did not persevere. That would have been the last opportunity for me to meet Idek in person.

Nonostante le ricerche non avessero dato i risultati sperati, Daria non si diede per vinta: nell’agosto 2022, infatti, decise di pubblicare un post su un gruppo Facebook di ebrei di Tarnòw dove esplicitava la propria volontà di trovare il ragazzo il ragazzo salvato dalla nonna il cui nome, ora lo sapeva, era Idek Wolfowicz: ottenne una risposta, il link di un funerale tenutosi a Roma nel 2017.

Le parole di Daria:

I turned to Facebook. To the group: Jews from Tarnow. Within minutes of posting a search for Idek Wolfowicz, I received a suggestion: a link to a funeral in Rome 2017. Was this the Idek I had been thinking about and looking for, for over 50 years? I felt mixed emotions. Joyful excitement that I was getting closer . . . and sadness and disappointment that I would never see Idek alive, if this was indeed his funeral notice.

Nell'avviso venivano però citati figli e nipoti: era chiaro ormai che vivevano tutti a Roma, la stessa città da cui Idek, nel 1947, aveva inviato la lettera a Stefania. Quella sera stessa Daria contattò i familiari di Idek.

Combinarono un incontro il giorno prima del compleanno di Stefania.

Come racconta Daria:

That same evening, I searched out all the 'Wolfowicz' surnames in Facebook whose FB page was in Italian or showed Italian residence. I found 3 and messaged them all. The first to respond was Idek's grandson, Simone. I couldn't believe my good fortune. Simone informed me that his father, William, was still alive (and, as it turned out, the 3rd person I had messaged was, Miriam, Idek's daughter). I was overjoyed that they were happy to hear from me and that Idek had never forgotten my grandmother (she didn't forget him either, but only my father got her story before the debilitating stroke silenced her. And we all found each other on the day before her birth date. My husband called it: 'cosmic poetry'.

William, il figlio di Idek, racconta che il padre non ha mai smesso di parlare di Stefania, la donna a cui doveva la vita.

Stefania Fedewicz e Idek Wolfowicz non riusciti a re-incontrarsi: per una serie di circostanze imprevedibili e imponderabili.

Desiderio condiviso della famiglia dell'una e dell'altro è quello che il nome di Stefania sia incluso tra i nomi dei *Giusti fra le nazioni* come "salvatrice ucraina", come avrebbe voluto Stefania: ne hanno fatto richiesta a Yad Washem, ma ancora non hanno ricevuto risposta.

Capitolo III

La codifica della testimonianza

3.1 L'importanza della codifica

Il processo di codifica di una fonte scritta è molto importante in quanto la quasi totalità dei testi presenti nella nostra cultura si trova in forma scritta, su supporti fisici che possono essere soggetti a danni o deterioramenti e quindi andare perduti. Tra i supporti fisici più comuni troviamo le iscrizioni su pietra, le pergamene e i libri a stampa. In quest'ottica diventa fondamentale il processo di codifica digitale di questi testi per evitare che vengano perduti e per rendere disponibile questo patrimonio in nuove forme tecnologicamente avanzate. La codifica di un testo permette anche di svolgere indagini linguistiche e computazioni sulle risorse acquisite.

Per la gestione delle informazioni digitali e computazionali è necessario effettuare un processo di transcodifica.²⁶ Il file così ottenuto si trova in un formato adatto alla conservazione elettronica e all'elaborazione automatica, detto *machine readable*.

L'attività di codifica è un metodo di analisi e di comprensione della conoscenza nell'ambito delle discipline che si occupano del testo, e può essere considerato come un vero e proprio linguaggio teorico in cui è possibile esplicitare e formalizzare le ipotesi interpretative di un certo oggetto o fenomeno di studio.

Ogni pezzo di informazione aggiunta ad un testo grezzo avviene mediante l'inserimento di dati meta-testuali²⁷ e costituisce il risultato di un'analisi e di un'interpretazione che è stata condotta (da un umano o da una macchina) al fine di esplicitarne e rappresentarne le informazioni attraverso il formato digitale prescelto nel modo più accurato e completo possibile.

Ad oggi, la codifica viene effettuata utilizzando dei *markup language* basati su XML, che è considerato lo *standard de facto* per la codifica dei testi.

²⁶ Procedimento di conversione dei dati codificati secondo un sistema iniziale verso un sistema diverso: dal loro supporto originario verso un supporto elettronico.

²⁷ Markup, annotazione, codifica.

XML necessita di uno schema di codifica per l'interscambio e il riuso. Quello messo a punto dalla *Text Encoding Initiative* (TEI)²⁸ è lo schema di codifica più adottato nell'ambito della rappresentazione digitale di testi d'interesse storico-letterario.

Si è scelto di utilizzare la tecnologia e il linguaggio XML perché offre la possibilità di creare dei linguaggi di marcatura personalizzati e specifici per ogni esigenza e dominio.

XML permette inoltre di:

- Strutturare i dati;
- Gestire in modo nativo strutture gerarchiche;
- Elaborare e presentare i dati con strumenti XML nativi;
- Validare i tipi di strutture e i tipi di dati;
- Gestire riferimenti incrociati tramite opportuni meccanismi di de-referenziazione;
- Aggiungere e gestire annotazioni a vari livelli di granularità

Una volta realizzato il file XML, si procede con la validazione della struttura e del modello del documento XML mediante un apposito modello di testo descritto da un ulteriore documento formalmente definito in formato ODD.²⁹

Per l'elaborazione dei documenti XML, al fine di garantirne una migliore visualizzazione, sono stati adottati appositi fogli di stile in linguaggio XSL.³⁰ La tecnologia XSL, mantenuta come le specifiche XML dal consorzio W3C, segue un

²⁸ TEI è un autorevole progetto internazionale a cui afferiscono varie organizzazioni e università, il cui scopo è quello di fornire agli studiosi di Informatica Umanistica uno strumento più espressivo e flessibile possibile per rappresentare qualsiasi aspetto di interesse relativo alla risorsa testuale da rappresentare digitalmente.

²⁹ "ODD" (One Document Does it All) è un formato TEI XML utilizzato per esprimere frammenti di schema, documentazione in prosa e documentazione di riferimento per qualsiasi schema di markup XML in un unico documento. È il linguaggio in cui la TEI è definita in sé e che dovrebbe essere utilizzato anche per esprimere una personalizzazione dello schema TEI. Una specifica ODD è un documento XML TEI normale che include componenti dal modulo tagdocs. Questo modulo è completamente descritto nel capitolo Elementi di documentazione delle Linee guida TEI. Le Linee guida descrivono anche in dettaglio come dovrebbe essere elaborato un file ODD. Il modo consigliato per personalizzare la TEI è creare una specifica formale che esprima le personalizzazioni come documento XML utilizzando il markup TEI ODD; questo può quindi essere compilato in un adeguato DTD, schema RELAX NG o schema W3C (insieme alla documentazione di riferimento appropriata) utilizzando il programma Roma. associare l'ODD a grammatiche e schemi di codifica è fondamentale per garantire la coerenza, la validità e l'interoperabilità dei documenti TEI e per facilitare il processo di creazione, gestione e condivisione di documenti codificati in TEI.

³⁰ La tecnologia XSL segue un approccio dichiarativo, ovvero si dichiarano varie istruzioni per specificare il risultato che si desidera ottenere in output, è basato sul modello pattern matching e permette di ottenere file di output in diversi formati (XML, HTML o TEXT) a partire dallo stesso documento XML di input.

approccio dichiarativo.³¹ è basata sul modello *pattern-matching* e permette di ottenere file di output in diversi formati³² a partire dallo stesso documento XML di input. XSL racchiude in sé tre differenti linguaggi:

1. XSL-T: permette la trasformazione di un documento XML in un altro formato;
2. XSL-FO. si occupa dell'applicazione degli stili e della resa grafica del documento XML;
3. XPath: seleziona nodi, elementi e attributi di un documento XML e richiama funzioni con cui elaborare in modo estremamente efficace i dati selezionati.

