ISSN: 2612-226X

Studi AISV 6

GLI ARCHIVI SONORI AL CROCEVIA TRA SCIENZE FONETICHE, INFORMATICA UMANISTICA E PATRIMONIO DIGITALE

AUDIO ARCHIVES AT THE CROSSROADS OF SPEECH SCIENCES, DIGITAL HUMANITIES AND DIGITAL HERITAGE

a cura di
Duccio Piccardi, Fabio Ardolino, Silvia Calamai

AISV
Associazione Italiana Scienze della Voce

GLI ARCHIVI SONORI AL CROCEVIA TRA SCIENZE FONETICHE, INFORMATICA UMANISTICA E PATRIMONIO DIGITALE

AUDIO ARCHIVES AT THE CROSSROADS OF SPEECH SCIENCES, DIGITAL HUMANITIES AND DIGITAL HERITAGE

a cura di
Duccio Piccardi, Fabio Ardolino, Silvia Calamai

Milano 2019

studi AISV - Collana peer reviewed

Studi AISV è una collana di volumi collettanei e monografie dedicati alla dimensione sonora del linguaggio e alle diverse interfacce con le altre componenti della grammatica e col discorso. La collana, programmaticamente interdisciplinare, è aperta a molteplici punti di vista e argomenti sul linguaggio: dall'attenzione per la struttura sonora alla variazione sociofonetica e al mutamento storico, dai disturbi della parola alle basi cognitive e neurobiologiche delle rappresentazioni fonologiche, fino alle applicazioni tecnologiche.

I testi sono sottoposti a processi di revisione anonima fra pari che ne assicurano la conformità ai più alti livelli qualitativi del settore.

I volumi sono pubblicati nel sito dell'Associazione Italiana di Scienze della Voce con accesso libero a tutti gli interessati.

c/o LUSI Lab - Dip. di Scienze Fisiche
Complesso Universitario di Monte S. Angelo
via Cynthia snc
80135 Napoli
email: presidente@aisv.it
sito: www.aisv.it

Edizione realizzata da
Officinaventuno
Via F.lli Bazzaro, 18
20128 Milano - Italy
email: info@officinaventuno.com
sito: www.officinaventuno.com

ISBN edizione digitale: 978-88-97657-38-5
ISSN: 2612-226X

Sommario

DOI: 10.17469/O2106AISV000000

DUCCIO PICCARDI, FABIO ARDOLINO, SILVIA CALAMAI

Prefazione

Questo volume nasce ad un anno dalla chiusura dei lavori del XV Convegno Nazionale AISV, organizzato presso il Dipartimento di Scienze della Formazione, Scienze umane e della Comunicazione interculturale dell'Università degli Studi di Siena (14-16 febbraio 2019).

Gli archivi sonori hanno rappresentato il fulcro tematico intorno al quale numerosissimi studiosi, provenienti da ambiti anche molto diversi, si sono riuniti nel Campus universitario del Pionta ad Arezzo; al contempo – come è tradizione nei Convegni AISV – nel corso delle sessioni hanno trovato spazio diversi temi relativi allo studio e alle tecnologie del parlato, in funzione delle proposte di comunicazione provenienti dai Soci. Con la partecipazione di oltre novanta studiosi provenienti da sette Paesi differenti, il convegno AISV 2019 ha costituito l'avvio di un'intensa fase di approfondimento e confronto tra esperti con orientamenti teorici e/o metodologici spesso differenti. Di tutto ciò questo volume vuole essere diretta, anche se parziale, testimonianza.

Come ben evidenzia il titolo del Convegno ("Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale"), gli archivi sonori rappresentano artefatti comuni a molti campi delle scienze umanistiche e sociali, dalle discipline linguistiche a quelle storiche, sociologiche, antropologiche e psicologiche; essi costituiscono inoltre strumenti di sempre più ampio utilizzo nello sviluppo delle tecnologie della voce (elaborazione del linguaggio naturale, riconoscimento automatico del parlato). Nonostante l'indiscutibile valore documentario che tali risorse costituiscono per la ricerca umanistica e sociale, il loro utilizzo appare ancora piuttosto limitato. Fra le molte ragioni alla base di questa mancata valorizzazione, due appaiono forse le più stringenti: da un lato, la frammentazione degli archivi, dei depositi istituzionali e dei centri di elaborazione, frammentazione che mina l'accessibilità e la riusabilità delle risorse orali; dall'altro, la legislazione in merito alla riservatezza e alla tutela dei dati vocali, essendo la voce un dato biometrico. In questo quadro, il convegno ha indagato la possibilità di una più stretta collaborazione tra studiosi della voce, esperti di analisi conversazionale e di corpora di parlato, ingegneri del suono e storici orali.

Le ricadute positive di queste prove di dialogo sono diverse e concrete. La creazione di una più stretta collaborazione con l'Associazione Italiana di Storia Orale (AISO) costituisce un primo motivo di soddisfazione. L'Associazione ha sostenuto, anche attraverso alcune borse di studio, la partecipazione di giovani studiosi in formazione all'evento ed è stata una delle parti attivamente coinvolte – nella persona del suo Presidente, Alessandro Casellato – nella Tavola Rotonda tenutasi nei giorni del Convegno. Inoltre, proprio la Tavola Rotonda ha permesso, lungo tutto il 2019, l'avvio di una serie di tavoli operativi, con Istituzioni, Istituti Centrali, Enti ed Associazioni avente come obiettivo la creazione di un Vademecum sugli ambiti della conservazione, della catalogazione, dell'uso e riuso degli archivi sonori. Naturalmente, l'AISV ha preso parte con costanza a siffatti tavoli di lavoro, portando in essi le istanze della comunità che rappresenta.

