

IL PROGETTO BOMARZO

Amedeo De Dominicis e Pamela Mattana¹
Laboratorio di Fonetica, Università degli Studi della Toscana
dedomini@unitus.it, pamela.mattana@gmail.com

1. SOMMARIO

Nel corso degli ultimi 50 anni sono stati raccolti, da vari centri di ricerca (soprattutto orientati allo studio dei dialetti), numerosi materiali che documentano l'uso della lingua da parte di parlanti diversi per provenienza geografica, età, livello di cultura. In non rari casi, le varietà dialettali documentate costituiscono la testimonianza viva di un uso ormai scomparso.

Tuttavia, i supporti materiali su cui tali raccolte sono conservate sono spesso obsoleti e ciò, oltre a renderne difficile la consultazione, mette a rischio la loro stessa conservazione. Sarebbe possibile e soprattutto auspicabile riversare su nuovi supporti, di tipo digitale, questo prezioso giacimento, creando una base di dati facilmente consultabile secondo procedure unitarie dalla comunità degli studiosi.

D'altro canto, sarebbe possibile e altrettanto auspicabile realizzare un archivio centralizzato, organizzato in una piattaforma di facile uso, in cui i punti di raccolta siano indicati sulla mappa, con la possibilità di *clickare* su un punto qualsiasi, vedere l'elenco dei materiali immagazzinati (ovunque possibile, anche sotto forma di trascrizione), e sentire la corrispondente realizzazione sonora.

In tal modo, questi materiali oggi virtualmente inesistenti per il mondo della ricerca, sarebbero nuovamente utilizzabili per l'intera comunità scientifica.

Dalla consapevolezza dell'importanza linguistica e storica di questo patrimonio a rischio, nel novembre 2004 nasce il progetto Calliope (le "soffitte" della voce), uno sviluppo del precedente progetto COVAID (aprile 2002)² (De Dominicis, 2002).

2. DEFINIZIONE DEI MATERIALI DI INTERESSE

I documenti vocali di interesse per i due progetti sono tutte le registrazioni analogiche di rilevanza linguistica, con particolare riguardo a dialetti, comprese le parlate alloglotte d'Italia, le varietà locali dell'italiano (incluse quelle di comunità estere), le varietà di italiano L2, il parlato registrato in manifestazioni folkloriche. In tale lista si comprendono non solo documenti puramente acustici, ma anche di tipo audiovisivo. Nel caso di materiali già digitalizzati, occorre valutare la compatibilità dei criteri adottati, per eventualmente predisporre la ridigitalizzazione a partire dall'originale.

3. PARTECIPANTI

Inizialmente, i partecipanti al progetto COVAID erano il Laboratorio di Fonetica dell'Università della Toscana (Amedeo De Dominicis), il Laboratorio di Linguistica della Scuola Normale Superiore di Pisa (Pier Marco Bertinetto), l'Università di Padova (Alberto Zamboni), la sezione di "Fonetica e Dialettologia" di Padova dell'Istituto di Scienze

¹ I paragrafi 1-4 sono di Amedeo De Dominicis, i paragrafi 5-8 di Pamela Mattana.

² **COVAID** è l'acronimo di **C**onservazione e **V**alorizzazione degli archivi vocali dell'**I**taliano e dei suoi **D**ialetti.

Tecnologie della Cognizione del CNR (I.S.T.C.) (Emanuela Magno Caldognetto) e la Fondazione "Ugo Bordoni" di Roma (Andrea Paoloni). In seguito, si è aggiunto il Laboratorio di Fonetica dell'Università della Calabria (Luciano Romito).

4. OBIETTIVO

L'obiettivo dei due progetti è la creazione di un archivio storico dell'italiano parlato e delle sue varietà, con particolare attenzione alle differenze diacroniche, diastratiche e diafasiche, che sia di facile consultazione da parte degli studiosi. Tuttavia, la mancata attuazione del progetto su scala nazionale ha rallentato la preliminare definizione degli *standard* di lavoro. In primo luogo, si sarebbe dovuto preliminarmente precisare la struttura della base di dati, le modalità di consultazione della stessa e come dinamicamente ampliarla ed aggiornarla. Allo stesso modo, avrebbero dovuto essere definite le modalità di individuazione, sul territorio nazionale, dei materiali fonici a vario titolo raccolti da organismi o persone. Per quanto riguarda il formato di archiviazione audio, lo standard di diffusione dei documenti audio su internet non è affidabile (il formato MP3 a 32 Kbyte sec., ad es., non consente un'analisi spettrografica attendibile) (Paoloni, 2004). Inoltre, sarebbe necessaria una scheda di trascrizione, altrimenti il materiale sarebbe simile a quello oggi udibile, ad es. sul sito delle Teche RAI (Parola, Santoli, 2004).