Grazie a XSLT e XPath è possibile operare in modo flessibile sul documento. Nella scrittura di un foglio di stile XSL devono essere utilizzati gli appositi *namespace*, i quali consentono di distinguere le istruzioni proprie³³, i nodi TEI³⁴ e l'output.³⁵

Questo progetto di tesi prende origine dalla codifica di una testimonianza orale inedita di un superstite alla Shoa, *Idek Wolfowicz*. La testimonianza fa parte del progetto "*Voci dall'Inferno*". La testimonianza è una fonte orale e non scritta, per cui ci troviamo davanti a problematiche di interpretazione più complesse rispetto a quelle di una testimonianza scritta: essa è stata registrata con un dispositivo non professionale e, per questo motivo la comprensione del parlato spesso risulta un'attività complessa e soggetta ad errori. Il supporto analogico che conserva la fonte sono due microcassette TDK MC-90. Alcuni punti della testimonianza risultano interrotti per poi riprendere, tagliando così parti della conversazione. La comunicazione è costituita anche da espressioni in linguaggio non verbale che spesso può essere poco percepibile o può essere soggetto a mal interpretazione. Ci sono anche incidenti e rumori di sottofondo (una porta che si apre, un colpo di tosse, il rumore di una macchina, etc.) che possono alterare la comprensione del messaggio comunicato.

Ci sono anche degli aspetti positivi nell'utilizzo di una fonte orale piuttosto di una fonte scritta: per esempio si può sentire la variazione del tono della voce del parlante, l'incertezza del parlante a seconda della domanda posta, la presenza di pause più o

³¹ si dichiarano varie istruzioni per specificare il risultato che si desidera ottenere in output.

³² XML, HTML, TEXT.

³³ preceduti da prefisso *xsl*:

³⁴ preceduti da prefisso *tei*:

³⁵ È stato inoltre utilizzato il processore XSLT Saxon, ovvero un processore XSLT che esegue le seguenti azioni: legge il documento XML e ne crea l'albero corrispondente; inizia a percorrere l'albero leggendo i singoli nodi; confronta i nodi con le regole indicate nel foglio di stile; produce un output secondo le istruzioni date; restituisce l'albero di output.

meno lunghe. Tutto ciò va a restituire una spontaneità che in un testo scritto viene persa.

3.2 Passaggi preliminari

La testimonianza di Idek Wolfowicz è stata di difficile consultazione, dato soprattutto il supporto su cui era conservata (due microcassette TDK MC-90): per poterlo fare era necessario avere a disposizione uno specifico dispositivo (nel mio caso un recorder dittafofo per microcassette Olympus S928 pearlcore). È stato quindi necessario, prima di tutto, convertire le due microcassette in un formato digitale (.mp3) per garantire una più semplice fruizione e condivisione e una migliore esperienza d'ascolto.

Per creare una versione digitale delle microcassette le ho inserite all'interno dell'apposito lettore che ho poi collegato a un computer utilizzando un cavo munito di un adattatore micro-jack (di misura 2.5 mm, inferiore alla misura standard di 3.5 mm) e, utilizzando il software *Audacity*,³⁶ ho realizzato una rudimentale conversione digitale delle registrazioni.

Grazie alla creazione del formato digitale delle microcassette l'ascolto e la condivisione sono notevolmente migliorati.

Una volta ascoltata e trascritta la testimonianza ho proceduto alla codifica mediante il linguaggio di markup XML nella versione dello schema TEI P5 con le modalità descritte dal consorzio TEI all'interno delle proprie linee guida.

Per la validazione della struttura e del modello dei documenti XML ho utilizzato un'apposita grammatica DTD generata da un documento ODD realizzato in un precedente progetto di tesi per il progetto di *Voci dall'Inferno*.³⁷ La validazione è stata poi eseguita attraverso il processore XML Xerces per verificare che il documento XML fosse aderente allo schema definito.

3.3 Linee guida per la rappresentazione dei fenomeni del parlato

³⁶ Software open source, free, per registrar ed editare audio.

³⁷ Cfr. Sofia Capone, Tesi di laurea: 'Archiviare' gli Inferni di ieri e di oggi, Pisa, Università di Pisa, 2021-2022.

All'interno del modulo otto delle linee guida TEI si trova descritto il modulo destinato alla rappresentazione dei fenomeni del parlato.³⁸

Gli elementi definiti per rappresentare i fenomeni del parlato che ho utilizzato sono:

- `<u>`: rappresenta le *utterances* e sono le porzioni in cui è suddiviso il testo. Ogn *utterance* corrisponde a un enunciato di varia lunghezza all'interno del quale possono essere comprese anche pause, suoni non lessicali e rumori di sottofondo. In genere una pausa più o meno lunga o il cambio di interlocutore possono essere considerati dei buoni punti di riferimento per capire quando inserire un nuovo elemento `<u>`. Ogni elemento `<u>` è provvisto di un attributo `@who` che permette di associare le singole frasi alla persona che le ha pronunciate, e degli attributi `@xml:id` e `@synch` per la propria identificazione e sincronizzazione;
- `<pause>`:³⁹ rappresenta le pause. È accompagnato dall'attributo `@type` per fornire un'indicazione approssimativa sulla lunghezza di ciascuna;⁴⁰
- `<vocal>`:⁴¹ rappresenta i fenomeni vocali non lessicali (sospiri, colpi di tosse, starnuti o interiezioni come “eh”, “ah”, “ehm”, “mh”, eccetera) piuttosto frequenti in qualsiasi discorso orale. È accompagnato dall'attributo `@who` e seguito dall'elemento `<desc>`⁴² attraverso cui è stato possibile dare una descrizione dettagliata del singolo fenomeno preso in esame;⁴³
- `<indicent>`:⁴⁴ rappresenta i rumori di sottofondo, più o meno forti. Anche questo *tag* è accompagnato dall'elemento `<desc>`;⁴⁵
- `<gap>`:⁴⁶ rappresenta i passi non udibili della testimonianza. È accompagnato dall'attributo `@reason`;

³⁸ <https://tei-c.org/release/doc/tei-p5-doc/it/html/TS.html>.

³⁹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-pause.html>.

⁴⁰ https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L892.

⁴¹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-vocal.html>.

⁴² <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-desc.html>.

⁴³ https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L818-L820.

⁴⁴ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-incident.html>.

⁴⁵ https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L1659-L1661.

⁴⁶ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-gap.html>.

- `<choice>`:⁴⁷ rappresenta il raggruppamento di più codifiche alternative per una stessa parola o espressione. È accompagnato dai tag `<orig>`⁴⁸ e `<reg>`⁴⁹: il primo per la marcatura di forme linguistiche particolari, il secondo per la correzione a cura dell'*encoder*;⁵⁰
- ``:⁵¹ rappresenta fenomeni di ripetizioni, troncamenti o ‘false partenze’. È accompagnato dall’attributo `@type` attraverso cui è stato possibile specificare di volta in volta il tipo di fenomeno esaminato;⁵²
- `<placeName>`:⁵³ rappresenta le entità nominate di luoghi;⁵⁴
- `<persName>`:⁵⁵ rappresenta le entità nominate di persone.⁵⁶

⁴⁷ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-choice.html>.

⁴⁸ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-orig.html>.

⁴⁹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-reg.html>.

⁵⁰ https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L1230-L1233.

⁵¹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-del.html>.

⁵² https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L1038.

⁵³ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-placeName.html>.

⁵⁴ https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L115.

⁵⁵ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-persName.html>.

⁵⁶ https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/Idek_Wolfowicz.xml#L2740-L2743.

Capitolo IV

L'applicazione

4.1 eXist-DB

eXist-DB è un software open source in Java. È disponibile gratuitamente sia come eseguibile che in codice sorgente. È stato concepito come un database nativo XML e definito come un database NoSQL basato su un modello a documento.

L'unità di lavoro fondamentale di eXist-DB è il documento: questo ha la funzione di gestione dei dati.⁵⁷

Al contrario della maggior parte dei database NoSQL, eXist-DB utilizza un linguaggio di query standardizzato da W3C chiamato XQuery⁵⁸ (del quale mi occuperò nel prossimo paragrafo). Grazie all'utilizzo di questo linguaggio di interrogazione è possibile realizzare qualsiasi tipo di analisi e poterla eseguire su qualsiasi processore o piattaforma che supporta XQuery.

eXist-DB offre svariati vantaggi, tra cui:

- Possibilità di sviluppare applicazioni: può essere utilizzato come piattaforma per lo sviluppo di applicazioni web (come nel mio caso). Ciò è possibile grazie alla capacità di eXist-DB di integrarsi con linguaggi di trasformazione e interrogazione (XQuery, XSLT). Permette quindi di creare applicazioni web basate su documenti XML permettendo di accedervi in modo semplice e intuitivo;
- Funzionalità di ricerca strutturata: eXist-DB possiede la facoltà di definire molteplici indici. Ciò permette di sviluppare una ricerca avanzata mantenendo comunque un'elevata flessibilità, efficacia ed efficienza;
- Mentre i database relazionali (SQL) necessitano di definire uno schema prima di poter procedere con la memorizzazione dei dati, eXist-DB non ha questa necessità.