Infine, il Convegno ha costituito l'opportunità, per la comunità di AISV, di conoscere più da vicino la Rete di Ricerca Europea CLARIN (*Common Language Resources and Technology Infrastructure*) attraverso due momenti fondamentali: la conferenza in plenaria dal titolo *Spoken Word Archives as Societal and Cultural Data* tenuta dal direttore esecutivo dell'Infrastruttura, Franciska de Jong, che ha illustrato al pubblico le potenzialità della ricerca sugli archivi orali all'interno di CLARIN, e il tutorial a numero chiuso, indirizzato ai giovani studiosi in formazione e finanziariamente supportato da CLARIN (nel programma *User Involvement*), dal titolo *Creating, Managing and Analysing Speech Databases using BAS Services and Emu: A Hands-On Tutorial*, a cura di Christoph Draxler e Florian Schiel (Bavarian Archive for Speech Signals, Ludwig Maximilian University Munich). Giovani linguisti, oralisti, tecnologi hanno potuto riflettere insieme in merito alla creazione e al trattamento di archivi di parlato (si veda il resoconto sottoforma di blog nel sito di CLARIN: https://www.clarin.eu/blog/clarin-tutorial-aisv-2019-arezzo).

Il volume, articolato in tre parti, include una selezione dei saggi presentati al Convegno e sottoposti a revisione fra pari. La prima delle tre sezioni contiene i contributi relativi alla sessione tematica sugli archivi orali.

Valentina Colonna e Antonio Romano (*VIP: un archivio per le voci della poesia italiana*) presentano *Voices of Italian Poets* (VIP), un archivio sonoro online di letture di poesia italiana del presente e degli ultimi quarant'anni. Lo strumento consta di 600 registrazioni di letture e si rivolge a una comunità scientifica di ambito fonetico sperimentale. Il contributo offre una panoramica sulla letteratura relativa a questo specifico interesse e su simili strumenti a livello internazionale. I primi risultati di studi basati su VIP sono infine riassunti per esporre le potenzialità dello strumento.

Carlo Geraci, Roland Pfau, Pietro Braione, Carlo Cecchetto e Josep Quer (*Hidden languages in a digital world: the case of sign language archives*) descrivono SIGN-HUB, un progetto europeo dedicato alle lingue dei non udenti. Il progetto ha quattro finalità: la compilazione di grammatiche, di un atlante linguistico, lo sviluppo di procedure diagnostiche per disturbi del linguaggio e la pubblicazione di un archivio online contenente narrazioni di non udenti. Quest'ultimo aspetto è descritto in dettaglio, con riferimenti tecnici sulla struttura della piattaforma e sui criteri di svolgimento delle 137 interviste. L'impatto storico, antropologico e sociologico dell'operazione è ampiamente messo in evidenza nel contributo.

Il contributo di Antonio Romano e Valentina De Iacovo (*La base di dati "Tramontane": dati di parlato su lingue, dialetti, etnoletti e interletti del laboratorio di fonetica sperimentale "Arturo Genre"*) descrive l'origine, gli sviluppi e le potenzialità di un vasto archivio di parlato costruito presso il Laboratorio di Fonetica Sperimentale "Arturo Genre" dell'Università di Torino e contenente dati etichettati per 528 fra lingue e varietà dialettali.

Manuela Soldi (*L'archivio sonoro di Festivaletteratura di Mantova*) riporta l'esperienza del *Festivaletteratura* di Mantova – rassegna culturale giunta alla sua ventitreesima edizione – nel raccogliere, preparare e redistribuire al pubblico l'imponente archivio sonoro costituito dalle voci e dai contributi susseguitisi nel corso delle diverse edizioni del Festival a partire dal 2009.

Rosanna Sornicola, Giovanni Abete, Elisa D'Argenio e Cesarina Vecchia (*Raccontare un archivio di fonti orali: il progetto "Voci, parole e testi della Campania"*) illustrano le strategie impiegate per garantire ai non specialisti (e in primo luogo alle stesse comunità informanti) l'accesso ai dati sociali, etnici e linguistici raccolti nel corso del progetto *Voci, parole e testi della Campania. Un archivio sociolinguistico ed etnografico digitale per la promozione culturale del territorio* sviluppato dal Dipartimento di Studi Umanistici dell'Università di Napoli "Federico II".

La riflessione proposta da Maria Francesca Stamuli (*Fonti orali, documenti e archivi: riflessioni e proposte per la nascita di un 'archivio vivo'*) offre un inquadramento delle nozioni teoriche e terminologiche correlate alle fonti orali, approfondendo limiti e potenzialità degli assunti fondamentali ad esse sottesi. A tale riflessione si accompagna la discussione delle principali metodologie di trattamento archivistico per tali fonti, ed il loro valore in termini di usabilità ed accessibilità del dato.

Il contributo di Daniela-Carmen Stoica e Anyla Saraçi (*Oral Histories and the Dialectal Research in Albania*) illustra come sessioni di parlato raccolte attraverso i principi metodologici della storia orale possano fornire una preziosa base di dati per lo studio di due parlate con status vulnerabile nell'Albania contemporanea (toskërisht e arumeno).

Graziano Tisato (*Acquisizione dell'intero AIS (Sprach- und Sachatlas Italiens und der Südschweiz)*) espone i risultati del progetto NavigAIS, avente come scopo l'acquisizione dell'intero Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale. Il progetto decennale (2009-2019) ha attraversato tre fasi: la produzione di una versione digitale ed esplorabile ad alta risoluzione dell'intero Atlante, la sua implementazione in rete e infine la progettazione di uno strumento OCR specifico per la rapida acquisizione del testo dalle tavole digitalizzate.

Il lavoro di Petra Vangelisti, Caterina Pesce, Marica Setaro, Greta Bianchini, Lucilla Gigli e Silvia Calamai (*Ritrovare voci. Il lavoro intorno all'archivio sonoro di Anna Maria Bruzzone*), traccia un primo bilancio dell'analisi linguistica e storico-sociale condotta sull'archivio di Anna Maria Bruzzone, donato all'Università di Siena nel 2017. Il materiale archivistico – costituito da quaderni manoscritti e dattiloscritti e registrazioni audio – rappresenta un'eccezionale testimonianza storica degli ultimi anni di attività dell'ospedale psichiatrico di Arezzo.