Inoltre, resta da decidere quali siano le modalità atte a motivare la collaborazione degli istituti e degli studiosi alla collezione dei dati. Infine, la costituzione di un archivio centralizzato creato grazie alle donazioni di un insieme di contributori pone alcune questioni.

Innanzitutto occorre stabilire dove collocare materialmente l'archivio. Probabilmente la soluzione migliore è di affidarlo ad un Ente tecnico *super partes*, come il Ministero delle Comunicazioni o la Discoteca di Stato.

Inoltre, occorrerebbe motivare i detentori (privati o pubblici) di giacimenti vocali a divenire contributori del progetto. A tal fine sarebbe utile che la proprietà dei documenti contenuti nell'archivio rimanesse ai contributori (con tutela del nome del proprietario, dell'istituto di appartenenza e del raccoglitore). A fronte della disponibilità a riversare i documenti sonori nell'archivio centralizzato, il progetto dovrebbe fornire ai contributori i seguenti vantaggi: riversamento su supporto non deperibile del proprio materiale, che resterebbe ovviamente a disposizione dell'ente contributore; documentazione e catalogazione (e, in prospettiva, trascrizione) dell'immenso patrimonio sommerso; in prospettiva, accesso gratuito all'intera banca dati da parte dei contributori, previo consenso - in forme da concordare - dei proprietari dei singoli fondi documentali (si può ipotizzare per es. un ragionevole arco di tempo, trascorso il quale il documento prima riservato all'uso esclusivo del proprietario possa divenire consultabile per tutti i contributori. Ciò permetterebbe di completare eventuali ricerche scientifiche in corso); possibilità di figurare all'interno del catalogo, con evidente guadagno di immagine.

5. COME NASCE IL PROGETTO BOMARZO

Il territorio della Tuscia è soggetto - seppur in forma poco pubblicizzata - ad indagini e raccolte a carattere etnografico. Questo materiale è eterogeneo: in parte composto da manufatti del vivere quotidiano in ambiente montano o rurale (confluiti poi nel museo di Canepina (VT), in parte da foto e video di rappresentazioni sacre o feste patronali, ma anche - e in parte rilevante - da registrazioni su supporto magnetico, come bobine o audiocassette. Molti dei materiali sopra elencati sono oggi nel fondo audio della Provincia

di Viterbo. Il Laboratorio di Fonetica dell'Università della Tuscia ha, quindi, stipulato un accordo con la Provincia di Viterbo al fine di intraprendere un lavoro che garantisca la salvaguardia dei documenti sonori e un minor ingombro fisico del materiale³.

Il Laboratorio di Fonetica ha così potuto effettuare un primo sondaggio sul patrimonio di questo fondo e sulle sue potenzialità. Di comune accordo con i tutori del fondo - che sono inoltre coloro che lo hanno costruito - abbiamo iniziato a trattare le registrazioni su supporto magnetico relative a Bomarzo (VT), il primo centro che è toccato dalle indagini promosse dal Centro di Catalogazione dei beni culturali della Provincia di Viterbo.

Il Progetto Bomarzo rappresenta il primo tentativo di dare vita al progetto Calliope dal quale trae le linee guida e al quale, nel suo piccolo, tenta di dare risposte in merito agli *standard*.

È nata, quindi, una collaborazione fondata sulla convinzione che la tutela dell'identità storica del territorio e la sua valorizzazione siano il fondamento del suo stesso sviluppo culturale.

6. L'ARCHIVIO AUDIO DELLA PROVINCIA DI VT

L'archivio audio della Provincia di Viterbo è costituito da molteplici tipologie di supporti: bobine, audiocassette, cassette video VHS, Compact Disc. Si tratta di un fondo in espansione: i tutori del fondo e suoi raccoglitori continuano nel tempo ad alimentarlo con nuovi materiali. Questi ultimi datano dalla fine degli anni sessanta ad oggi.

Il fondo ha dei registri di inventario e schede documentarie nelle quali figurano non solo l'anagrafico del documento, ma anche le trascrizioni di documentazione (cfr. Figura 1).

Figura 1: Scheda cartacea.