⁵⁷ <https://www.oreilly.com/library/view/exist/9781449337094/>.

⁵⁸ <https://www.w3.org/TR/xquery-31/>.

- È orientato al documento: i classici database relazionali (RDBMS) come MySQL⁵⁹, SQL Server⁶⁰, PostgreSQL⁶¹ e Oracle⁶² sono invece orientati a schemi rigidi basati sul concetto di tabelle e relazioni.
- A differenza di altri database NoSQL che archiviano documenti in formato JSON, come Apache Cassandra,⁶³ MongoDB⁶⁴ e Neo4j,⁶⁵ eXist-DB utilizza il formato XML. Ciò permette la gestione di strutture complesse che combinano dati e testo in modo semi-strutturato e misto.

eXist-DB permette inoltre di realizzare query sofisticate e portabili. I RDBMS usano tipicamente un linguaggio di query standardizzato SQL, può essere però difficile eseguire le stesse query SQL in diversi prodotti dei database RDBMS. Allo stesso modo, la maggior parte dei sistemi NoSQL possiedono i propri linguaggi di query proprietari che sono interamente specifici per il prodotto. eXist-DB adotta un approccio diverso e fornisce linguaggi standard quali XQuery e XSLT, che sono linguaggi di query e trasformazione definiti e mantenuti dal consorzio W3C, il che significa che si possono eseguire query eXist-DB su ogni altro prodotto che fornisca un processore XQuery e/o XSLT.⁶⁶

4.2: Il linguaggio XQuery

XQuery è un linguaggio di programmazione standard sviluppato e mantenuto dal consorzio W3C (XML Query language) che permette di selezionare e interrogare collezioni di documenti XML (e non solo): permette anche di riorganizzarli, trasformarli e restituire l'output in diversi formati.⁶⁷

Alcune caratteristiche di XQuery sono:

- Espressività e flessibilità: questo linguaggio, infatti, può essere usato per interrogare molteplici strutture dati, tra le quali troviamo dati gerarchici o tabellari;⁶⁸

⁵⁹ Cfr. <https://www.mysql.com/it/>.

⁶⁰ Cfr. <https://www.microsoft.com/en-us/sql-server/sql-server-2022>.

⁶¹ Cfr. <https://www.postgresql.org/>.

⁶² Cfr. <https://www.oracle.com/it/>.

⁶³ Cfr. https://cassandra.apache.org/_/index.html.

⁶⁴ Cfr. <https://www.mongodb.com/it-it>.

⁶⁵ Cfr. <https://neo4j.com/>.

⁶⁶ Siegel, E. e Retter, A. (2015). eXist. United States of America, O'Reilly Media, pp. 3-4.

⁶⁷ Walmsley Priscilla, XQuery Search Across a Variety of XML Data, p. 2.

⁶⁸ Cfr. <https://en.wikibooks.org/wiki/XQuery/Benefits>.

- Brevità: le istruzioni XQuery sono più brevi rispetto a quelle definite dal linguaggio XSLT o SQL;⁶⁹
- Universalità: è un linguaggio riconosciuto dal W3C;
- Semplicità della sintassi;
- Possibilità di estrarre dati che provengono da documenti diversi e aggregarli;
- Possibilità di filtrare le informazioni a disposizione in base a criteri specifici definiti dal programmatore durante la costruzione della query;
- Capacità di creare applicazioni web;
- Capacità di creare direttamente pagine HTML da poter usare all'interno di siti web;
- Possibilità di manipolare stringhe o svolgere espressioni aritmetiche o calcoli su dati numerici.

La selezione di elementi o attributi da un documento XML di input rappresenta il tipo di interrogazione fondamentale realizzata tramite XQuery: questo linguaggio utilizza le espressioni *XPath*⁷⁰ per costruire le clausole di interrogazione.

XQuery utilizza anche il cosiddetto meccanismo di espressioni *FLWOR*: questo è il costruito base del linguaggio, supporta l'iterazione e permette di formulare selezioni molto complesse manipolando documenti XML.

FLWOR è l'acronimo di “*For, Let, Where, Order by, Return*” ed è composto quindi da cinque parti di cui l'unica obbligatoria è return:

- FOR: definisce l'iterazione che permette di valutare il resto del FLWOR tante volte quanti sono gli elementi presenti nella sequenza di selezione;
- LET: si utilizza per associare un valore a una variabile (è utile per evitare di ripetere la stessa espressione più volte);
- ORDER BY: ordina l'output in base a una o più condizioni di ordinamento (devono essere separate da una virgola);
- WHERE: filtra gli elementi in base a uno o più criteri. Può essere composta da più espressioni legate tra loro da *or* e *and*;
- RETURN: costruisce il risultato finale dell'espressione FLWOR.

In XQuery è anche possibile usare le espressioni condizionali: il costruito più utilizzato è *if-then-else*.

⁶⁹ Cfr. <https://en.wikibooks.org/wiki/XQuery/Benefits>.

⁷⁰ XPath è un linguaggio che permette di selezionare, tra gli altri dati, gli elementi e gli attributi da un documento XML sui quali verranno, poi, applicate le operazioni per la presentazione dei dati.

Un altro elemento importante che compone questo linguaggio è la possibilità di usare e di definire funzioni.

Le funzioni possono essere definite come blocchi di codice riutilizzabili che ricevono uno o più argomenti e restituiscono un risultato.

Esistono due tipi di funzioni: le funzioni integrate (predefinite), chiamate *built-in* e le funzioni definite dal programmatore.

Per definire una propria funzione si seguono le indicazioni:

1. La parola chiave è *declare*;
2. Si può aggiungere un'annotazione facoltativa;
3. La parola chiave *function*;
4. Il nome completo della funzione preceduto da un *namespace*, che viene definito come un prefisso associato a un URI e viene utilizzato per identificare in modo univoco le funzioni usate o gli elementi XML, anche se la funzione è creata dall'utente;
5. Un elenco di parametri racchiusi tra parentesi tonde e separati dalla virgola. È possibile inserire il tipo di dato del parametro.
Può essere specificato anche il tipo di dato che restituisce la funzione (facoltativo);
6. Il corpo della funzione racchiuso tra parentesi graffe e seguito da un punto e virgola.

Un esempio di funzione XQuery è il seguente:

```
declare function app:cercaSurname($cognome as xs:string?) as element()* {  
  for $doc in collection("/db/apps/proget/xml")//tei:listPerson/tei:person[tei:persName/tei:surname = $cognome]  
  (: Verifica se il tag surname è uguale a $cognome :)  
  return $doc  
  (: Restituisce i documenti che soddisfano la condizione :)  
};
```

Figura 3: funzione *cercaSurname*⁷¹

In cui ritroviamo tutte le indicazioni sopracitate.

4.3 Stadio iniziale

Per la realizzazione dell'applicazione web per il progetto di *Voci dall'Inferno* ho lavorato con altri colleghi del gruppo di lavoro, in particolare con colleghi che si sono occupati principalmente di sviluppare un modulo per l'interrogazione avanzata.

⁷¹ <https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/app.xqm#L303-L307>.

L'obiettivo è quello di creare un'applicazione che armonizzasse le sperimentazioni e i prototipi esistenti creati in precedenti lavori di tesi⁷² e che portasse avanti le indagini testuali sui file XML presenti nel catalogo di *Voci dall'Inferno*. Di vitale importanza è anche il lavoro di “generalizzazione” (iniziato già precedentemente da un collega):⁷³ tutte le funzioni create sono infatti valide per tutti i documenti XML presenti nell'archivio. In quest'ottica i lavori dei prossimi colleghi seguiranno lo schema ODD creato e modificato da precedenti lavori di tesi nel contesto del progetto di ricerca⁷⁴ e vedranno comparire le proprie testimonianze in modo dinamico all'interno del catalogo. I dati restituiti dalle funzioni implementate non sono inseriti staticamente, ma le funzioni recuperano autonomamente, automaticamente e dinamicamente i dati dai valori inseriti dal gruppo di lavoro durante la codifica XML delle varie testimonianze.

