

# Scrivere (quasi) la stessa cosa

La trascrizione come atto interpretativo nella pratica della storia orale

a cura di **Francesca Di Meo, Roberta Garruccio, Francesca Socrate**

[*ride*]



## **Storia Orale**

**Questa collana è uno spazio dove riflettere sul passato attraverso il racconto e la memoria. Mettendo al centro la relazione tra testimoni e studiosi, la storia orale ragiona criticamente sulle dimensioni del linguaggio e della trasmissione del ricordo nel tempo e nella società, analizza gli eventi e i territori a partire dalle persone che li hanno attraversati. La collana, raccogliendo il testimone di una solida tradizione italiana e internazionale di ricerca sociale, vuole aprirsi alla pluralità di forme narrative e temi che meglio possono aiutare a comprendere la dimensione umana della storia.**

**Direttrice: Gabriella Gribaudi**

**Coordinatore: Giovanni Pietrangeli**

**Comitato scientifico:**

**Stefano Bartolini**

**Bruno Bonomo**

**Andrea Brazzoduro**

**Marco Buttino**

**Antonio Canovi**

**Alessandro Casellato**

**Giovanni Contini**

**Caterina Di Pasquale**

**Antonio Fanelli**

**Roberta Garruccio**

**Martina Giuffrè**

**Enrico Grammaroli**

**Gloria Nemeč**

**Sandro Portelli**

**Gabriele Progljo**

**Omerita Ranalli**

**Francesca Socrate**

**Anna Maria Zaccaria**

**Sara Zanisi**

**Le pubblicazioni sono sottoposte a peer review, a cura del comitato scientifico della collana.**

# **Scrivere (quasi) la stessa cosa**

**La trascrizione come atto interpretativo  
nella pratica della storia orale**

a cura di

**Francesca Di Meo**

**Roberta Garruccio**

**Francesca Socrate**

Proprietà letteraria riservata  
© 2022 editpress, Firenze  
Via Lorenzo Viani, 74  
50142 Firenze - Italy  
[www.editpress.it](http://www.editpress.it)  
[info@editpress.it](mailto:info@editpress.it)  
Printed in Italy

Scrivere (quasi) la stessa cosa /  
a cura di F. Di Meo, R. Garruccio, F. Socrate-  
Firenze : editpress, 2022. -  
204 p. ; 21 cm  
(Storia orale ; 6.)  
ISBN 979-12-80675-20-0  
Permalink formato digitale:  
<[digital.casalini.it/9791280675200](http://digital.casalini.it/9791280675200)>

## Sommario

- 7 Prefazione. Scrivere (quasi) la stessa cosa  
*Bianca Pastori*
- 13 Leggere come un discorso parlato. La riflessione sulla trascrizione nella storia orale anglosassone e italiana  
*Patrick Urru*
- 29 L'infrastruttura CLARIN e il servizio di trascrizione multilingue T-Chain  
*Stefania Scagliola, Silvia Calamai*
- 51 Indicizzazione o trascrizione? Indicizzare per tutelare l'ascolto. L'esempio di *Patrimonio Orale*  
*Donatella Orecchia*
- 67 Il parlato narrativo tra oralità e scrittura: caratteristiche lessico-sintattiche  
*Alberto Bramati*
- 97 La simulazione del parlato nella narrativa italiana del secondo Novecento  
*Gianni Turchetta*
- 117 Scritture, parole, voci e figure. Interviste di artisti migranti a confronto  
*Graziella Bonansea*

- 133 Trascrivere e interpretare. Tradire per restare fedeli?  
*Fabio Dei*
- 149 Nelle mani giuste. Trascrivere e sottotitolare una fonte  
orale in lingua dei segni  
*Luca Des Dorides*
- 169 Voci dal manicomio di Maggiano rievocate in  
eterodirezione  
*Marco Cavalcoli*
- 173 Prove tecniche di trascrizione: un'esperienza personale  
*Patrik Urru*
- 195 Autori e curatori

# L'infrastruttura CLARIN e il servizio di trascrizione multilingue T-Chain

*Stefania Scagliola, Silvia Calamai*

Questo contributo mira a presentare alla comunità italiana degli oralisti l'infrastruttura di ricerca europea CLARIN e il prototipo di un servizio di trascrizione multilingue chiamato T-chain. Questo servizio è stato sviluppato con il supporto di CLARIN da una rete internazionale di esperti provenienti da diversi campi tecnologici e umanistici di cui le autrici fanno parte, con l'obiettivo di rafforzare l'integrazione degli strumenti digitali nella ricerca umanistica, storica e sociale. Le attività del gruppo di ricerca sono descritte sul sito <[www.speechandtech.eu](http://www.speechandtech.eu)>.

La descrizione di CLARIN e della T-chain è accompagnata da riflessioni sui prerequisiti per una collaborazione interdisciplinare, onde attuare una integrazione consapevole della tecnologia all'interno della ricerca umanistica. In più, vengono spiegate le funzioni elementari dei sistemi di riconoscimento vocale e le condizioni per arrivare ad ottenere i migliori risultati. Sono descritti anche i rischi e i limiti associati alla dipendenza dal settore privato per questo tipo di tecnologia. Il contributo si conclude con un esempio diretto: come usare il servizio di trascrizione partendo da una intervista condotta in italiano dalla prima autrice.

## 1. Infrastrutture di ricerca digitali per la storia orale e la linguistica

### *Il significato e l'obiettivo della infrastruttura CLARIN*

CLARIN è una infrastruttura di ricerca internazionale, attiva dal 2012 con lo scopo di creare uno spazio pubblico virtuale

dove si possono unire servizi, dati e strumenti relativi allo studio della lingua, sia in forma scritta, che parlata e multimodale. Le risorse sono accessibili tramite il sito web <www.clarin.eu>, a cui si può accedere sulla base di un sistema di accreditamento. Centri di ricerca e università situati in Europa forniscono il loro contributo sia in forma di dati e servizi che in forma finanziaria<sup>1</sup>.



Immagine 1. Homepage del sito di CLARINEU.

Questo tipo di collaborazione tramite federazione di servizi infrastrutturali virtuali entra nel quadro della politica dell'Unione Europea, che mira a creare uno spazio comune per ricercatori e cittadini in tutti i paesi dell'Europa: l'European Open Science Cloud<sup>2</sup>. Altre simili infrastrutture per la ricerca umanistica pan-europee sono <<https://www.dariah.eu>> e <<https://www.operas-eu.org>>.