Lo stato di conservazione e documentazione del materiale, all'epoca del progetto, non rispettava gli *standard* di conservazione dei materiali sia cartacei che non cartacei (Canazza, 2003; Carucci, 1993). Il fondo presenta lacune sia dal punto di vista

³ Infatti a conclusione del lavoro la Provincia è stata dotata di una copia del materiale audio riversato su formato dvd in formato 22.05 kHz 16 bit.

documentaristico (descrizione interna alle schede), che archivistico (codici delle schede, numerazioni, divisioni in faldoni e inventario).

7. PROGETTO BOMARZO

Il periodo di elicitazione dei dati relativi a Bomarzo è ristretto in un arco di tempo che va dal 31-12-1978 al 31-12-1979. Gli intervistatori erano principalmente due donne e un uomo originari della zona della Tuscia; lo scopo del loro lavoro è stato quello di raccogliere materiale per una descrizione delle tradizioni popolari nel detto territorio, al fine di tutelare un patrimonio culturale che ha come suo veicolo primario l'oralità.

Per quanto concerne gli informatori censiti, possiamo rilevare che sono un numero piuttosto elevato, ben 61 divisi tra 38 uomini e 23 donne. Per quel che riguarda la loro età, è rilevante il dato che 11 fossero nati negli ultimi anni dell'800 e 50 nei primi decenni dello scorso secolo (Arduini, Leuzzi, Palmisciano 1983).

Il lavoro consta di un totale registrato pari a 1934.769.15 minuti, cioè circa 32 ore, suddiviso in 24 cassette audio di tipo ferro-cromo da 90 minuti, registrate su ambedue i lati, con larghezza di banda di 8000 Hz.

Per quanto concerne le schede anagrafiche corredate di trascrizioni, il Laboratorio di Fonetica dispone di una copia sia cartacea che su formato PDF, mentre gli originali sono rimasti in possesso della Provincia.

Il nostro lavoro si è articolato in cinque fasi.

7.1 Fase prima

La prima fase ha richiesto competenze di tipo archivistico. Infatti, abbiamo rilevato l'intero fondo audio della provincia di Viterbo, effettuando il sondaggio del fondo e la consultazione degli inventari cartacei ad esso relativi. Attraverso la loro messa a confronto siamo giunti alla stima specifica: quantità, stato di conservazione, luogo di conservazione, lacune o materiale disperso, ecc.. È stata la fase più delicata del lavoro, senza la quale non sarebbe possibile giungere alla selezione del materiale che è stato successivamente sottoposto al trattamento di conservazione e documentazione. La scelta della porzione di fondo sulla quale operare è stata effettuata sulla base di parametri di opportunità; con questo termine ci riferiamo allo stato di conservazione del materiale, al suo prestigio e alla sua quantità (Carucci, 1993; Carucci, 1998; Londei, 2004; Zanni Rosiello, 1996).

7.2 Fase seconda

La seconda fase ha richiesto competenze di carattere acustico e di restauro dell'audio. Il materiale in nostro possesso era conservato in audio-cassette del tipo ferro-cromo della durata di 90 minuti. In primo luogo, è stata verificata la qualità del segnale analogico, che in alcune parti risulta saturato, mentre in altre presenta una dinamica tanto insufficiente da comprometterne l'udibilità. Di conseguenza, è stato effettuato il riversamento da analogico a digitale (formato WAV, frequenza di campionamento 44.1 kHz e quantizzazione 16 bit) e successivamente abbiamo sottoposto il segnale così digitalizzato a filtraggio ed a normalizzazione in ampiezza. Facciamo presente che esistono quindi due copie digitali: la copia conforme all'originale e la copia che ha subito il filtraggio e la normalizzazione.

Il materiale raccolto è affetto da una serie di imperfezioni e disturbi acustici, dovuti alla impreparazione dei raccoglitori originari e alle finalità amatoriali che essi si prefiggevano. I segnali sono spesso saturati e i rumori ambientali contenuti nelle registrazioni (passaggio di bus, mezzi agricoli o altri macchinari) li rendono poco intelligibili. Le registrazioni sono state condotte in luoghi conviviali e in occasioni particolari come feste e celebrazioni, quindi il parlato è spesso sovrapposto e in alcuni casi più parlanti si alternano nel corso

della stessa sessione di registrazione. Si tratta di disturbi per i quali non è facile attuare un protocollo di restauro (Cosi, 2001; Canazza, 2004; De Dominicis, 2006; De Mezzo, 2004).

7.3 Fase terza

La terza fase ha carattere specificamente documentale. Le schede originali di documentazione, redatte dai raccoglitori su supporto cartaceo, sono state digitalizzate in formato PDF.