Per fare un passo in più rispetto ai lavori dei miei colleghi ho posto l'accento principalmente su:

- La creazione di funzioni che permettano la ricerca nel catalogo per mezzo di dati descrittivi presenti nel *teiHeader* dei file XML;
- Migliorare la visualizzazione della trascrizione del linguaggio parlato permettendo di scegliere tra due convenzioni di trascrizione differenti: lo standard *CHAT*⁷⁵ e lo standard *DTI*.⁷⁶

⁷² Cfr. Andrei Mihaila, Tesi di laurea: Inferno di spine: Nedo Fiano e la sua testimonianza, Pisa, Università di Pisa, 2022-2023.

Cfr. Erika Deboni, Tesi di laurea: Una voce dall'inferno: la testimonianza di Ida Marcheria, Pisa, Università di Pisa, 2020-2021.

⁷³ Cfr. Andrei Mihaila, Tesi di laurea: Inferno di spine: Nedo Fiano e la sua testimonianza, Pisa, Università di Pisa, 2022-2023.

Cfr. Capone Sofia, Tesi di laurea: 'Archiviare' gli Inferni di ieri e di oggi, Pisa, Università di Pisa, 2021-2022.

⁷⁴ Cfr. Andrei Mihaila, Tesi di laurea: Inferno di spine: Nedo Fiano e la sua testimonianza, Pisa, Università di Pisa, 2022-2023.

⁷⁵ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, Journal of the Text Encoding Initiative [Online].

⁷⁶ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, Journal of the Text Encoding Initiative [Online].

Capitolo V

Standardizzazione della trascrizione del linguaggio parlato

5.1 Cenni preliminari

Ho lavorato su un metodo di visualizzazione/presentazione dei dati trascritti che seguisse uno standard ben preciso, in modo che i fenomeni codificati venissero rappresentati secondo criteri rigorosi e convenzioni standard riconosciute dalla comunità degli studiosi di trascrizioni da fonti orali.

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

Sono state sviluppate, negli ultimi anni, molteplici convenzioni di trascrizione inerenti a lingue, interessi e tradizioni diverse. Le convenzioni vengono implementate per mezzo dell'utilizzo di una varietà di strumenti informatici, ognuno con il proprio modello di dati e formati. Non esiste quindi un metodo dominante o standardizzato per la trascrizione del linguaggio parlato, è però necessario trovarne uno.⁷⁷

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

- I trascrittori devono lavorare con strumenti informatici adeguati;
- La standardizzazione dovrebbe distinguere attentamente tra la variabilità presente nelle convenzioni di trascrizione e nei formati degli strumenti che può derivare da differenze di interesse di ricerca e approcci teorici o da idiosincrasie individuali e proporre unificazioni solo per il primo caso;
- Ogni sforzo di standardizzazione deve lasciare spazio a negoziazioni tra le parti coinvolte.⁷⁸

5.2 Gli strumenti di trascrizione

⁷⁷ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, Journal of the Text Encoding Initiative [Online].

⁷⁸ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, Journal of the Text Encoding Initiative [Online].

Gli strumenti di trascrizione devono essere il più precisi possibile e supportare l'utente nel collegare le descrizioni testuali alle parti selezionate di una traccia audio. Tra gli strumenti a disposizione troviamo, per esempio:

- EXMERaLDA: strumento specializzato nella dialettologia, nella ricerca sul multilinguismo e nell'analisi pragmatica del discorso;
- CLAN/CHAT: originariamente sviluppato per estrapolare dati per la trascrizione e la codifica dei dati sulla lingua dei bambini;
- ELAN: utilizzato per la trascrizione della lingua dei segni.

Nonostante le differenze in termini di design e nell'implementazione, questi strumenti possono essere visti come varianti dello stesso modello di base: questo rappresenta le informazioni attraverso un'annotazione allineata temporalmente che consiste in un punto di fine e una parte contenente la trascrizione o l'annotazione.⁷⁹

Gli strumenti, però, nonostante condividano uno stesso modello di base, presentano numerose differenze e una completa assimilazione dei diversi modelli di dati non è possibile.

5.3 Convenzioni di trascrizione

Esistono diverse convenzioni per rappresentare la trascrizione di una risorsa orale: purtroppo molte di esse sono state sviluppate per specifici corpora o progetti e non sono adatte ad essere adottate da un pubblico più ampio.

Tuttavia, alcune convenzioni sono state pubblicate e possono rappresentare un possibile riferimento:

- HIAT⁸⁰ (Halbinterpretative Arbeitstranskriptionen): sistema utilizzato nella comunità di ricerca sulla pragmatica funzionale;
- GAT⁸¹ (Gesprächsanalytisches Transkriptionssystem): utilizzato in Germania per la trascrizione della lingua tedesca;

⁷⁹ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, *Journal of the Text Encoding Initiative* [Online].

⁸⁰ Ehlich, K. 2003. "HIAT: A Transcription System for Discourse Data." In *Talking Data: Transcription and Coding in Discourse Research*, ed. J. Edwards and M. Lampert, 123–148. Hillsdale, NJ: Erlbaum.

⁸¹ Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluff, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock, and S. Uhmann. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)". *Gesprächsforschung* 10: 353–402.

- CHAT⁸² (Codes for Human Analysis of Transcripts): utilizzato nei campi di ricerca per l'analisi del linguaggio dei bambini e CA-CHAT, il cui fine è l'analisi delle conversazioni;
- CGAT: utilizzato per l'analisi genomica comparativa;
- DT1⁸³ (Discourse Transcription): usato per la trascrizione del Santa Barbara Corpus of Spoken American English;
- CTS⁸⁴ (Göteborg Transcription Standard): utilizzato in Svezia presso l'università di Göteborg e fondamentale per il Corpus Parlatto Svedese.

Le convenzioni sopra elencate difficilmente si integrano/allineano l'una con l'altra. Inoltre, non sono stati compiuti sforzi significativi per stabilire uno standard comune a tutti gli ambiti e alle necessità.

Tutte le convenzioni condividono due entità di base: le pause non riempite e gli eventi ubili non linguistici (per es. tosse, risata, respiro). Inoltre, tutte specificano alcuni metodi per la rappresentazione dell'incertezza nella trascrizione e per la rappresentazione di passaggi non comprensibili. Un'altra classe che di entità che si trova nella maggior parte dei sistemi di trascrizione riguarda le caratterizzazioni prosodiche di parole o porzioni di parole. Questa classe può includere fenomeni come l'allungamento delle sillabe o lo stress enfatico.

Nella seguente tabella relativa a un segmento di discorso trascritto secondo cinque sistemi diversi di trascrizione, presa dall'articolo scritto da *Thomas Schmidt* per *Journals of the Text Encoding Initiative*, si possono vedere alcuni importanti differenze tra le convenzioni di trascrizione.

HIAT	((coughs)) You must/ you (should) let • it be. ((laughs)) Please!
GAT	((coughs)) you must- you (should/could) let (-) it be; ((laughs)) plea:se-
CHAT	&=coughs you must... you should let # it be. &=laughs please!
DT1	(COUGH) you must-- you <X should X> let .. it be. @@ please?
cGAT	((coughs)) you must you (should/could) let (-) it be ((laughs)) please

Figura 4: differenze tra alcune convenzioni di trascrizione

⁸² MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. 3rd edition Mahwah, NJ: Lawrence Erlbaum Associates.

⁸³ Discourse Transcription: DuBois et al. 1993.

⁸⁴ Göteborg Transcription Standard: Nivre et al. 1999.