Nonostante l'orientamento verso la ricerca linguistica (come del resto rivela l'acronimo *CLARIN* - Common Language Resources and Technology Infrastructure) l'obiettivo del consorzio è più ampio: si mira cioè a facilitare la ricerca nell'intero campo delle scienze umanistiche e sociali. La messa a disposizione dei dati è il primo, fondamentale, passo: i ricercatori e le ricercatrici possono accedere al catalogo virtuale di CLARIN per selezionare e scaricare "oggetti" utili alle loro finalità di ricerca (per esempio, lo stori-

co tedesco potrebbe scegliere di studiare una collezione di libri di cucina<sup>3</sup>, il sociologo anglosassone potrebbe analizzare una collezione di blogs sul tema del cambiamento climatico<sup>4</sup>, i ricercatori nel campo della linguistica hanno l'imbarazzo della scelta fra corpora di parlato spontaneo<sup>5</sup> e controllato in varie lingue).

Ma le aspirazioni di CLARIN vanno oltre questo servizio: l'obiettivo ultimo riguarda anche la possibilità di produrre idee e percorsi innovativi nel campo della ricerca, usando strumenti e dati da discipline diverse per paragonare temi sociali e culturali.

### *Il valore aggiunto di attraversare i confini fra le varie discipline*

Per realizzare questo obiettivo ambizioso occorre ben capire le pratiche quotidiane eterogenee dei diversi ricercatori, e provare a superare i vincoli imposti dalle singole discipline<sup>6</sup>. Proprio per intercettare i bisogni di comunità differenti, e per trovare un punto di partenza comune, nonché un vocabolario condiviso, sono stati organizzati diversi workshops dalla nostra rete *speech and text*<sup>7</sup>: ricercatori di varie discipline – sociolinguisti, linguisti computazionali, storici, sociologi, tecnologi e informatici ed archivisti, che in qualche forma lavorano con fonti parlate – sono stati invitati a lavorare con strumenti digitali che occorrono nelle fasi successive del processo di ricerca: strumenti per la trascrizione, l'annotazione, l'analisi del contenuto e per il riconoscimento di emozioni<sup>8</sup>. L'immagine 2 mostra il poster che sintetizza i lavori, gli obiettivi e i commenti emersi durante il workshop tenuto all'università di Monaco di Baviera nel 2016.

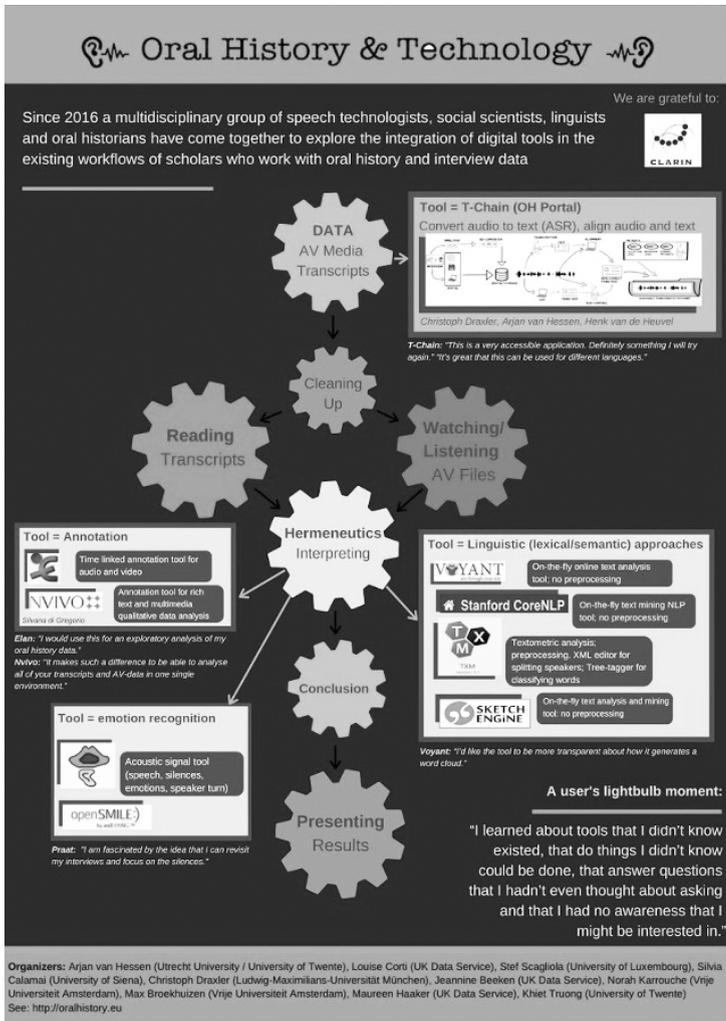


Immagine 2. Poster del workshop multidisciplinare *Oral History and Technology*, Monaco di Baviera (2016).

L'entusiasmo e l'interesse reciproco durante il workshop erano promettenti, come si può leggere nella citazione di uno dei partecipanti sul poster: «A user's lightbulb moment».

D'altra parte, gli ostacoli che non permettono l'integrazione di nuove tecnologie nel proprio flusso di lavoro non sono di poco

conto. Non è facile modificare il proprio vocabolario di riferimento, non è facile imparare ad usare strumenti nuovi e integrarli nel proprio flusso di lavoro: l'investimento di tempo richiesto appare considerevole, e il ricercatore è disposto a modificare le proprie abitudini solo se il valore aggiunto dell'approccio alternativo è ovvio. Ricercatori maturi sono già legati a tutta una serie di criteri riguardo la validità e qualità dei risultati della loro ricerca entro la propria comunità accademica. Uscire da questa "comfort zone" e trovare nuove strade non solo è impegnativo, ma è anche rischioso, in quanto il cosiddetto "paradigma esplorativo", di solito, non possiede uno stato riconosciuto nel proprio campo accademico<sup>9</sup>. Forse il pubblico più adatto a queste esplorazioni è quello al livello di dottorati all'inizio della loro carriera, che hanno più spazio per esplorare un approccio multidisciplinare predisposto a studiare modi diversi di interpretazione e a paragonare i risultati. Arricchisce il potenziale analitico nel senso che aiuta a capire come lo stesso fenomeno può risultare in significati molto diversi<sup>10</sup>.