Chiudono la prima sezione due interventi scritti per la Tavola Rotonda, cui hanno partecipato (in ordine alfabetico) Simonetta Buttò, direttore dell'Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU); Alessandro Casellato, Presidente dell'Associazione Italiana di Storia Orale; Laura Gavioli, coordinatore del Centro Interuniversitario di Analisi dell'Interazione e della Mediazione (AIM); Barbara Gili Fivela, Presidente dell'Associazione Italiana di Scienze della Voce; Sabina Magrini, direttore dell'Istituto Centrale per i Beni Sonori ed Audiovisivi; Piero Cavallari, storico contemporaneo presso il medesimo Istituto; Monica Monachini, coordinatore Nazionale CLARIN-IT; Maria Palmerini, settore Ricerca e Sviluppo di *Cedat85*, società che si occupa di servizi e tecnologie per il trattamento del parlato.

L'idea di una Tavola Rotonda è nata dall'osservazione di una distonia: mentre si registra nel panorama internazionale un interesse sempre più forte nei confronti dei dati (a livello mondiale, attraverso la *Research Data Alliance*; a livello europeo, nell'ambito di infrastrutture di ricerca come CLARIN, DARIAH, CESSDA…), nel nostro Paese l'attenzione verso il patrimonio intangibile rappresentato dalla voce (in tutte le sue declinazioni) non appare accompagnato da un sostegno altrettanto robusto per ciò che riguarda la sua conservazione e la sua tutela. Al contempo, diventa sempre più pressante, da parte di istituzioni, ricercatori, appassionati, la richiesta di avere procedure condivise, strategie comuni e filiere di lavoro standardizzate che permettano a chi raccoglie sul campo materiale audio-visivo di gestire complessivamente il flusso relativo ai dati (dalla raccolta sul campo alla metadatazione, dalla conservazione a lungo termine al trattamento delle questioni etiche e legali che regolano il rapporto tra raccoglitore e fonte, sino al riuso futuro dei dati raccolti). Si registra, più in generale, la necessità di offrire risposte chiare e scientificamente fondate ai ricercatori in formazione e ai cittadini che vogliono rendere disponibile il loro materiale sonoro per ulteriori ricerche e ulteriori utilizzi, anche extra-accademici. Le domande che in quella sede sono state poste agli invitati si possono riassumere nella maniera seguente:

a. Come favorire la conservazione, la descrizione, il riutilizzo di risorse orali raccolte in ambito di progetti di ricerca passati e presenti?
b. Quale metadatazione usare? È possibile proporre un elenco minimo di metadati essenziali?
c. È possibile creare una rete tra associazioni, istituzioni, enti di ricerca che si prenda cura di questo problema, trasversale a molte e diverse discipline?
d. In che modo è possibile rispondere ai molti ricercatori che chiedono cosa fare, a chi affidare, dove depositare i propri archivi sonori raccolti nel corso di anni e talvolta decenni?
e. Esistono in Italia delle istituzioni attrezzate per accogliere questi archivi? È preferibile puntare su un archivio unico nazionale o su una rete di archivi locali o tematici? In questo secondo caso, come far sì che questi possano lavorare insieme, in maniera coordinata, seguendo procedure condivise, condividendo i rispettivi cataloghi?
f. Quali indicazioni di minima è possibile dare a queste istituzioni che in gran parte già esistono, spesso attive su base locale?

L'obiettivo di lunga durata che si è posto il gruppo dei presenti, e che si è concretizzato in una serie di incontri nei mesi successivi (di cui ha dato conto anche *Il*

Mondo degli archivi alla pagina http://www.ilmondodegliarchivi.org/rubriche/in-italia/747-verso-un-vademecum-per-le-fonti-orali) è il seguente: creare delle *linee guida* per il giovane ricercatore in formazione e per i conservatori di archivi orali che possano offrire informazioni su:

i. Metadatazione (gli standard esistenti, con un elenco di metadati essenziali);
ii. Sedi di conservazione (chi a lungo termine si prende cura dei dati e dei metadati: quali procedure attivare, quali i 'requisiti minimi');
iii. Modalità di accesso e di riuso dei dati stessi;
iv. Questioni etiche e legali implicate nel rapporto con la fonte orale, anche nell'ottica del Regolamento generale sulla protezione dei dati (reg. n. 2016/679 - DGPR).

La seconda parte del volume raccoglie i contributi derivanti dalla sessione generale sulle scienze del parlato.

Fabio Ardolino (*Adapting the Implicit Association Protocol to speech: Issues and methodological proposals*) apre la sezione con un contributo teorico sull'*Implicit Association Test* (IAT). Nata nell'ambito della psicologia sociale, questa procedura è di interesse in ambito linguistico per la sua capacità di verificare ipotesi specifiche sugli atteggiamenti impliciti. L'autore sintetizza la struttura della procedura sperimentale trattata; infine, commentando soluzioni operative messe in atto in due ricerche in corso, Ardolino discute tre aspetti critici relativi alla progettazione di protocolli IAT, ovvero la durata, il contenuto e la familiarità degli stimoli.

Lo studio di Violetta Cataldo, Loredana Schettino, Renata Savy, Isabella Poggi, Antonio Origlia, Alessandro Ansani, Isora Sessa e Alessandra Chiera (*Phonetic and functional features of pauses, and concurrent gestures, in tourist guides' speech*), descrive, nell'ambito del progetto CHROME, le fasi di progettazione e di verifica di una metodologia per la raccolta e l'analisi di dati relativi alle pause e ai gesti concomitanti in tre guide turistiche, allo scopo di migliorare lo sviluppo del design di guide virtuali. Lo studio evidenzia una cesura fra pause silenti (grammaticalmente connotate ed intenzionali) e pause vocalizzate (espressione di fenomeni esitativi). Tale distinzione trova conferma nell'analisi dei gesti concomitanti, semanticamente carichi solo quando associati al primo tipo di pausa.