Possiamo notare alcune differenze simboliche tra le convenzioni utilizzate nei diversi sistemi di trascrizione, ad esempio HIAT, cGAT e GAT utilizzano le doppie parentesi per descrivere gli incidenti non verbali (in questo caso colpi di tosse o risate) mentre CHAT utilizza i simboli prefissati &= e DT1 utilizza lettere maiuscole tra parentesi singole. Inoltre, DT1 ha simboli speciali predefiniti per alcuni incidenti specifici, ad esempio i simboli @@ per rappresentare una risata. Allo stesso modo ogni sistema ha i propri simboli per rappresentare una breve pausa non misurata: in HIAT si rappresenta con “•”, in cGAT e in GAT si rappresenta con “(-)”, in DT1 si rappresenta con “:(periodi)” e in CHAT con “#”.

Le convenzioni variano anche per quanto riguarda i fenomeni che vengono rappresentati nella trascrizione: per esempio l’allungamento della vocale nella parola *please* viene indicato in HIAT duplicando il simbolo della vocale e in GAT viene inserito un doppio punto, negli altri tre sistemi questo fenomeno non viene rappresentato.

Anche l’incertezza del trascrittore rispetto a una specifica parola può essere segnalata in HIAT, GAT, cGAT e DT1 utilizzando le parentesi singole o, nel caso di DT1, con “<X” e “X>”, solo GAT e cGAT però permettono di specificare alternative di trascrizione aggiuntive per una parola incerta, aggiungendole dopo una barra e all’interno delle parentesi.⁸⁵

Un’altra forma di variazione riguarda l’organizzazione delle unità di base della trascrizione in strutture più ampie.

Anche questa forma di variazione è evidente nei simboli di punteggiatura utilizzati nella figura 1:

- HIAT: il tratto di discorso viene diviso in due entità chiamate *enunciati*. Gli enunciati sono unità pragmatiche del discorso che sono identificate e classificate. Il primo enunciato termina con un punto, indicando che in modalità dichiarativa, mentre il secondo termina con un punto esclamativo, indicando un umore esclamativo. Un altro simbolo di punteggiatura è una barra obliqua dietro la parola *must*: indica una correzione autonoma ma non ha funzione di segno di fine enunciato. È importante inoltre notare che HIAT utilizza la lettera maiuscola all’inizio di ogni enunciato;

⁸⁵ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, Journal of the Text Encoding Initiative [Online].

- GAT: lo stesso tratto di discorso viene suddiviso in tre entità chiamate *frasi di intonazione*. Le frasi di intonazione sono unità prosodiche del discorso che sono identificate e classificate in base alla loro forma (in particolare al contorno intonativo). La prima e la terza frase di intonazione terminano con un trattino, indicando una modulazione finale del tono a livello costante. La seconda frase, invece, termina con un punto e virgola: ciò indica un movimento finale del tono in caduta;
- CHAT: la suddivisione è simile a quella che troviamo nel sistema HIAT ma è formata da tre enunciati. Il primo enunciato termina con un simbolo di ellissi (tre puntini): ciò indica un enunciato interrotto. Gli altri due enunciati sono contrassegnati rispettivamente da un punto (periodo) e un punto esclamativo;
- DT1: le entità corrispondenti sono chiamate *unità di intonazione*. La prima termina con due trattini, la seconda con un punto e la terza è contrassegnata da un punto interrogativo;
- cGAT: non presenta una suddivisione dei blocchi di base della trascrizione in entità più ampie.

Se volessimo standardizzare le informazioni codificate nelle convenzioni di trascrizione sarebbe necessario prendere in considerazione questi diversi tipi di variazione tra i sistemi.⁸⁶ L'idea di una standardizzazione formale e condivisa sembra ancora lontana.

5.4: Convenzioni utilizzate all'interno del progetto

All'interno del progetto, per la trascrizione della testimonianza, ho deciso di utilizzare le convenzioni offerte da CHAT⁸⁷ e quelle offerte da DT1⁸⁸ (DT1 è stato superato dalla convenzione DFN).

Ho inserito alcuni fenomeni, i più ricorrenti all'interno delle testimonianze, ovvero quelli relativi alle ripetizioni effettuate dai parlanti, alle parole o porzioni non chiare, all'enfasi sul tono della voce (in crescendo) e alle pause.

In CHAT le convenzioni utilizzate sono di seguito elencate:

⁸⁶ Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, *Journal of the Text Encoding Initiative* [Online].

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

⁸⁸ <https://github.com/digitallinguistics/DFT/blob/master/Version%20Mapping.csv>.

- Registrazione non chiara: [?];
- Parte mancante: xxx;
- tono in crescendo: ↑;
- pausa: (..)

In DT1, invece, gli stessi fenomeni si indicano nel modo seguente:

- Registrazione non chiara o parte mancante: xxx;
- tono in crescendo: /;
- pausa: ...

Nel progetto che ho realizzato, nella pagina dedicata al singolo testimone scelto, si può scegliere di visualizzare:

- la trascrizione dell'intervista con convenzione CHAT;
- la trascrizione dell'intervista con convenzione DT1.

Per rendere possibile questa scelta ho realizzato innanzitutto un menu di scelta a *discesa*: si può scegliere così quale delle tre trascrizioni visualizzare. Il menu è accompagnato da un *button* che permette di nascondere la trascrizione e la legenda che stiamo visualizzando.

Una volta effettuata la scelta, oltre alla trascrizione, comparirà anche una breve legenda che indica i fenomeni analizzati e il significato delle convenzioni di trascrizione.

Per inserire le convenzioni ho creato delle funzioni XQuery che, a seconda della scelta fatta dall'utente, generano in output un frammento *HTML* utilizzando il modulo di trasformazione offerto da eXist-DB:

```
let $xslt := doc("/db/apps/proget/xslt/xslt_chat.xsl")
let $newHTML := transform:transform($fileXML, $xslt, ())
```

Figura 5: esempio modulo di trasformazione eXist-DB

Il modulo di trasformazione eXistDB permette di eseguire trasformazioni XSLT dal codice XQuery utilizzando il processore Saxon.

La funzione *transform:transform* esegue una trasformazione XSLT e restituisce il risultato. In questo caso il risultato viene associato alla variabile *\$newHTML*.

Nel caso riportato nella figura si è creato un output *HTML* utilizzando un foglio di stile *XSLT* contenente i fenomeni analizzati secondo le regole della convenzione CHAT: si viene così a creare una mappatura dei fenomeni dall'originale codifica XML-TEI alla convenzione CHAT prescelta.

```

<xsl:template match="//tei:u/tei:unclear">
  <span class="unclear1">
    [?]
  </span>
</xsl:template>

<xsl:template match="//tei:u/tei:gap">
  <span class="gap1">
    xxx
  </span>
</xsl:template>

<xsl:template match="//tei:u/tei:emph">
  <span class="emph1">
    ↑
  </span>
</xsl:template>

<xsl:template match="//tei:u/tei:pause">
  <span class="pause1">
    (...)
  </span>
</xsl:template>

```

Figura 6: esempio di match XSL con convenzione CHAT

Nel foglio di stile XSLT ho creato dei *match* per ogni elemento mappato con convenzione CHAT.

La gestione della scelta fatta nel menu a discesa è stata effettuata con *Javascript*.⁸⁹

```

function selezionaLegenda() {
  var select = document.getElementById("legendaSelect");
  var scelta = select.value;
  var contenutoLegenda = document.getElementById("contenutoLegenda");
  var divchat = document.getElementsByClassName("intervistachatchat");
  var div2 = document.getElementsByClassName("intervista2");

  // Nascondi tutti i div chat e div2
  for (var i = 0; i < divchat.length; i++) {
    divchat[i].style.display = "none";
  }
  for (var i = 0; i < div2.length; i++) {
    div2[i].style.display = "none";
  }

  // Nascondi il contenuto precedente della legenda
  contenutoLegenda.style.display = "none";

  if (scelta == "legenda2") {
    // Inserisci il contenuto per Legenda 1
    contenutoLegenda.innerHTML = "<h2>Legenda CHAT</h2><ul><li>Registrazione non chiara: [?]</li><li>Parte mancante: xxx</li><li>tono in crescendo: +</li><li>pausa: (...)</li></ul>";
    contenutoLegenda.style.display = "block"; // Mostra il contenuto della legenda
    for (var i = 0; i < divchat.length; i++) {
      divchat[i].style.display = "block"; // Mostra i div chat
    }
  } else if (scelta == "legenda2") {
    // Inserisci il contenuto per Legenda 2
    contenutoLegenda.innerHTML = "<h2>Legenda DT1</h2><ul><li>Registrazione non chiara: xxx</li><li>tono in crescendo: /</li><li>pausa: ...</li></ul>";
    contenutoLegenda.style.display = "block"; // Mostra il contenuto della legenda
    for (var i = 0; i < div2.length; i++) {
      div2[i].style.display = "block"; // Mostra i div2
    }
  }
}

```

Figura 7: funzione Javascript che gestisce la scelta della convenzione⁹⁰

La funzione nella figura permette di selezionare un'opzione (da un menu a discesa) e mostrare il contenuto relativo a quella scelta, nascondendo contemporaneamente il contenuto precedente e gli elementi non pertinenti. La funzione è progettata per gestire un menu a discesa (select) nella pagina web e modificare dinamicamente il contenuto della pagina in base alla scelta fatta dall'utente nel menu a discesa.