## 2. La trascrizione, l'innovazione tecnologica e la dipendenza dall'industria

### *Perché occorre la trascrizione*

I profili eterogenei dei ricercatori che lavorano con i dati del parlato sono rispecchiati nei vari tipi di trascrizioni utilizzate. È ben noto che le convenzioni trascrittive sono specifiche di singole comunità di ricerca, le quali elaborano un proprio armamentario di simboli e di strategie per rappresentare l'evanescenza dell'atto di parola. Studiare il parlato limitandosi ad ascoltarlo è quasi impossibile, in quanto il nostro udito è capace solo di catturare queste onde istantaneamente. Per poter dunque studiare e paragonare quello che viene detto, è necessario trasformare quello che sentiamo in una forma tangibile e durevole. Questo si fa, usando simboli come

caratteri, numeri e segni ortografici, seguendo una convenzione che appartiene ad una specifica comunità di ricerca<sup>11</sup>.

Nel passato, quando si usavano la penna e l'inchiostro o la macchina da scrivere, i termini "tangibile" e "durevole" avevano un significato letterale, ma con l'introduzione del computer, del web e dell'ipertesto il significato è diventato figurativo in quanto la materialità della trascrizione è stata sostituita da dati digitali<sup>12</sup>. Tale caratteristica rende possibile leggere e rileggere la trascrizione su uno schermo in dettaglio, o rapidamente scansionarla, alla ricerca di un termine o di un tema specifico. In teoria si dovrebbe sempre verificare un passaggio tornando al parlato, ma molto spesso questo lavoro impegnativo viene saltato. La trascrizione può essere dunque considerata come una rappresentazione durevole del segnale acustico, o in carta o in forma digitale, a cui sono state tolte tutte le caratteristiche effimere e non verbali.

### 3. La trascrizione automatica del linguaggio

La trascrizione automatica del linguaggio (in sigla, ASR, Automatic Speech Recognition) produce una trascrizione scritta del parlato attraverso l'uso di software, senza intervento umano diretto.

Il software è in grado di riconoscere solo ciò che già conosce sulla base di una sorta di "allenamento". Questo significa che deve essere alimentato con molti, moltissimi dati per arricchire la sua potenza. Per arrivare ad una percentuale di riconoscimento soddisfacente occorre un addestramento con almeno 50-100 ore di parlato che è stato trascritto manualmente, in combinazione con un vocabolario di termini specifici sul soggetto del documento orale, in forma sia fonetica che ortografica, incluso il modo in cui vengono pronunciati, e la struttura grammaticale in cui i termini vengono usati<sup>13</sup>.

Il software di riconoscimento ASR compie due azioni allo stesso tempo. In primo luogo, paragona il segnale acustico ad un modello probabilistico di segnali-fonemi che è stato creato tramite l'ad-

destramento; in secondo luogo, paragona il linguaggio che viene espresso ad un modello probabilistico basato su un vocabolario specifico. Più sono i dati disponibili per l'addestramento, più la diversità può essere rappresentata nel modello, e più alta sarà la percentuale del riconoscimento<sup>14</sup>. Ricercatori del parlato potrebbero essere di aiuto, se disposti a condividere le loro trascrizioni manuali insieme alle loro registrazioni<sup>15</sup>.

In un articolo in cui viene spiegato a non esperti come funziona l'ASR, Christoph Draxler, tecnologo del Bavarian Speech Centre a Monaco di Baviera e uno delle forze motrici dietro lo sviluppo della T-chain, descrive le condizioni per poter usufruire massimamente della tecnologia ASR. Occorre un contesto comunicativo e ambientale "controllato". I migliori risultati vengono ottenuti registrando una sola voce che detta un messaggio usando un microfono di buona qualità in una stanza senza altri segnali acustici. Questo è naturalmente il contesto opposto a quello in cui lavora l'oralista o il sociolinguista: chi lavora sul campo è infatti coinvolto in un discorso spontaneo con molte ripetizioni, interruzioni, rumori di fondo, uso di dialetto e riferimenti a luoghi e persone spesso sconosciuti al software<sup>16</sup>. Sono molte le situazioni in cui occorre "improvvisare" per poter condurre un'intervista, ed accettare la presenza di terze persone, animali domestici, lavori in corso, tazze che vengono spostate sul tavolo, pioggia. Per poter "riconoscere" il segnale acustico e poterlo distinguere da tutti i "rumori", c'è ancora molto lavoro da fare.

#### 4. Il rapporto ambiguo con l'industria

Non è un mistero che gli investimenti siano fatti in applicazioni per attività che offrono il maggior profitto. Anche nell'ambito della ricerca sul parlato, per mobilitare risorse umane e tecnologiche è necessario un interesse commerciale o politico. La storia della tecnica è piena di esempi al riguardo: lo sviluppo di una versione del registratore con nastro magnetico a prezzo moderato si colloca nel

contesto di un vantaggio tecnologico degli Stati Uniti sulla Germania, durante la seconda guerra mondiale<sup>17</sup>; lo sviluppo della macchina da scrivere è scaturito dall'obiettivo di organizzare il lavoro in una azienda in modo più efficiente, per aumentare il profitto<sup>18</sup>.

Anche nel contesto dello sviluppo del riconoscimento automatico del parlato, le prime applicazioni miravano a guadagnare tempo sostituendo personale o a monitorare comunicazioni che avrebbero potuto rappresentare una minaccia per lo Stato (intercettazione delle comunicazioni di un nemico, attivisti o terroristi). In effetti, uno degli investimenti più cospicui nello sviluppo dell'ASR è stato fatto dalla Defense Advanced Research Projects Agency negli Stati Uniti (DARPA) negli anni Settanta<sup>19</sup>. La ricerca accademica, in particolare nelle scienze umane, ha sempre potuto beneficiare di questi progressi.

Tuttavia, il passaggio alla tecnologia digitale e all'elaborazione probabilistica, che si basa sull'addestramento da grandi dati, ha cambiato il gioco. Per applicare questa tecnologia con successo, è necessario il controllo dell'intera catena di elaborazione, dal momento in cui vengono creati dati, fino alla consegna di un prodotto o servizio. Offrendo servizi gratuiti, le aziende permettono agli utenti inesperti di accedere facilmente ai loro servizi, usufruendo in questo modo dei dati degli utenti stessi. Con questo vantaggio e il loro successo commerciale viene determinata la direzione dello sviluppo della tecnologia. Nel campo accademico mancano le risorse per sviluppare una tecnologia con lo stesso livello di perfezione che poi si orienti verso le esigenze specifiche dell'area umanistica (con le questioni etiche e legali ad essa associate)<sup>20</sup>.