Claudia Roberta Combei e Giovanna Marotta (*Where do accents come from? Factors affecting the degree of foreign-accented Italian*) espongono i risultati di uno studio sperimentale sulla percezione dell'accento straniero di parlato italiano da parte di partecipanti madrelingua. Registrazioni di parlato italiano letto e spontaneo da parte di madrelingua inglesi, francesi, spagnoli, tedeschi, romeni e russi sono state sottoposte a 288 ascoltatori, che avevano il compito di valutare l'intensità dell'accento straniero percepito. I risultati ottenuti tramite regressioni lineari mostrano effetti significativi sia per le variabili riguardanti il parlante, sia per quelle relative all'ascoltatore.

Piero Cossu (*The metathesis in the dialect of Suelli (CA): A descriptive analysis of the rhotic consonant*) riporta i risultati di un'analisi fonologica dei nessi consonantici metatetici in sardo campidanese. Tramite specifici test condotti su un corpus di frasi tradotte nella varietà locale da parlanti di Suelli (intrusioni vocaliche, lenizione in sandhi esterno), l'autore interpreta i nessi come attacchi sillabici complessi. I materiali sono anche oggetto di analisi acustica, che evidenzia l'estrema variabilità allofonica delle rotiche nel dialetto in questione.

Lorenzo Filipponio, Davide Garassino e Dalila Dipino (*Between phonology and typology. Consonant duration in two Gallo-Italian dialects*) comparano dati acustici provenienti dall'area bolognese e da Porto Maurizio nella Liguria occidentale. A livello segmentale, gli autori saggiano le durate delle consonanti postoniche e delle vocali che le precedono al fine di esplorare il ruolo di questi due fattori nella realizzazione di opposizioni fonologiche di lunghezza. I due dialetti si comportano similmente: a una durata vocalica fonologica si accompagna una maggiore lunghezza consonantica dopo vocale breve. Ciononostante, sostanziali differenze emerse da un'analisi qualitativa di parametri d'intensità vocalica inducono gli autori a considerare le strutture segmentali dei due dialetti come il risultato di diverse fenomenologie ritmiche.

Manuela Frontera, Andrea Tarasi e Elvira Graziano (*Le consonanti occlusive sorde aspirate in Calabria: un confronto tra aree dialettali*) propongono una prima mappatura del fenomeno dell'aspirazione delle occlusive sorde in posizione forte in tre aree dialettali della Calabria meridionale. Le medie delle durate denotano che il tratto è ben presente in tutto il territorio considerato. Tuttavia, gli autori evidenziano differenze interareali riguardo al rapporto tra le durate del rilascio e la posizione accentuale. Vengono prese in considerazione anche macrovariabili sociali (sesso, età), evidenziando tendenze distribuzionali che saranno oggetto di future analisi.

Riccardo Orrico, Renata Savy e Mariapaola D'Imperio (*Salerno Italian: intonational phonology and dimensions of variation*) fanno uso degli strumenti del modello autosegmentale per proporre una descrizione dei profili intonativi nell'italiano di Salerno. I dati, provenienti da due tipologie di compito, concernono affermazioni e domande suddivise in un totale di sei sottotipologie. La prospettiva dello studio è in parte comparativa con dati provenienti da altre zone d'Italia, in parte interpretativa di fenomeni interni all'italiano di Salerno. In particolare, viene rilevata una grande variabilità interindividuale nella selezione degli schemi interrogativi, sia in generale che in relazione a particolari condizioni pragmatiche. Gli autori concludono a favore dell'integrazione di parametri sociofonetici e cognitivi negli studi sull'intonazione per meglio rendere conto dei criteri di variazione.

Nel contributo di Duccio Piccardi (*Bilingual Frequency in a Favorable Context (BFFC) in the Italian dialectal area. Theoretical preliminaries to the analysis of geminate lateral retroflexion and voiceless plosives aspiration in Antona (MS)*) viene delineato un quadro teorico utile all'applicazione della metodologia per il calcolo della *Bilingual Frequency in Favorable Context* in contesti bilingue italiano/dialetto. L'applicazione della formula, sviluppata al fine di quantificare il peso che la frequenza di parole imparentate invarianti esercita nei fenomeni variazionali in comunità bilingui, si scontra nel contesto italofono con l'assenza di corpora di frequenza per quanto riguarda le varietà dialettali. Il lavoro specula sulla possibilità di superare tali lacune e testa l'efficacia del BFFC nello spazio bilingue di Antona (MS). Il lavoro si conclude con la discussione dei principali tratti fonetici sui quali estendere l'indagine, aprendo la strada ad applicazioni pratiche del principio metodologico proposto.

Lo studio condotto da Daniele Santini, Giovanna Lenoci, Giovanna Marotta e Pier Marco Bertinetto (*Repeated vowels in Italian: An exploratory study*) esamina i correlati articolatori della produzione di vocali identiche consecutive in Italiano. Sebbene a livello strutturale tale sequenza dia origine ad uno iato, nella pratica essa si risolve in una serie di realizzazioni divergenti, talvolta caratterizzate da forme di riarticolazione fra i due componenti vocalici. La fenomenologia sottesa a tali realizzazioni è discussa nei suoi esiti principali, anche con il ricorso ad analisi UTI (*Ultrasound Tongue Imaging*) durante la produzione di parlato controllato.

Il lavoro di Martina Turconi e Giovanna Marotta (*Speech and music rhythm: A pilot study on English and Italian songs*) esplora le relazioni ritmiche fra composizione musicale e madrelingua del compositore. Lo studio, in particolare, si propone di esaminare le relazioni che si producono fra il ritmo delle diverse madrelingue di compositore ed interprete, e di come la seconda influenzi il ritmo musicale ed il cantato all'atto dell'esecuzione. La duplice analisi di canzoni italiane e inglesi, e delle rispettive rese in lingua straniera, suggerisce che il ritmo isoaccentuale eserciti una maggiore influenza di quello isosillabico, condizionando la resa musicale italiana di canzoni in inglese. Viceversa, entrambi i ritmi sembrano influire ugualmente sulla resa ritmica del cantato.