⁸⁹ linguaggio di programmazione per lo sviluppo di web applications.

⁹⁰ <https://github.com/DilettaLelli99/Vocidall-Inferno/blob/687c334806a6761bd089bc161ea02b3335f2874e/legenda.js#L1-L36>.

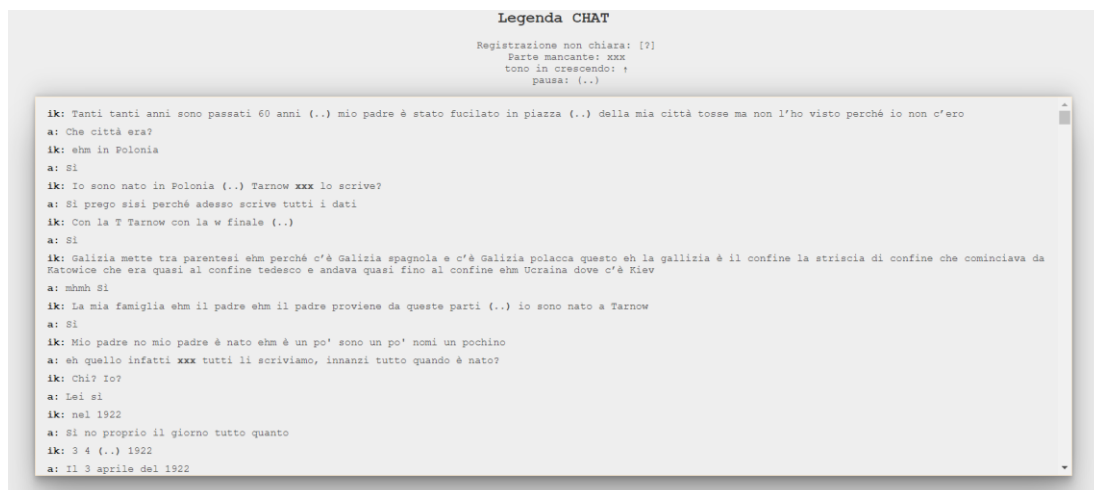


Figura 8: *legenda CHAT*

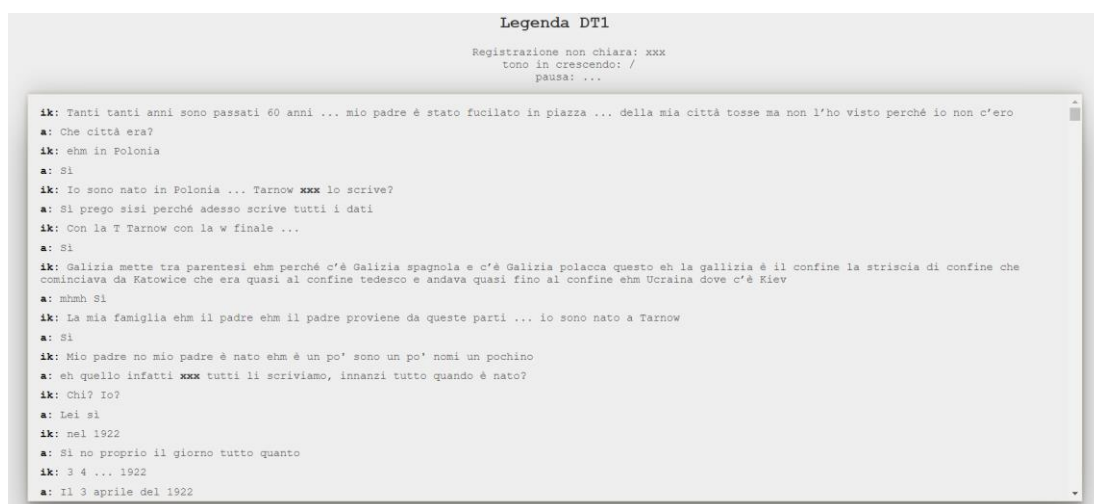


Figura 9: *Legenda DT1*

Capitolo VI

La ricerca nel catalogo

6.1 La ricerca

Il catalogo delle testimonianze ha necessità di essere filtrato secondo chiavi di ricerca specifiche. Ho implementato una prima versione di questa funzionalità che permette di svolgere ricerche su parametri presenti nel *teiHeader* dei documenti XML. È importante specificare che si tratta di una prima versione di ricerca nel catalogo che sfrutta alcuni metadati presenti nel *teiHeader* che potrà poi essere implementata in una ricerca filtrata ispirata a cataloghi noti come, per esempio, il catalogo ICCU SBN.⁹¹ Per farlo ho usato un pacchetto predefinito in eXist-DB che usa il noto tool di indicizzazione Apache Lucene.

Attraverso indici *Lucene* ho creato una sezione di search avanzato in cui la ricerca può essere fatta attraverso alcuni parametri contenuti all'interno del *teiHeader* dei file XML.

I parametri che sono stati scelti per la prima versione della ricerca in catalogo sono:

- nome;
- cognome;
- nazionalità;
- età;
- anno di nascita ed eventuale anno di morte;
- sesso.

Ho scelto questi parametri piuttosto che altri perché permettono di dar vita a ricerche sulle caratteristiche dei vari testimoni. Sarebbe interessante, in un'ottica futura, dar vita anche a ricerche basate su parametri inerenti alla prigionia dei testimoni o sui presenti durante le interviste, in modo da trovare nuovi filoni d'indagine basati su caratteristiche comuni ai testimoni.

6.2 Apache Lucene

⁹¹ <https://www.iccu.sbn.it/it/>.

La piattaforma eXist-DB offre, integrate al suo interno, delle funzionalità di indicizzazione del testo che consentono di cercare parole e frasi all'interno dei documenti, usando un linguaggio di interrogazione la cui espressività consente di costruire espressioni molto articolate. Queste funzionalità si realizzano tramite la libreria Apache Lucene⁹², ovvero un indicizzatore full-text scritto in Java che può essere usato all'interno di qualsiasi applicazione per implementare funzioni di ricerca avanzata su testo piano. L'ambiente eXist-DB, inoltre, aggiorna le modifiche in maniera automatica, quindi, non è necessaria la re-indicizzazione manuale.

La configurazione dell'indice avviene tramite la modifica del file di configurazione *collection.xconf* che viene generato automaticamente in fase di creazione dell'applet dal tool Yeoman.⁹³ Questo file è un documento XML standard: l'elemento radice del documento è *<collection>* e il figlio è l'elemento *<index>*. All'interno di questo elemento vengono definiti gli elementi figli che servono per specificare i vari tipi di indice che verranno generati e usati.

eXist-DB offre la possibilità di definire l'indice Lucene su un singolo elemento mediante l'utilizzo del tag *<qname>* oppure un percorso del nodo con carattere jolly usando l'elemento *<match>*.⁹⁴

eXist-DB dà inoltre all'utente la possibilità di determinare aspetti relativi all'analisi del testo prima della ricezione dei risultati. L'analizzatore di base è lo *StandardAnalyzer*, il quale tokenizza il testo, elimina i segni di punteggiatura e converte le lettere maiuscole in minuscole. L'utente può inoltre definire il proprio analizzatore seguendo le istruzioni presenti nella documentazione di eXist-DB.⁹⁵

La tokenizzazione⁹⁶ è il processo di suddivisione del testo di input in piccoli elementi di indicizzazione, chiamati token. Il modo in cui il testo di input viene suddiviso in token influisce notevolmente su come le persone saranno in grado di cercare quel testo. Ad esempio, l'identificazione dell'inizio e della fine delle frasi può essere utile per

⁹² <http://lucene.apache.org/>.