In più esiste il problema cruciale della mancanza di trasparenza dal lato delle aziende commerciali – conseguenza non secondaria di ciò che potremmo definire “comodità digitale”. Le aziende hanno sempre protetto la loro posizione sul mercato, mantenendo segreta la chiave del loro successo, ma nell'epoca analogica esisteva una distanza fra il mondo del ricercatore umanistico e quello dell'azienda in cui veniva sviluppata una tecnologia. Con l'introduzione del web questa distanza è (apparentemente) scompar-

sa. Ora è facilissimo accedere a una piattaforma di un'azienda commerciale online che offre la possibilità di elaborare una quantità limitata di parlato con un software ASR.

Queste sono alcune tra le piattaforme più conosciute che offrono questo tipo di servizi per alcune delle lingue europee. È da notare come per la lingua italiana siano disponibili solo software proprietari.

Per l'inglese e l'olandese esistono servizi open source:

- LST by Radboud University Nijmegen (open source and free) <[https://webservices-lst.science.ru.nl/oral\\_history](https://webservices-lst.science.ru.nl/oral_history)>.
- WebASR <<https://www.webasr.org>> by Sheffield University.

Per l'Italiano esistono solo servizi commerciali:

- IBM Watson <<https://www.ibm.com/cloud/watson-speech-to-text>>.
- Google, YouTube <[cloud.google.com/speech-to-text/](https://cloud.google.com/speech-to-text/)>.
- European Media Laboratory <[www.eml.org](http://www.eml.org)>.

Anche se i risultati dei servizi commerciali sono ottimi, non si sa come siano stati ottenuti, né è sempre chiaro cosa succeda con i dati dopo che sono stati processati. Nel caso della storia orale, i dati includono nomi di persone e dialoghi su esperienze di carattere molto personale. Il bisogno di trasparenza riguardo a come si è arrivati ad un certo risultato e l'attenzione da riservare alla tutela della riservatezza sono due questioni che la tecnologia del riconoscimento del parlato automatico ancora non è stata in grado di risolvere. E proprio per questa ragione che CLARIN ha deciso di finanziare lo sviluppo della T-chain. La dipendenza da meccanismi e tecnologie il cui sviluppo è inevitabilmente legato allo sfruttamento commerciale rende la ricerca accademica non completamente libera. Dipendere da servizi esterni significa anche non avere il controllo dell'intero ciclo di vita dei dati (Fortunato et al., 2020)<sup>21</sup>.

## 5. La catena di trascrizione (T-Chain)

La T-Chain è stata sviluppata con l'obiettivo di fornire una piattaforma multilingue e facile da usare per trascrivere il parlato<sup>22</sup>. Una versione Beta è stata presentata da Christoph Draxler per la prima volta durante un CLARIN workshop organizzato ad Arezzo il 10 maggio 2017 per ricercatori di storia orale, linguisti, umanisti e tecnologi del parlato.

La T-Chain utilizza alcuni servizi già esistenti sviluppati al Bavarian Archive of Speech, come l'allineamento di parole a suoni. Di conseguenza, la prima versione del servizio era adeguata alle esigenze e abitudini dei tecnologi del parlato. Attraverso una serie di workshop che si sono svolti dopo Arezzo nel 2017 l'interfaccia della T-Chain è stata adattata alle esigenze dei ricercatori con un orientamento non tecnologico. È stata poi resa accessibile attraverso l'infrastruttura CLARIN, con l'obiettivo di offrire una risorsa che rende i ricercatori meno dipendenti dal software commerciale<sup>23</sup>.

Naturalmente questo è solo un primo passo e ci sono ancora molti ostacoli da superare. Prima di tutto, l'idea di un servizio libero universale per tutte le lingue non è ancora fattibile. La T-Chain tenta di massimizzare il componente open source, presentando una combinazione di strumenti in una sola piattaforma che sono in parte gratuiti, in parte ancora forniti dall'industria. L'ideale sarebbe riuscire gradualmente a sostituire i servizi commerciali con quelli open source.

In questo momento la piattaforma è capace di processare solo files WAV. Quindi per far funzionare l'ASR le registrazioni devono essere prima convertite in questo formato, un processo abbastanza facile da eseguire utilizzando strumenti come GoldWave, Audacity o `To_Wave_Convertor`.

L'inclusione di servizi commerciali è necessaria per poter processare varie lingue. D'altra parte, il vantaggio di avere una serie di strumenti e servizi per varie lingue in un solo ambiente, ha come svantaggio la limitata dimensione dei file (da caricare e dunque da pro-

cessare): tale limite di dimensione è infatti posto dalle aziende per poter usufruire di una quota mensile gratuita di ASR. Questo significa che i file devono essere tagliati in frammenti più piccoli. La dipendenza da servizi esterni senza legami diretti o responsabilità verso il centro di ricerca BAS significa che, in caso di errore, può risultare difficile risolvere il problema. Di seguito vengono descritte le varie fasi del processo, in cui si è tentato di processare un piccolo frammento di un'intervista condotta in italiano con una persona anziana.

### *Le varie fasi: guida all'utilizzo della T-Chain<sup>24</sup>*

Il servizio della T-Chain è composto dalle seguenti fasi: l'utente **AGGIUNGE** il file, poi deve essere **VERIFICATO**, e per ultimo viene **PROCESSATO**.

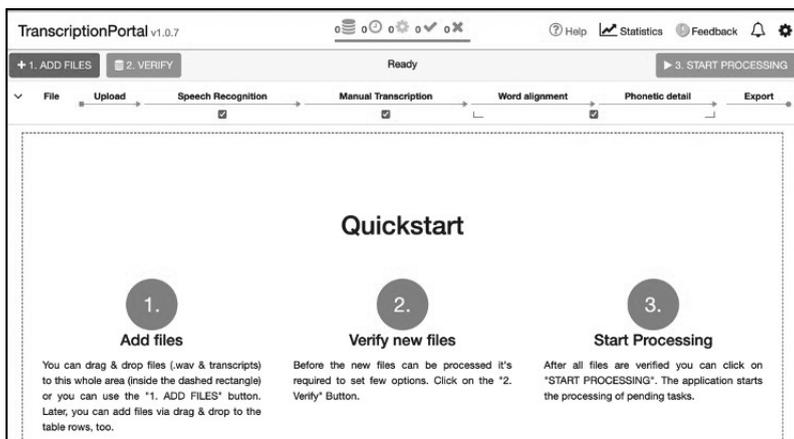


Immagine 3.

Nella prima riga in mezzo, vengono rispecchiate le fasi in una serie di simboli: un fusto grigio rappresenta il caricamento del file, l'orologio blu indica che un compito deve essere eseguito, la ruota dentata gialla gira durante l'elaborazione, la spunta verde indica il completamento del compito, e la croce rossa indica la presenza di un errore.