Claudio Zmarich, Erica Bonato, Roberto Bovo, Vincenzo Galatà e Alessandro Pinton (*A new phonological discrimination test for children aged 48-72 months*) presentano alcuni risultati di un vasto progetto finalizzato alla creazione di un test di discriminazione fonologica per soggetti in età prescolare, utile alla diagnosi precoce del disturbo primario del linguaggio. Il test sviluppato nello studio valuta la capacità di discriminare coppie di non-parole (uguali/diverse), ricomprendendo tutti i processi fonologici potenzialmente presenti nella produzione di un bambino in età prescolare.

Il contributo di Enrico Zovato, Vito Quinci e Paolo Mairano (*Modelling sentiment analysis scores and acoustic features of emotional speech with neural networks: A pilot study*) analizza l'accuratezza di una rete neurale nella predizione del sentimento sulla base dei dettagli acustici, stabilendo la non scindibilità delle componenti acustiche e lessicali nell'espressione dell'emozione nel campione in analisi (audiolibri).

La terza sezione del volume è dedicata alle tecnologie del parlato.

Il contributo di Francesco Cangemi, Jessica Fründt, Harriet Hanekamp e Martine Grice (*A semi-automatic workflow for orthographic transcription and syllabic segmentation*) presenta, seguendo un'intelaiatura sperimentale, una *pipeline* semiautomatizzata per l'annotazione di file audio, con particolare riferimento al parlato spontaneo. La *pipeline* rifunzionalizza la generazione automatica di sottotitoli di Youtube per ottenere un primo livello di trascrizione, che viene poi allineato al file audio usando gli strumenti della piattaforma WebMAUS. Gli autori comparano il prodotto finale a una annotazione manuale degli stessi file, evidenziando un notevole abbreviamento dei tempi senza un peggioramento della qualità.

Maria Di Maro, Giusy Gaudino e Francesco Cutugno (*Conversing on Artistic Representation Topics (CART): An Android audio guide with dialogic skills*) descrivono l'architettura di CART, un'audioguida concepita per migliorare l'esperienza dei visitatori della Certosa napoletana di San Martino. Lo strumento, sviluppato tramite Android Studio, integra la trasmissione di contenuti culturali ad un sistema dialogico multimodale, in grado di interagire con i visitatori. I feedback generalmente positivi ottenuti dalla messa in opera del dispositivo costituiscono un risultato promettente per l'integrazione di simili tecnologie in altri circuiti museali.

Philippe Martin (*Prosodic annotation of oral archives*) espone le potenzialità del software WinPitch nell'analisi prosodica di archivi orali. Il software, offrendo una scelta tra sette algoritmi e la possibilità di annotazione supervisionata, è particolarmente adatto per la sua flessibilità in caso di materiali audio datati e di bassa qualità. Integrando anche strumenti per la segmentazione automatica e l'estrazione dei dati, WinPitch si propone come *pipeline* completa per gli studi prosodici. Il contributo è corredato di esempi presi da registrazioni d'epoca.

Alessandro Russo, Niccolò Pretto, Antonio Rodà e Sergio Canazza (*L'informatica per la gestione e la conservazione di informazioni acustiche (musica e voce)*) presentano le ricerche del Centro di Sonologia Computazionale (CSC) dell'Università di Padova, incentrate sullo sviluppo di metodologie per il restauro, la conservazione e la riproduzione di archivi audio. Grande attenzione viene dedicata agli aspetti contestuali dei supporti fisici, che possono contenere importanti porzioni di metadati. Il contributo descrive inoltre alcune delle innovazioni in sviluppo presso il CSC, quali sistemi basati su *neural network* e *computer vision* per l'individuazione di punti di interesse su nastri, tecniche di *machine learning* per il riconoscimento automatico delle curve di equalizzazione, algoritmi per il restauro audio e interfacce scheumorfiche che simulano le funzionalità di strumentazioni analogiche per la riproduzione di materiali datati.

I curatori ringraziano sentitamente i membri del Comitato Scientifico del XV Convegno AISV e tutti gli studiosi che hanno agito come revisori dei contributi:

Cinzia Avesani (ISTC-CNR, Padova), Pier Marco Bertinetto (Scuola Normale Superiore), Sergio Canazza (Università degli Studi di Padova), Francesco Cangemi (Universität zu Köln), Alessandro Casellato (Università Ca' Foscari Venezia), Chiara Celata (Scuola Normale Superiore), Francesco Cutugno (Università degli Studi di Napoli "Federico II"), Mariapaola D'Imperio (Rutgers University), Amedeo De Dominicis (Università della Tuscia), Anna De Meo (Università degli Studi di Napoli L'Orientale), Christopher Draxler (LMU München), Lorenzo Filipponio (Humboldt Universität zu Berlin), Francesca Frontini (Université Paul-Valéry, Montpellier 3), Vincenzo Galatà (ISTC-CNR, Padova), Barbara Gili Fivela (Università del Salento), Véronique Ginouvès (Phonothèque MMSH, Aix-

en-Provence), Mirko Grimaldi (Università del Salento), Henk van den Heuvel (Radboud University, Nijmegen), Anne Karpf (London Metropolitan University), Michele Loporcaro (Universität Zürich), Giovanna Marotta (Università di Pisa), Pietro Maturi (Università degli Studi di Napoli "Federico II"), Daniela Mereu (Libera Università di Bolzano), Laura Mori (Università degli Studi Internazionali, Roma), Rosalba Nodari (Università degli Studi di Siena), Antonio Origlia (Università degli Studi di Napoli "Federico II"), Bianca Pastori (Associazione Italia di Storia Orale), Elisa Pellegrino (Universität Zürich), Antonio Rodà (Università degli Studi di Padova), Antonio Romano (Università di Torino), Alessandro Russo (Università degli Studi di Padova), Luciano Romito (Università della Calabria), Niccolò Pretto (Università degli Studi di Padova), Alessandro Russo (Università degli Studi di Padova), Florian Schiel (LMU München), Giancarlo Schirru (Università degli Studi di Napoli L'Orientale), Stephan Schmid (Universität Zürich), Francesca Socrate (Università di Roma "La Sapienza"), Patrizia Sorianello (Università di Bari), Ottavia Tordini (Università di Pisa), Alessandro Vietti (Libera Università di Bolzano), Giulia Zitelli Conti (Associazione Italia di Storia Orale), Claudio Zmarich (ISTC-CNR, Padova).