Successivamente, i dati contenuti nelle schede sono stati inseriti in un foglio *Excel*. I campi già presenti sono stati integrati con dati relativi alla durata, alla qualità del segnale, ecc. (cfr. Figura 2). Infine, le schede in formato *Excel* sono state convertite in formato *Access*, al fine di consentirne una più agevole consultazione attraverso *query*.

Figura 2: Scheda *Access*.

7.4 Fase quarta

Nella quarta fase si è sviluppato un prototipo – assai primordiale – di trascrizione linguistica del *corpus*.

Le schede cartacee di documentazione non costituiscono un valido punto di partenza in tal senso, in quanto redatte da mani diverse, con competenze amatoriali e con criteri non scientifici. La trascrizione è estremamente disomogenea: in genere, è ispirata all'ortografia dell'italiano, talvolta con tentativi di rappresentare le forme dialettali con segni diacritici.

Inoltre le schede cartacee forniscono una trascrizione del contenuto parlato della registrazione spesso lacunosa: rappresentano solo il turno di parola dell'intervistato e non dell'intervistatore, descrivono sommariamente l'argomento elencando solo le parole ritenute "interessanti", trascrivono solo il *cammeo* (barzelletta, canzone) e non l'intero svolgersi della conversazione in cui è inserito.

Perciò, il lavoro di trascrizione è stato svolto *ex novo*. Il segnale è stato etichettato attraverso le risorse offerte dal *software* di elaborazione audio *Praat*. Ai segnali audio sono stati associati file *TextGrid* dotati di due *Tiers* di annotazione: uno per la trascrizione ortografica e l'altro per annotazioni generali, ove vengono indicati vuoti di segnale (0), sovrapposizioni di più persone (sov.), risate (ridono), ma anche rumori (R). Ovviamente, è possibile aggiungere altri *Tiers* per segnalare, ad es., il cambio di locutore o semplicemente per distinguere un rumore ambientale da un colpo di tosse o da una sovrapposizione. La scelta di *Praat* è stata dettata proprio per questa possibilità di poter aggiungere livelli di annotazione (cfr. Figura 3).

La trascrizione ortografica è stata effettuata su tutto il materiale ad eccezione di quello cantato.

7.5 Fase quinta

Il Laboratorio di Fonetica ha intrapreso la trascrizione e l'annotazione fonetica del materiale di Bomarzo al fine di compiere un lavoro di descrizione della varietà locale. Ad oggi questo materiale non è mai stato oggetto di studio linguistico, ma da una prima analisi del *corpus* emergerebbero i tratti caratteristici dell'area centro-meridionale (Grassi *et al.*, 2001: 71-160), come il raddoppiamento fonosintattico, la sonorizzazione delle occlusive intervocaliche e il rotacismo della laterale preconsonantica. Certamente queste prime rilevazioni attendono la validazione dai dati del *corpus*.

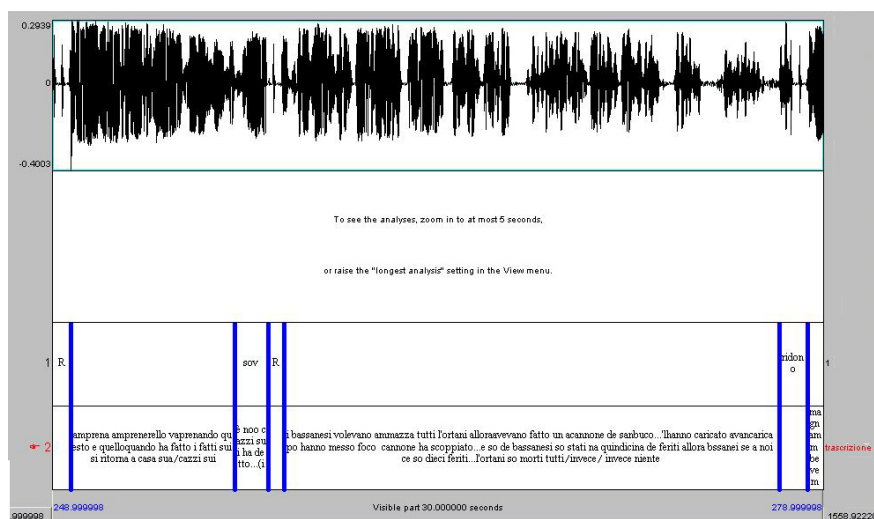


Figura 3: Schermata Praat con trascrizione e annotazioni.