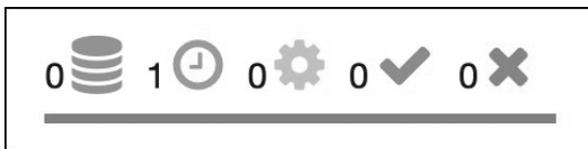


Immagine 4.

Nelle colonne nere vengono descritte le varie funzioni nel processo di elaborazione: il processo di ASR (Speech Recognition), la possibilità di correggere ASR manualmente (Manual Transcription), la possibilità di allineare il testo all'audio (Word Alignment), la possibilità di produrre una trascrizione con dettagli fonetici (Phonetic detail).

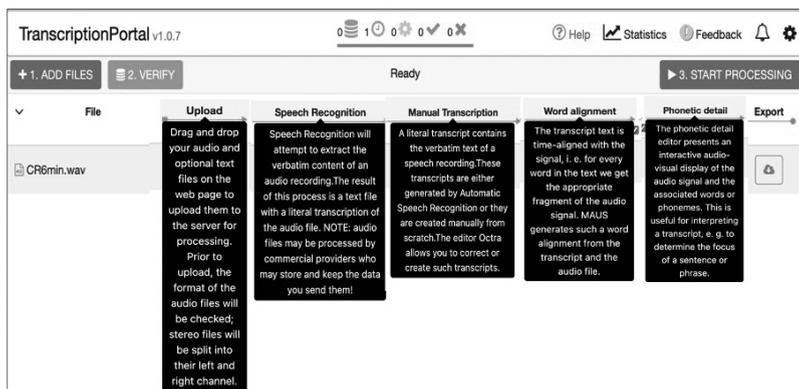


Immagine 5.

Prima di iniziare il corso del processo occorre verificare il file.



Immagine 6.

Cliccando su VERIFY si apre una finestra dove viene spiegato come funziona il processo. Compare anche un menu per la scelta dei servizi, della lingua.

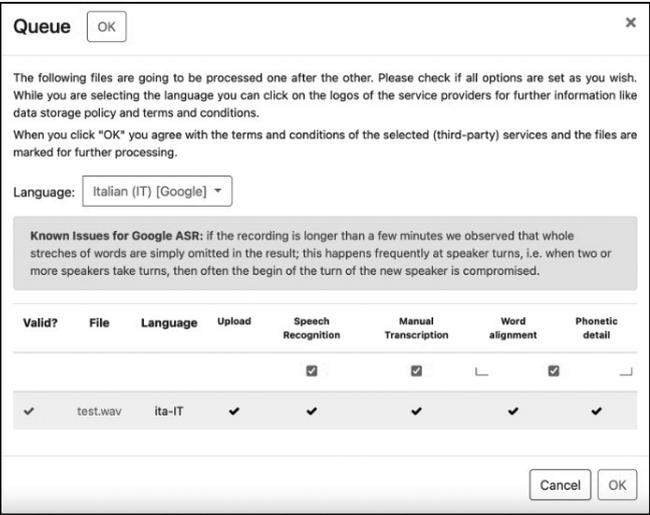


Immagine 7.

Accanto alla campo LANGUAGE si apre un menu con i vari servizi di ASR che esistono per le lingue inglese, olandese, italiano e tedesco.

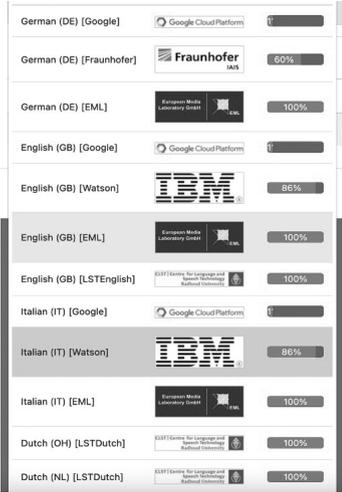


Immagine 8.

Scegliendo una opzione, appaiono caratteristiche e condizioni del servizio.

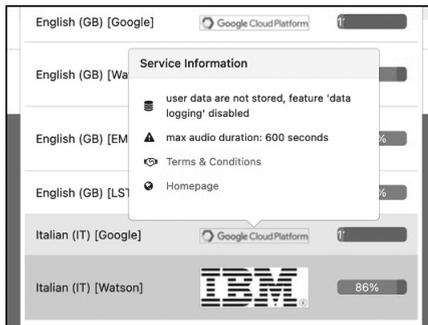


Immagine 9.

Il primo test fatto attraverso Italian Google con un file troppo grande ha prodotto il messaggio visibile nel box che segnala un errore con la croce rossa.

The following files are going to be processed one after the other. Please check if all options are set as you wish. While you are selecting the language you can click on the logos of the service providers for further information like data storage policy and terms and conditions.

When you click "OK" you agree with the terms and conditions of the selected (third-party) services and the files are marked for further processing.

Language:

One or more audio files do not fit the ASR provider's requirements:

- Maximum audio duration: 00:10:00 (H:m:s)

Incompatible audio files are marked with a red cross on the left. These audio files are not going to be processed.

**Known Issues for Google ASR:** if the recording is longer than a few minutes we observed that whole stretches of words are simply omitted in the result; this happens frequently at speaker turns, i.e. when two or more speakers take turns, then often the begin of the turn of the new speaker is compromised.

Valid?	File	Language	Upload	Speech Recognition	Manual Transcription	Word alignment	Phonetic detail
✘	test2.wav	ita-IT	✓	✓	✓	└	✓

Immagine 10.

Riducendo la grandezza del file, Google accetta il file e si può procedere a processarlo cliccando su OK.

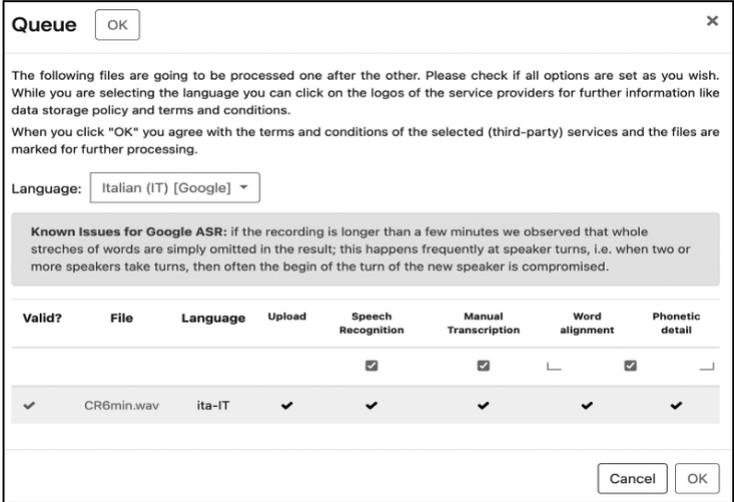


Immagine 11.