I curatori ringraziano infine Letizia Cirillo e Caterina Pesce, per il loro preziosissimo contributo alla macchina organizzativa del Convegno, gli studenti dell'Università di Siena Lorenzo Russo e Martina Vilucchi, che hanno svolto il loro tirocinio curriculare a supporto del convegno, Vincenzo Galatà per la cura del sito internet dedicato, Antonio Origlia e Maria Di Maro per le attività della segreteria dell'Associazione.

Con il XV Convegno AISV si è chiuso il mandato di Presidente di Franco Cutugno. A lui il più sentito ringraziamento e alla nuova Presidente gli auguri di buon lavoro. Alla memoria di Renata Savy, che tanto ha contribuito all'Associazione, è dedicato questo volume.

PARTE I

ARCHIVI ORALI
(SESSIONE TEMATICA)

DOI: 10.17469/O2106AISV000001

VALENTINA COLONNA, ANTONIO ROMANO

VIP: un archivio per le voci della poesia italiana

Reading is one of the most interesting aspects to observe the sound dimension of poetry. For this purpose, the support of poetry vocal archives is fundamental. After a review of the major international databases, together with the most significant related studies, this article aims to present *Voices of Italian Poets*, a research project dedicated to the study about the voice in Italian poetry reading and an Italian national sound archive, available on the website of the Laboratory of Experimental Phonetics "Arturo Genre" of the University of Turin. VIP constitutes a platform of interpretations by Italian living and Twentieth century's poets, together with actors, accompanied by phonetic annotations, analysis and comparisons. In addition to the possible methods of consultation, in this work we will present the perspectives that this research proposes and opens.

Key words: phonetics, poetry, poetry vocal archive, prosody, Voices of Italian Poets.

1. *Introduzione*

Per affrontare la questione cruciale della musica della poesia contemporanea, ovvero approfondire l'ossatura sonora che le dà forma e la rende tale, la lettura ad alta voce della stessa può costituire, al giorno d'oggi, un valido supporto. Lo sviluppo difatti delle tecniche di registrazione, unitamente alla fitta diffusione di manifestazioni artistiche dedicate alla poesia e alla sua lettura in pubblico, insieme anche al perfezionamento della strumentazione e della metodologia di studio della Linguistica e, in particolare, della Fonetica, rappresentano elementi consistenti per avvalorare la necessità di intraprendere ricerche che esplorino queste realtà, seguendo così strade ancora poco battute. Al fine dunque di intraprendere lavori di questo tipo e al fine anche di poter conservare e usufruire di preziosi materiali sonori, la creazione di archivi vocali che raccolgano registrazioni di letture originali di poeti e altre eventuali prospettive di lettura rappresenta un elemento cruciale, che ha avuto modo di svilupparsi nel tempo, facilitando anche la consultazione della stessa documentazione, nel passaggio dalla conservazione analogica a quella digitale. Determinante è inoltre la diffusione che è avvenuta a livello internazionale, attraverso modalità che, seppure differenti, convergono in una stessa direzione di valorizzazione di un patrimonio culturale che consente non solo una maggiore sensibilizzazione alla poesia ma anche di esplorare ambiti di ricerca trasversali, utili a una maggiore comprensione su più livelli della dimensione poetica attuale e della sua storia.

2. *Cenni generali sugli archivi vocali di poesia*

Prima di soffermarci su una questione centrale in questo articolo, come quella degli archivi vocali relativi a letture di poesie, è opportuno premettere che diversi studi sono stati condotti nel panorama internazionale, relativamente alla lettura del verso libero e alla sua delicata condizione musicale che gli appartiene. Tra questi menzioniamo la teoria di Cohen (1966), a sostegno di un carattere inespressivo del verso, che apparirebbe a esso congenito e risulterebbe evidente nella sua lettura, a cui si affianca la teoria italiana di Beccaria (1964, 1972). Se alla prosodia del verso libero si è dedicato anche Hartman (1980), alla voce nella poesia sono stati dedicati anche lavori che seguissero un approccio di tipo fonetico sperimentale, come quelli di Crystal (1975), Fónagy (2000), Schirru (2004) e Pamies Bertrán (1999, 2010). Tuttavia, risultano ancora alquanto limitati i lavori sperimentali effettivi dedicati alla questione ritmico-melodica della poesia, in particolare in ambito italiano, dove pionieristici sono stati invece i contributi di Bertinetto (1973, 1978).

2.1 Archivi internazionali di letture poetiche

I maggiori archivi internazionali che raccolgono letture di poesie nascono con l'obiettivo di preservare e conservare il materiale, costituendo in questo modo banche-dati di registrazioni di diverse ampiezze, spesso anche rivolte alla diffusione della poesia stessa. A tal fine sono presenti archivi che includono al loro interno registrazioni audio e, talvolta, anche audiovideo, non solo fisicamente depositati presso enti istituzionali, ma anche presenti *online* e consultabili dagli utenti in forme diverse e in modo gratuito. È difatti appartenente a una recente tradizione la volontà di digitalizzare anche i materiali sonori della poesia. In questo panorama si collocano gli archivi che includono letture originali di autori contemporanei e che, delle volte, offrono anche la possibilità di ascoltarne le letture, a cura di professionisti, in traduzione ad altre lingue. Spesso questi *database* costituiscono parte di progetti internazionali imponenti, che offrono un ventaglio molto largo di dati raccolti e si presentano in continua crescita. Tra le significative realtà internazionali di archivi digitali di letture di poesia si menzionano brevemente alcune tradizioni più rappresentative, che hanno permesso di conservare una tradizione di lettura poetica nazionale, unitamente ad archivi che lavorano invece su un piano diverso, rivolgendosi a più realtà nazionali al contempo. In tal senso, è opportuno passare in rassegna alcuni degli archivi più considerevoli al momento nel panorama mondiale, mirati alla lettura della poesia e che individueremo in quattro filoni principali: ispanofono, anglofono, francofono e germanofono. Nell'area ispanofona sono andati affermandosi archivi ampi come *A media voz, Voces que dejan huellas – Cecilia.com.mx, Biblioteca nacional digital de Chile, The Booksmovie.* In area anglofona menzioniamo invece, tra i numerosi archivi presenti, quello della *Library of Congress of Washington* che, oltre all'*Archive of Recorded Poetry and Literature* e al *The Poet and the Poem Series*, ospita anche il *The South Asian Literary Recordings Project*, l'*Archive of the Now*, il *PennSound*, l'*UbuWeb*, il *The Poetry Archive*, il *Woodberry Poetry Room*, *The Poetry Archive*, *Raleigh Review*, *Academy of American Poets*, *The Cortland Review*, *Essential*