8. COSA RIMANE DA FARE? IL FUTURO

Il progetto Bomarzo, in quanto parte del più ampio progetto Calliope, per svilupparsi ha bisogno di potersi avvalere di *standard* non solo relativi ai criteri di trascrizione fonetica, ma anche alla struttura del data-base e delle gerarchie classificatorie sottostanti, ai livelli di analisi; come pure sarebbe necessario disporre di un *thesaurus* classificatorio omogeneo e stabile che permetta una facile interrogazione dei dati (Weston, 2002).

Per quanto riguarda specificamente l'annotazione relativa alla qualità del segnale e alla sua intelligibilità, sarebbe auspicabile la possibilità di poter etichettare i due fattori con criteri graduali, potendo inoltre segnalare per chi è fruibile un determinato *frame*. Ad esempio, sarebbe utile poter indicare che una data porzione di segnale è una fonte valida per la ricostruzione lessicale oppure per la rilevazione della F0.

Ovviamente, quanto detto finora è attuabile, nelle migliori prospettive, solo nella lunga durata e con largo investimento di risorse; infatti, si richiede la digitalizzazione di tutto il patrimonio che allo stato attuale non è stato ancora interamente censito, la sua schedatura, trascrizione, inserimento in una banca dati ed etichettatura in senso documentario-multimediale (*tag e link*) (Calvo, Roncaglia, 2000).

Si sta lavorando inoltre al fine di una messa in rete del materiale.

9. BIBLIOGRAFIA

- Arduini M, Leuzzi M. D., Palmisciano M.G. (1983), *Tradizioni orali a Bomarzo alcuni repertori di una ricerca*, Viterbo: Amministrazione Provinciale di Viterbo.
- Calvo M., Roncaglia G. (2000), *Internet 2000: manuale per l'uso della rete*, Bari: Laterza.
- Canazza S. (2003), <http://www.dei.unipd.it/~musica/Dispense/Restauro.pdf>
- Canazza S. (2004), Il restauro del segnale vocale mediante proiezione locale, in: De Dominicis A., Mori L., Stefani M. (editors) (2004), 39-52.
- Carucci P. (1993), *Le fonti archivistiche: ordinamento e conservazione*, Roma: La Nuova Italia Scientifica.
- Carucci P. (1998), *Manuale di archivistica per l'impresa*, Roma: Carocci.
- Così P. (2001), Il restauro digitale del segnale audio, in: De Dominicis A. (ed.) (2001), 189-199.
- De Dominicis A. (2001), La voce come bene culturale, la linguistica come strumento di catalogazione e restauro, in: De Dominicis A. (ed.) (2001), 7-26.
- De Dominicis A. (2002), Co.Va.I.D. (COnservazione e VAlorizzazione degli archivi vocali dell'Italiano e dei suoi Dialetti), *La comunicazione*, numero unico speciale a cura di Giuseppe Rinaldo e Roberto Piraino contenente gli Atti della Conferenza *TIPi (Tecnologie Informatiche nella Promozione della lingua Italiana)*, LI, 97-98.
- De Dominicis A. (2006), Il restauro "perfetto". In Canazza, Sergio & Casadei Turrone Monti, Mauro (eds.) *Ri-mediazione dei documenti sonori*, Udine: Forum, 603-608.
- De Dominicis A. (ed.) (2001), *La voce come bene culturale*, Roma: Carocci.
- De Dominicis A., Mori L., Stefani M. (editors) (2004), Costituzione, gestione e restauro di corpora vocali, Atti delle XIV giornate di studio del GFS (A.I.A.), Roma: Esagrafica.
- De Mezzo G. (2004), Conservazione dei documenti sonori: il riversamento conservativo di registrazioni su nastro magnetico, in: De Dominicis A., Mori L., Stefani M. (editors) (2004), 27-38.
- Grassi C., Sobrero A.A., Telmon T. (2001), *Fondamenti di dialettologia italiana*, Roma-Bari: Laterza.
- Londei L. (2004), *Elementi di archivistica*, Roma: Jouvence.
- Paoloni A. (2004), Restauro e memorizzazione del segnale vocale, in: De Dominicis A., Mori L., Stefani M. (editors) (2004), 53-66.
- Parola L., Santoli P. (2001), Il catalogo multimediale della RAI. Le Teche radiofoniche, in: De Dominicis A. (ed.) (2001), 201-208.
- Weston P. G. (2002), *Il catalogo elettronico: dalla biblioteca cartacea alla biblioteca digitale*, Roma: Carocci.
- Zanni Rosiello I. (1996), *Andare in archivio*, Bologna: il Mulino.