Qui si ritorna allo schermo centrale dove si deve attivare il processo cliccando su: 3. START PROCESSING.

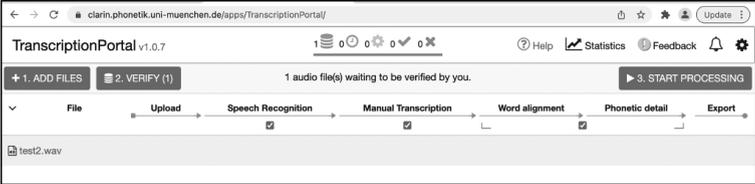


Immagine 12.

Mentre il processo è attivo si vede girare la ruota gialla sopra, e una ruota grigia nella colonna del Speech Recognition.

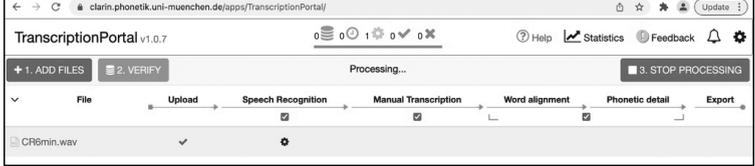


Immagine 13.

Quando il processo è terminato appare un segno di spunta verde nella colonna Speech Recognition. Cliccando sulla colonna si apre un menu che offre l'occasione di scaricare i risultati in vari formati. In questo esempio si sceglie di continuare il processo con la correzione manuale.

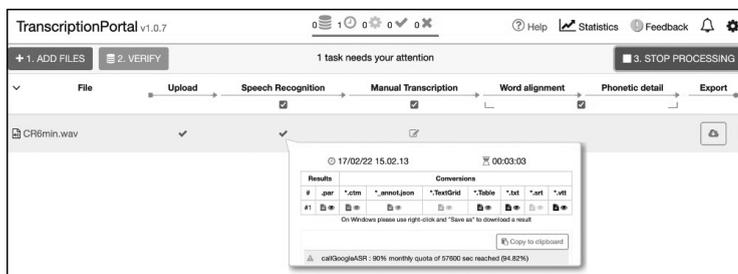


Immagine 14.

Cliccando sotto Manual Transcription si apre il menu del servizio OCTRA (Pömp et al., 2017), l'editor di trascrizione integrato all'interno del browser. Esso offre la possibilità di vedere i risultati in vari modi. Qui sotto si vede il segnale acustico unito alla trascrizione.

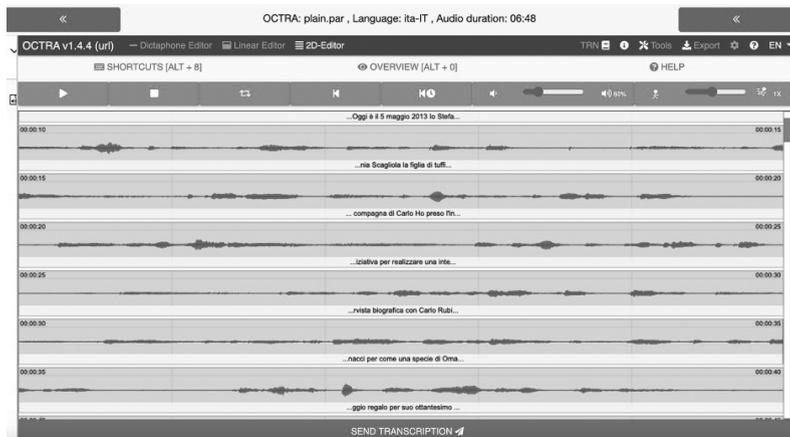


Immagine 15.

Si può anche scegliere per OVERVIEW, che offre una finestra in cui audio e testo sono allineati e divisi in unità più piccole, per essere in grado di riascoltare e correggere il testo in dettaglio.

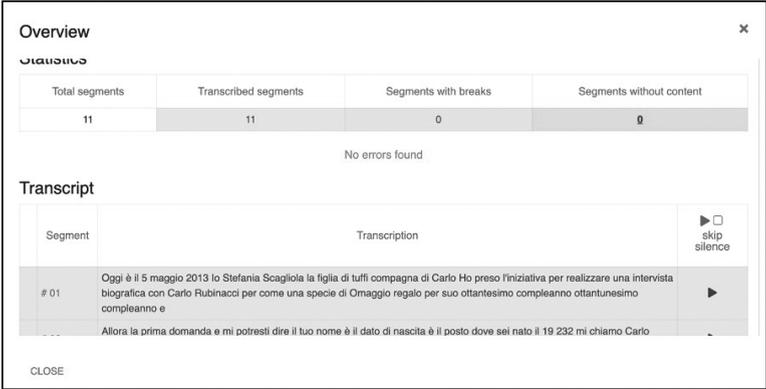


Immagine 16.

Una volta corretto il testo, per poter creare un nuovo documento con il testo corretto, e segmentato in modo da poter essere utile per l'analisi linguistica, occorre cliccare sulla riga verde nella prima finestra.



Immagine 17.

Quando la segmentazione è stata eseguita dal servizio WebMAUS (Kisler et al., 2012) appare una spunta verde sotto Word Alignment, e l'icona della trascrizione nella colonna Phonetic detail.

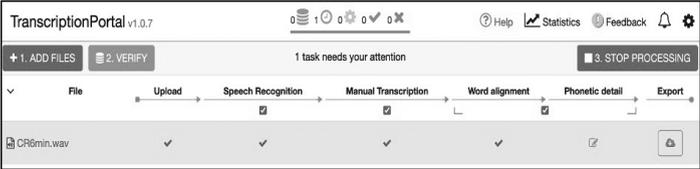


Immagine 18.

Aprendo l'icona appaiono il segnale acustico e il testo corrispondente al livello del segmento, cliccando su “in” e “out” si possono ingrandire o ridurre le dimensioni dei dati.