American Poets, *From the Fishouse*, *The Internet Poetry Archive*, *IPRA – Irish Poetry Reading Archive*, *Podcasts.ie: Voices from Ireland*, *The Poetry Foundation*, *voca – University of Arizona Poetry Center's Audio Video Library*, *Canadian Poetry Audio Archives*, *Woodberry Poetry Room – Harvard University*[1]. Nell'ambito francofono menzioniamo *Le patrimoine sonore de la poésie*, mentre in quello germanofono ricordiamo *Lyrikline*, forse il più rappresentativo e aggiornato archivio rivolto alla raccolta di dati sonori di diverse letterature. Quest'ultimo, difatti, nato a partire dal 1999 a Berlino e aggiornato settimanalmente e quotidianamente nel corso di questi anni, include un numero imponente di dati di tutto il mondo, così da compiere anche una missione transculturale non secondaria e dando al progetto un taglio non di raccolta e proposta prevalentemente nazionale ma mondiale. Sempre in una prospettiva internazionale si consideri il recente tentativo di mappa sonora della poesia della *Poetry Sound Library*.

Offre un'ampia raccolta di materiale aggiuntivo anche il generale *The Internet Archive*, nella sua sezione specifica *Audio Books and Poetry*, e a questi si possono aggiungere, per ciascuna delle varie realtà nazionali, gli archivi delle rispettive radio-televisioni di Stato e private, insieme ai numerosi centri studi dedicati ai singoli autori.

Tuttavia, in uno scenario così variegato e che dimostra di avere sviluppato al suo interno una tradizione, appare ancora fondamentale e non sempre esauriente lo sviluppo di archivi nazionali vocali più mirati, in grado di avere, oltre a un fine documentale, l'obiettivo di rappresentare materiale di ricerca scientifica. Sebbene le realtà geografiche precedentemente menzionate siano da più tempo sensibili alla questione e si siano dedicate alla raccolta e alla digitalizzazione di testimonianze nel corso di questi anni, il mondo italiano presenta, a sua volta, una sua specifica tradizione di conservazione di materiale prezioso di documentazione. Si tratta di fonti principalmente presenti in fisico in teche dedicate, non sempre di facile utilizzo *online* o molto limitate nella loro parte di libero accesso nella rete e di recupero dei documenti: tuttavia esse costituiscono un riferimento utile per approcciare allo studio di una tradizione di lettura poetica nazionale. Tra i maggiori archivi italiani di riferimento si considerino quelli di Rai Teche, ICBSA Istituto Centrale per i Beni Sonori ed Audiovisivi, quelli di diversi Centri Studi dedicati ad autori nazionali, tv e radio locali, RSI Radio Svizzera Italiana, festival letterari (si pensi, tra gli altri, a quello del *Festivaletteratura* di Mantova) e, in particolare, quello dell'*Archivio della voce dei poeti* (situato a Pontassieve).

In comparazione con le differenti tradizioni sviluppate in altri Paesi nello sviluppo di archivi digitali *online*, in Italia appare degno di nota l'archivio del progetto *Phonodia*, curato dall'Università degli Studi di Venezia (Dipartimento di Filosofia e Beni Culturali), rivolto non solo alla poesia italiana ma anche straniera e, in particolare, spagnola.

Un archivio nazionale esclusivamente rivolto alla raccolta delle letture di poeti italiani, disponibile *online* e in grado di fornire allo studioso il materiale di ascolto

[1] Si considerino in questo elenco i cataloghi di letture singole degli autori o interpretate da attori e ascoltabili *online*.

in modo immediato, è finora mancato. Per questo, con l'intento di costituire un riferimento per l'ascolto delle voci della poesia italiana e facilmente utilizzabile ai fini di Ricerca, è stato sviluppato il progetto *Voices of Italian Poets*, nato nel Laboratorio di Fonetica Sperimentale "Arturo Genre" dell'Università degli Studi di Torino, che andremo ad approfondire nei paragrafi successivi.

2.2 Archivi di registrazioni poetiche e progetti di Ricerca

Globalmente, come visto finora, gli archivi citati rappresentano banche di dati, che includono al loro interno anche digitalizzazioni o dati raccolti appositamente, e costituiscono vere e proprie audioteche, non rivolte però prettamente alla Ricerca. Tuttavia, la ricchezza di materiali simili si presta a una sinergia con il mondo accademico e con l'approfondimento e lo studio delle sue potenzialità. Risultando però ancora questo campo di studio alquanto inesplorato, pochi sono ancora i lavori rivolti a questa tematica e che prendano forma a partire da questi dati.

Progetto di rilievo nel panorama europeo degli studi afferenti alla lettura del testo è, ad esempio, quello inglese di *Hearing the Voice*, finalizzato allo studio della voce dell'interiorità dello scrittore e concentrato sulla dimensione endofasica. Interessante è anche il lavoro interdisciplinare svolto sull'ascolto nell'ambito del progetto londinese *Listening across Disciplines*.