⁹³ Yeoman è un sistema node.js che consente la creazione della struttura di base di diverse applicazioni, usando un meccanismo generico che fa uso dei generatori. In questo modo, si facilita il lavoro degli sviluppatori, i quali possono concentrarsi sulla scrittura del codice di dominio, lasciando a Yeoman la parte generica e ripetitiva finalizzata alla generazione della struttura del filesystem e dei file di configurazione del progetto.

⁹⁴ Generalmente utilizzato nel caso in cui il programmatore abbia la necessità di definire un indice su tutti gli elementi figli di un determinato tag.

⁹⁵ <https://exist-db.org/exist/apps/doc/lucene>.

⁹⁶ https://lucene.apache.org/core/9_8_0/core/org/apache/lucene/analysis/package-summary.html#package.description.

ricerche di frasi e prossimità più accurate (anche se l'identificazione delle frasi non è fornita da Lucene).

In alcuni casi, semplicemente suddividere il testo di input in token non è sufficiente: potrebbe essere necessaria un'analisi più approfondita. Lucene include strutture di analisi sia pre-tokenizzazione che post-tokenizzazione.

L'analisi pre-tokenizzazione può includere (ma non è limitata a) la rimozione del markup HTML e la trasformazione o la rimozione del testo corrispondente a modelli arbitrari o insiemi di stringhe fisse.

Ci sono molte fasi di post-tokenizzazione che possono essere eseguite, tra cui (ma non limitate a):

- Stemming: Sostituzione delle parole con le loro radici. Ad esempio, con lo stemming in inglese, "bikes" viene sostituito con "bike"; ora la query "bike" può trovare documenti contenenti sia "bike" che quelli contenenti "bikes";
- Filtraggio degli stop words: Parole comuni come "il", "e" e "un" raramente aggiungono valore a una ricerca. Rimuovendo queste parole si riduce la dimensione dell'indice e si migliora le prestazioni. Può anche ridurre un po' di "rumore" e migliorare effettivamente la qualità della ricerca;
- Normalizzazione del testo: La rimozione degli accenti e di altre marcature caratteristiche può migliorare la ricerca;
- Espansione dei sinonimi: Aggiungere sinonimi nella stessa posizione del token corrente può migliorare la corrispondenza quando gli utenti cercano con parole nel set di sinonimi.

Nell'applicazione che ho realizzato l'indice è stato definito sui seguenti elementi:

- u;
- person;
- persName;
- forename;
- surname;
- birth/@when;
- death/@when;
- sex;
- nationality.

In questo modo il testo che viene passato a Lucene per essere indicizzato è il valore di tutti gli elementi sopracitati.

```
<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index xmlns:tei="http://www.tei-c.org/ns/1.0" xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <.lucene>
      <analyzer class="org.apache.lucene.analysis.standard.StandardAnalyzer"/>
      <text qname="tei:u"/>
      <text qname="tei:person"/>
      <text qname="tei:persName"/>
      <text qname="tei:forename"/>
      <text qname="tei:surname"/>
      <text qname="tei:birth/@when"/>
      <text qname="tei:death/@when"/>
      <text qname="tei:sex"/>
      <text qname="tei:nationality"/>
    </.lucene>
  </index>
</collection>
```

Figura 10: indice realizzato per l'applicazione web

6.3 teiHeader

Un documento conforme allo schema TEI contiene, in generale, un singolo elemento `teiHeader` (ci sono casi in cui, avendo un testo composito come un corpus o una raccolta, possiamo avere diversi header), seguito da un singolo elemento `text` o `facsimile` o entrambi.

L'elemento `teiHeader` (intestazione TEI) fornisce metadati descrittivi e dichiarativi associati a una risorsa digitale o a un insieme di risorse. È formato da cinque componenti principali:⁹⁷

- `fileDesc` (descrizione del file): contiene una descrizione bibliografica completa del documento elettronico;
- `encodingDesc` (descrizione della codifica): documenta la relazione tra un testo elettronico e la fonte (o le fonti) da cui è derivato;
- `profileDesc` (descrizione del profilo del testo): fornisce una descrizione dettagliata degli aspetti non bibliografici di un testo, in particolare le lingue e le sottolingue utilizzate, la situazione in cui è stato prodotto, i partecipanti e il contesto;
- `xenData` (metadati non-TEI): fornisce un elemento contenitore in cui è possibile inserire metadati in formati non-TEI;
- `revisionDesc` (descrizione delle revisioni): riassume la cronologia delle revisioni per un file.

⁹⁷ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD11>.

I dati su cui poter impostare la ricerca avanzata che ho realizzato si trovano all'interno del *profileDesc*.

```
<person xml:id="ik" role="testimone" n="3">
  <persName>
    <forename>Idek</forename>
    <surname>Wolfowicz</surname>
  </persName>

  <birth when="1922"/>
  <death when="2017"/>
  <sex>Maschio</sex>
  <nationality>Polacca</nationality>
</person>
```

Figura 11: esempio *teiHeader*

In questo caso ci troviamo all'interno del file XML relativo al testimone di cui mi sono occupata, Idek Wolfowicz. All'interno del tag `<person>`⁹⁸ (fornisce informazioni relative a un soggetto specifico, come i partecipanti a un'interazione verbale o le persone identificate da una fonte storica) troviamo tutti gli elementi utilizzati in un secondo momento per realizzare la ricerca:

- `<persName>`: contiene un nome proprio o un sintagma identificabile come nome proprio, che si riferisce a una persona e può includere qualsiasi o tutti i prenomi, cognomi, titoli onorifici, o nomi aggiunti della persona in questione;⁹⁹
- `<forename>`: contiene un nome di battesimo;¹⁰⁰
- `<surname>`: contiene un cognome;¹⁰¹
- `<birth>`: contiene informazioni relative al luogo e alla data di nascita di una persona;¹⁰²
- `<death>`: contiene informazioni relative al luogo e alla data di morte di una persona;¹⁰³
- `<sex>`: specifica il sesso di un organismo;¹⁰⁴
- `<nationality>`: contiene una descrizione della nazionalità o cittadinanza presente o passata di una persona.¹⁰⁵

⁹⁸ <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-person.html>.

⁹⁹ <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-persName.html>.

¹⁰⁰ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-forename.html>.

¹⁰¹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-surname.html>.

¹⁰² <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-birth.html>.

¹⁰³ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-death.html>.

¹⁰⁴ <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-sex.html>.

¹⁰⁵ <https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-nationality.html>.

6.4 Realizzazione della ricerca

Per la realizzazione della ricerca per mezzo dei metadati sono partita creando una form HTML con la quale interagisce l'utente:



Figura 12: form ricerca catalogo

Nella figura ho riportato la form HTML realizzata. Come si può notare l'utente può inserire i dati da ricercare negli appositi campi di input.

Una volta compilata la form HTML, grazie al button *cerca*:

```
<button id="cerca_catalogo" onclick="cerca_catalogo()">Cerca</button>
```

Si avvia la funzione JavaScript¹⁰⁶ *cerca_catalogo()*:

```
function cerca_catalogo() {  
  
    form.submit(); // Invia il modulo  
  
}
```

Figura 13: funzione *cerca_catalogo*

Invia il modulo al server per l'esecuzione della ricerca nel catalogo.

Per ricercare i dati inseriti in input nella form HTML ho implementato specifiche funzioni XQuery.

¹⁰⁶ <https://github.com/DilettaLelli99/Vocidall-Inferno>.

Per ogni campo di ricerca si crea un frammento HTML contenente un template che ospiterà i risultati della ricerca sul campo.

Prendiamo come esempio la ricerca effettuata sul nome:

Inserendo “Idek” all’interno del campo nome (compilando anche gli altri campi) otterremo il risultato esemplificato:



Figura 14: risultato testimonianza Idek Wolfowicz

Cliccando sull’immagine si apre la pagina relativa alla testimonianza di Idek Wolfowicz.