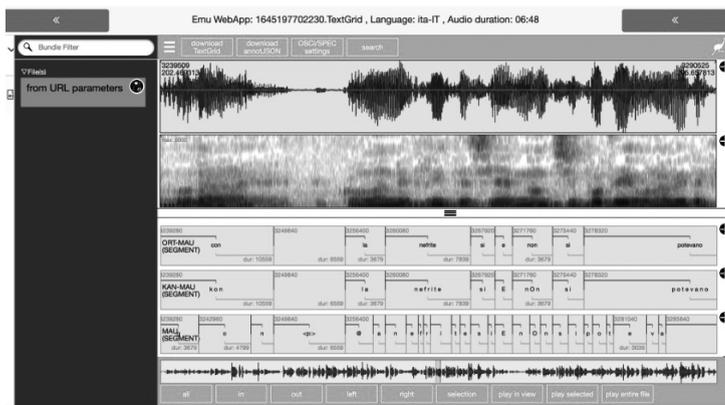


Immagine 19.

La colonna Export serve per scaricare i risultati. Questo può essere fatto in ogni fase del processo sul computer locale in diversi formati: testo semplice, tabelle separate da virgola, Praat TextGrid, Emu Annot-JSON, e ELAN per ulteriori elaborazioni e analisi. Queste esportazioni possono essere fatte in ordine di file o di colonna, cioè molti file in una volta sola.

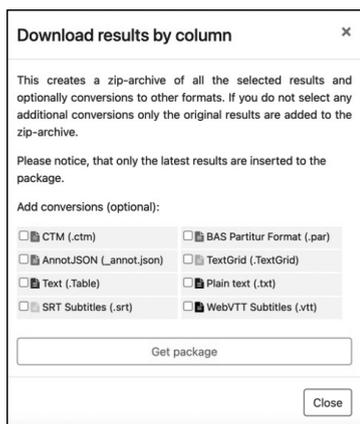


Immagine 20.

Quello che occorre aggiungere dalla prospettiva dello storico orale, è una trascrizione con time codes, che può essere caricata in uno spazio proprio, dove possono essere di nuovo uniti suono e testo, come succede nel servizio per correggere la trascrizione automatica, con l'aggiunta di una interfaccia che offre la possibilità di fare annotazioni.

#### 4. Conclusione

Il percorso appare senz'altro complesso e non facile, ma si tratta di una sfida da affrontare, in una sinergia di intenti, tra tecnologi del parlato e comunità di ricercatori delle scienze umanistiche e sociali che usano il parlato come fonte di ricerca. Una collaborazione intensa e necessaria per perfezionare lo stato della trascrizione in modo efficace, etico e multifunzionale. Intravediamo al momento due necessità: 1) sviluppare un modello adatto per poter processare anche dati non puliti, quelli spesso prodotti in inchieste sul campo o presenti all'interno di archivi orali del passato; 2) formare comunità di giovani ricercatori relativamente all'utilizzo consapevole della tecnologia (come utilizzare adeguatamente microfoni, come registrare, come creare spazi che favoriscono una registrazione "pulita").

Solo un dialogo tra scienze "dure" e scienze "morbide" è in grado di risolvere questa impasse. Gli umanisti devono essere "generosi" e organizzare la loro ricerca in un modo che renda possibile offrire, in futuro, i loro dati per allenare i sistemi, senza che dati privati vengano esposti al pubblico, e con la consapevolezza di permettere l'avanzamento della conoscenza. Dall'altra parte i tecnologi devono rispettare e comprendere queste preoccupazioni e spiegare in modo chiaro in quale modo si possano prendere misure adeguate per tutelare la fonte. Spesso le incomprensioni originano da scarsa conoscenza reciproca: da parte degli umanisti limitata competenza tecnologica e informatica, e nel mondo della ricerca tecnologica poca conoscenza delle questioni etiche e metodologiche

della storia orale. Bisogna anche creare spazio per soluzioni intermedie riguardo la tecnologia, quindi anche ripensare modi per facilitare la trascrizione manuale, come per esempio creare un progetto di crowdsourcing, in cui un gruppo di interessati adotta una collezione di interviste per poterle trascrivere in modo collaborativo.

Anche un sistema di “trriage” sarebbe utile per fare un passo avanti: essere in grado di selezionare sulla base di criteri concreti, quali dati siano adatti da processare con il ASR.

E come ultimo, investire in educazione sia nelle scuole secondarie, che all’università, in una formazione che insegni a gestire dati storici audiovisivi.

La ragione fondamentale per continuare a sviluppare progetti come la T-chain è tentare di proteggere gli interessi di varietà linguistiche, ambiti, tematiche e settori apparentemente privi di potenziale commerciale.

L’ASR accademico può concentrarsi sul parlato non standard, il parlato di persone anziane o su dialetti e lingue di minoranze con poche risorse. I fornitori di ASR finanziati pubblicamente possono diventare parte di un’infrastruttura di dati di ricerca in modo libero, aperto, disponibile a lungo termine, come molti altri servizi pubblici essenziali. È fondamentale che anche la comunità italiana contribuisca a raggiungere questo scopo, a beneficio delle generazioni future.

## Note

<sup>1</sup> F. de Jong, B. Maegaard, D. Fišer, D. Van Uytvanck, A. Witt, *Interoperability in an Infrastructure Enabling Multidisciplinary Research: The case of CLARIN*, in *Proceedings LREC 2020, 12th International Conference on Language Resources and Evaluation*, ELRA 2020, <<https://aclanthology.org/2020.lrec-1.417/>>.

<sup>2</sup> <[https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en)>.

<sup>3</sup> <<http://fedora.clarin-d.uni-saarland.de/sacoco/>>.

<sup>4</sup> <<https://clarino.uib.no/korpuskel/clarino-metadata?identifier=ntap>>.

<sup>5</sup> <<https://www.clarin.eu/resource-families/spoken-corpora>>.

<sup>6</sup> F.D. Jong, R. Ordelman, S. Scagliola, *Audio-visual Collections and the User Needs of Scholars in the Humanities: a Case for Co-Development*, in *Proceedings of the 2nd Conference on Supporting Digital Humanities*, SDH 2011.

<sup>7</sup> I promotori di questo sito web, Christoph Draxler (tecnologia del parlato), Silvia Calamai (sociolinguista), Henk van de Heuvel (tecnologia del parlato), Arjan van Hessen (tecnologia del parlato), Jeannine Beeken (linguista e specialista in archiviazione di dati delle scienze sociali), Stefania Scagliola (storica orale), sono un gruppo di esperti interessati ad esplorare come la tecnologia possa essere integrata nella ricerca che coinvolge le narrazioni parlate: <<https://speechandtech.eu>>.

<sup>8</sup> S. Scagliola, L. Corti, S. Calamai, N. Karrouche, J. van Beeken, A. van Hessen, C. Draxler, H. van den Heuvel, H. Broekhuizen, *Cross disciplinary overtures with interview data. Integrating digital practices and tools in the scholarly workflow*, in K. Simov, M. Eskevich (a cura di), *Proceedings of CLARIN Annual Conference*, 2019, pp. 126-136 [S.l. : s.n.].

<sup>9</sup> A. Woods, *Interdisciplinary authorship*, in *Working Knowledge, Hearing the Voice*, University of Durham 2015.

<sup>10</sup> H. Van den Berg, F. Wester, S. Scagliola (a cura di), *Wat Veteranen Vertellen, verschillende perspectieven op biografische interviews over ervaringen tijdens militaire operaties*, Amsterdam University Press, Amsterdam 2010. Questa è stata la mia esperienza dopo aver concluso un progetto in cui 25 interviste con reduci di guerra olandesi sono state analizzate da un gruppo di ricercatori di diverse discipline: discourse analysis, teologia, antropologia, sociologia, psicologia e storia orale. Vedi anche: <<https://www.watveteranenvertellen.nl/WatVeteranenVertellen.pdf>>, che si può leggere tradotto in italiano sul proprio schermo installando l'applicazione "language" nel browser di CHROME.

<sup>11</sup> L. Spreafico, *Corpora di parlato o corpora di ascoltato?*, in «Rivista italiana di dialettologia», 44, 2020, pp. 37-51, <<https://hdl.handle.net/10863/18479>>.

<sup>12</sup> J. Hughes, K.R. Lang, *Transmutability: Digital Decontextualization, Manipulation, and Recontextualization as a New Source of Value in the Production and Consumption of Culture Products*, in «Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)», 8, 2006, pp. 165a-165a.

<sup>13</sup> Per una spiegazione molto accessibile sulla tecnologia del riconoscimento del parlato che usa animazioni, vedi questo filmato prodotto dalla piattaforma digitale educativa Acadaimy e pubblicato su Youtube: <<https://youtu.be/6altVgTO9s>>. Si possono creare sottotitoli in italiano tramite l'interfaccia delle impostazioni di Youtube.

<sup>14</sup> C. Draxler, *Automatic Transcription of Spoken Language Using Publicly Available Web Services*, in J. Saturno, L. Spreafico (a cura di), *Fare linguistica applicata con le digital humanities*, Officinaventuno, Milano 2022.

<sup>15</sup> C. Draxler, *Automatic Transcription...* cit.

<sup>16</sup> S. Scagliola, F. de Jong, *Taking notes about ringing doorbells and barking dogs; the value of context for the reuse of Oral History data*, in N. Apostolopoulos, M. Barricelli, G. Koch (a cura di), *Preserving Sur-*

*vivors Memories: Digital Testimony Collections about Nazi*, EVZ Foundation Remembrance, Responsibility and Future 2018.

<sup>17</sup> Sul rapporto fra la guerra e la produzione di narrative e storia orale vedi anche: A. Casellato, *Le guerre non finiscono mai. Fonti orali, storiografia, culture di guerra*, in P. Del Negro, E. Francia (a cura di), *Guerre e culture di guerra nella storia d'Italia*, Unicopli, Milano 2011, pp. 179-196.

<sup>18</sup> S. Scagliola, F. de Jong, 9.4. *Clio's Talkative Daughter Goes Digital: The Interplay between Technology and Oral Accounts as Historical Data*, in R. Bod, T. Weststeijn, J. Maat (a cura di), *The Making of the Humanities, Volume III: The Modern Humanities*, Amsterdam University Press, Amsterdam 2014, pp. 511-526, <<https://doi.org/10.1515/9789048518449-034>>.

<sup>19</sup> L.R. Rabiner, B.H. Juang, *Automatic Speech Recognition. A Brief History of the Technology Development*, Georgia Institute of Technology, Atlanta; Rutgers University and the University of California, Santa Barbara 2004.

<sup>20</sup> M. Whittaker, *The steep cost of capture*, in «IX Interactions», XXVIII.6, November-December 2021, <[https://mags.acm.org/interactions/november\\_december\\_2021/MobilePagedReplica.action?=-undefined&pm=2&folio=50#pg52](https://mags.acm.org/interactions/november_december_2021/MobilePagedReplica.action?=-undefined&pm=2&folio=50#pg52)>.

<sup>21</sup> L. Fortunato, M. Galassi, M., *The case for free and open source software in research and scholarship*, 2020, <<https://doi.org/10.1098/rsta.2020.0079>>.

<sup>22</sup> C. Draxler, H. Van den Heuvel., A. Van Hessen, S. Calamai, L. Corti, S. Scagliola, *A CLARIN Transcription Portal for Interview Data*, Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC2020), 2020, pp. 3346-3352.

<sup>23</sup> Draxler et al., 2020.

<sup>24</sup> Per ragioni editoriali le immagini sono in bianco/nero. I riferimenti ai diversi colori utilizzati nella T-Chain possono essere verificati provando direttamente sul sito della T-Chain che è accessibile tramite il sito di CLARIN in paesi le cui università fanno parte del consorzio.



Questo volume raccoglie saggi che riguardano la trascrizione delle fonti orali. L'obiettivo è di mettere in comune vantaggi e limiti della trascrizione come tentativo di dire quasi la stessa cosa (secondo un'espressione utilizzata da Umberto Eco) rispetto al parlato, e di rendere nello scritto non solo l'orale, ma anche un orale "non vocale".

Il libro ha l'ambizione di contribuire a colmare un vuoto. La trascrizione del parlato è un passaggio spesso fondamentale, talvolta ineludibile, del lavoro degli storici orali sulle proprie fonti. Eppure, nella vasta letteratura sul metodo della storia orale, e salvo alcune interessanti eccezioni, alla trascrizione è sempre stata dedicata un'attenzione relativamente limitata, sia in Italia sia in ambito internazionale. È un gap che si spiega con l'adesione comune all'assunto implicito che la trascrizione sia una sorta di atto meccanico, un passaggio obbligato e anche un po' noioso, che segue regole piuttosto ovvie. Una più attenta riflessione spinge invece a considerare che di ovvio nella trascrizione c'è poco o nulla. Essa è sempre un atto di natura interpretativa, un nodo che intreccia un piano tecnico, teorico, etico; una ricerca continua, condotta tra molte mediazioni, di un punto di equilibrio tra fedeltà, leggibilità, non dispersione del senso e significato trasmesso a voce. In questa direzione il volume si propone anche come un prezioso contributo didattico.

ISBN 979-12-80675-20-0



9 791280 675200

€ 20,00