Figura 15: Schermata testimonianza Idek Wolfowicz

Il codice XQuery per effettuare questa ricerca è il seguente:

```

declare function app:cercaForename($nome as xs:string) as element()* {
  for $doc in collection("/db/apps/proget/xml")//tei:person[tei:persName/tei:forename = $nome] (: Itera su tutti i documenti nella collezione :)
  return $doc
  (: Restituisce i documenti che soddisfano la condizione :)
};

declare function app:cerca_catalogo_nome($node as node(), $model as map(*), $nome as xs:string?) {
  let $result := app:cercaForename($nome)

  return
  if (exists($result)) then
    <div class="risultato">
      {
        for $doc in $result
        let $forename := $doc//tei:persName/tei:forename
        let $surname := $doc//tei:persName/tei:surname
        let $nome-format0 := concat($forename, '_', $surname)
        let $nome-format := replace($nome-format0, '_', ' ')
        let $imageName := concat($forename, '_', $surname, '.jpeg')
        return
        <div class="testimonianzacatalogo" onclick="riportaAllaTestimonianza('{ $nome-format }')'" style="background-color: #BCDCDC; color: black;">
          <h3 style="font-size: 16px;">{concat("Testimonianza di ", $nome-format)}</h3>
          
        </div>
      }
    </div>
  else
    "file non presente"
};

```

Figura 16: funzione ricerca nome¹⁰⁷

Queste funzioni sono utilizzate per effettuare una ricerca all'interno di una collezione di documenti XML basata sul valore dell'elemento *forename* e restituire i risultati formattati in una struttura all'interno di un frammento HTML. La funzione *app:cercaForename* cerca i documenti, mentre *app:cerca_catalogo_nome* gestisce la presentazione dei risultati della ricerca.

¹⁰⁷ <https://github.com/DilettaLelli99/Vocidall-Inferno/blob/2a2e87edec6b989e2e0397a0b4d5de0f663ab444/app.xqm#L268-L295>.

Capitolo 7

Conclusioni e sviluppi futuri

L'obiettivo del lavoro di tesi è stato quello di aggiungere nuove funzionalità rispetto a quelle implementate precedentemente da altri progetti di tesi. In particolare, si voleva dare un'armonia generale al progetto *Voci dall'inferno*, unendo i prototipi di applicazioni pre-esistenti. Tra le funzionalità più importanti da integrare c'è stata la sezione di search avanzato che sfruttasse i metadati presenti nel *teiHeader* e la possibilità di visualizzare la trascrizione delle testimonianze con alcuni fenomeni codificati per mezzo delle convenzioni CHAT e DT1. Il lavoro svolto si riassume mediante i seguenti punti:

- Acquisizione in formato digitale della testimonianza registrata grazie all'utilizzo del software *Audacity*;
- Processo di ascolto e trascrizione della testimonianza;
- Codifica XML (secondo lo standard TEI p5) della testimonianza;
- Creazione dell'applicazione web (basata sulle precedenti esperienze di progetto);
- Indicizzazione dei tag presenti nel *teiHeader* al fine di svolgere la ricerca avanzata nel catalogo;
- Creazione di una sezione di search avanzato per la ricerca nel catalogo attraverso i metadati presenti nel *teiHeader*;
- Realizzazione della vista testuale con la possibilità di scegliere tra due convenzioni di codifica differenti per la visualizzazione di alcuni fenomeni codificati dopo aver trascritto della testimonianza

In futuro altri possibili sviluppi sono necessari per rendere l'applicazione competa e pronta per il rilascio. Innanzitutto, bisognerà rivedere le codifiche XML-TEI ed inserire i tag necessari nei vari *teiHeader* che permettano di effettuare le ricerche nel catalogo. Sarà così possibile effettuare ricerche basate su metadati o su argomenti riassunti all'interno del regesto: in quest'ottica sarebbe interessante vedere chi erano i presenti durante la registrazione della testimonianza o fare analisi su temi ricorrenti nelle varie testimonianze (come, per esempio i "tipi" di inferno vissuti). Inoltre, si potrà lavorare sulle funzioni già realizzate per la ricerca nel catalogo per trasformarla in una vera e propria ricerca filtrata.

Per quanto riguarda la visualizzazione delle trascrizioni con convenzioni differenti, infine, sarebbe interessante sviluppare una funzione in grado di scegliere dinamicamente, da un catalogo di convenzioni, quella da utilizzare. A tal fine, ulteriori sviluppi potranno prevedere la gestione di nuovi fenomeni da visualizzare con le opportune convenzioni.

Bibliografia

Del Grosso, A.M., a.a. 2021-2022, Slide Codifica di Tesi. Introduzione XML Markup.

Del Grosso A.M., a.a. 2021-2022 Slide Codifica di Tesi. Presentazione del Corso Codifica di Testi.

Bianconi V., Tesi di laurea: L'applicazione "Voce dei Margini": presentazione interrogazione e analisi linguistica delle postille di Giorgio Bassani al volume La scuola dell'uomo di Guido Calogero, Pisa, Università di Pisa, 2021-2022.

Capone S., Tesi di laurea: 'Archiviare' gli Inferni di ieri e di oggi, Pisa, Università di Pisa, 2021-2022

Schmidt, T. (June 2011). A TEI-based Approach to Standardising Spoken Language Transcription, Journal of the Text Encoding Initiative [Online]

Walmsley Priscilla, XQuery Search Across a Variety of XML Data, p. 2

Siegel, E. e Retter, A. (2015). eXist. United States of America, O'Reilly Media, pp. 3-4

Andrei Mihaila, Tesi di laurea: Inferno di spine: Nedo Fiano e la sua testimonianza, Pisa, Università di Pisa, 2022-2023

Erika Deboni, Tesi di laurea: Una voce dall'inferno: la testimonianza di Ida Marcheria, Pisa, Università di Pisa, 2020-2021

Ehlich, K. 2003. "HIAT: A Transcription System for Discourse Data." In Talking Data: Transcription and Coding in Discourse Research, ed. J. Edwards and M. Lampert, 123–148. Hillsdale, NJ: Erlbaum

Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock, and S. Uhmann. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)". *Gesprächsforschung* 10: 353–402.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. 3rd edition
Mahwah, NJ: Lawrence Erlbaum Associates

Discourse Transcription: DuBois et al. 1993

Göteborg Transcription Standard: Nivre et al. 1999

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

Erik Siegel, Adam Retter (2014). "eXist", O'Reilly Media, Inc.

Sitiografia

Memoriarchivio:

<https://memoriarchivio.org/>

CDEC, Centro di Documentazione Ebraica – Digital Library:

<http://digital-library.cdec.it/cdec-web/>

Anna Segre:

<http://www.annasegre.it>

Archivi Arolsen:

<https://arolsen-archives.org/en/>

International Holocaust Remembrance Alliance:

<https://www.holocaustremembrance.com/>

Sito ufficiale città Tarnow:

<https://www.tarnow.pl/>

Il patto tedesco sovietico:

<https://encyclopedia.ushmm.org/content/it/article/german-soviet-pact>

Sito ufficiale città di Przemyśl:

<https://przemysl.pl/70302/miasto-przemysl-strona-oficjalna.html>

Yad Vashem:

<https://www.yadvashem.org/>

Wikibook XQuery:

<https://en.wikibooks.org/wiki/XQuery>

Yeoman Generator:

<https://yeoman.io/generators/https://github.com/eXist-db/generator-exist>

MySQL:

<https://www.mysql.com/it/>

SQL Server:

<https://www.microsoft.com/en-us/sql-server/sql-server-2022>

PostgreSQL:

<https://www.postgresql.org/>

Oracle:

<https://www.oracle.com/it/>

Cassandra apache:

https://cassandra.apache.org/_/index.html

Mongodb:

<https://www.mongodb.com/it-it>

Neo4j:

<https://neo4j.com/>

Lucene:

<https://exist-db.org/exist/apps/doc/lucene>

TEI:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD11>

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

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

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

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

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

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

<https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-sex.html>
<https://tei-c.org/release/doc/tei-p5-doc/it/html/ref-nationality.html>
<https://tei-c.org/release/doc/tei-p5-doc/it/html/TS.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-del.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-placeName.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-persName.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-gap.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-choice.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-orig.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-reg.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-vocal.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-desc.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-incident.html>
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-pause.html>

Catalogo ICCU SBN:

<https://www.iccu.sbn.it/it/>