Studi invece che mantengano un impianto prettamente fonetico rivolto alla lettura della poesia, invece, sono ancora limitati, non solo sul piano nazionale ma a livello internazionale. Recentissimo è però il nuovo progetto di Ricerca che, utilizzando l'archivio di *Lyrikline* nella sua parte tedesca e inglese, investiga la questione ritmica della poesia attraverso metodologie di fonetica e grazie allo sviluppo di tecnologie specifiche: il gruppo tedesco di *Rhythmicalizer* ha sede nella *Freie Universität Berlin* e si afferma nel 2018 come tale.

In ambito italiano il progetto di *Phonodia* (Mistrorigo, 2018), al contempo, si finalizza non solo alla conservazione e diffusione del materiale sonoro ma anche a uno studio specifico del dato sonoro, con una metodologia letteraria che si avvicina alla linguistica e, in particolare, alla strumentazione della Fonetica. In questo contesto si colloca dunque anche il progetto *Voices of Italian Poets*, che vuole unire nella sua natura una declinazione archivistica conservativa insieme a una vocazione investigativa, dedicata alla ricerca della realtà sonora della poesia italiana del Novecento e dei contemporanei, vista attraverso le metodologie della Fonetica sperimentale e dei suoi strumenti.

Oltre a questi gruppi di ricerca menzionati di area europea si citino anche quello americano di MacArthur *et al.* (2018) e il lavoro di Mustazza sull'archivio PennSound.

3. *Voices of Italian Poets*

3.1 Il progetto: linee generali dell'archivio

Il progetto VIP – *Voices of Italian Poets* nasce con l'intenzione di costituire una piattaforma di riferimento per l'ascolto della poesia italiana e, soprattutto, per il suo studio scientifico, declinabile in più direzioni, seppure mirato, nello specifico, a un approccio fonetico di tipo sperimentale.

La piattaforma è ospitata *online* dal sito del Laboratorio di Fonetica Sperimentale "Arturo Genre" dell'Università degli Studi di Torino e si presenta in tre sezioni principali: l'archivio, la sezione relativa alle metodologie di studio e infine quella dedicata alla giornata internazionale di studi, ideata e organizzata annualmente in occasione della giornata mondiale della poesia.

Prendendo in considerazione più nel dettaglio la parte di archivio del progetto, presentiamo brevemente la piattaforma così come appare allo stato attuale. Essa accoglie al suo interno registrazioni della poesia italiana del Novecento e del nostro tempo, corredate da annotazioni, analisi e comparazioni fonetiche che fanno uso di una metodologia di studio esplicata nella piattaforma stessa. Sono incluse al momento oltre 600 registrazioni di letture, non solo di poeti che leggono sé stessi ma anche di attori e poeti che leggono altri poeti, scelti all'interno di una selezione basata sulle letture originali di alcuni autori del Novecento e su una rappresentativa varietà stilistica e metrica di scrittura. Nel complesso quella che si presenta è una istantanea della poesia italiana allo stato attuale e degli ultimi quarant'anni circa, che include letture di poeti di differenti generazioni, col fine di offrire un ventaglio più aperto di ascolto e non escludere la possibilità di documentare anche eventuali fasi diverse di lettura, relative a diversi periodi della vita, in un medesimo autore. A tal proposito, sono inclusi, tra i viventi, poeti di chiara fama, poeti più giovani già affermati e con solidi percorsi poetici e infine anche un ristretto numero di esordienti, già riconosciuti a livello nazionale dalla Critica.

È possibile inoltre, relativamente al materiale di autori del Novecento, compiere almeno una parziale rassegna del materiale originale presente in altri archivi, di cui si menziona la presenza, laddove non sia stato possibile inserire *online* il materiale originale per motivi di diritti d'autore. Il materiale audio raccolto è frutto di registrazioni effettuate in cabina silente modulare, presso il Laboratorio di Fonetica Sperimentale, oppure a mezzo di registratori ambientali e con studio mobile. Si può trattare talvolta anche di riversamenti di materiali analogici recuperati (come ad esempio l'archivio del Centro di Poesia Contemporanea dell'Università di Bologna, di cui sono in corso di inserimento alcune parti del materiale) o di materiale già pervenuto in formato digitale, tra cui si menzionano anche le registrazioni effettuate e inviate, raramente, dagli autori stessi. I materiali sonori vengono conservati e studiati in formato .wav, sebbene siano inseriti *online* in un formato .mp3, uniformati a 16000 Hz: di tutti gli audio sono inseriti, secondo il protocollo seguito[2], il

[2] L'utilizzo del materiale, come segnalato nella pagina stessa dell'archivio, necessita di una specifica formula indicata. Sono inoltre presenti, a fianco dei *metadati* della piattaforma, alcuni asterischi, in-

nome dell'autore del testo, il titolo della poesia, il nome dell'interprete, il credito del raccoglitore, l'audio[3] e l'eventuale .txt[4] presente. Le letture sono a cura di autori e anche di professionisti della voce, come attori, doppiatori e *speaker* radiofonici. Difatti sono presenti due principali approcci di consultazione dell'archivio: uno comparativo, che permette l'ascolto di differenti voci, alle prese con una selezione di uguali testi, e uno invece specifico, che permette l'ascolto di più letture originali di uno stesso autore. Un approccio comparativo di questo tipo, particolarmente utile in studi di tipo fonetico, in modo da mettere in evidenza elementi ricorrenti e variabili, costituisce anche un elemento unico e nuovo nell'ambito degli studi sulla lettura della poesia, utile in particolare per evidenziare la connessione tra testo scritto e vocale.

Un archivio di tali dimensioni è reso possibile solamente grazie alla collaborazione con alcuni enti coi quali sono state stabilite delle *partnership* importanti ai fini di una più esaustiva rappresentazione, non solo del panorama contemporaneo italiano ma anche di quello del nostro Novecento: è possibile in questo modo incrociare più fonti, così da presentare l'eventuale compresenza di materiale in più banche di dati. Oltre alla già menzionata collaborazione col Centro di Poesia dell'Università di Bologna, si ricordano anche le collaborazioni con l'Associazione Mendrisio Mario Luzi nel mondo, con il Centro Studi Mario Luzi «La Barca» e la cooperazione anche con rassegne poetiche nazionali.

3.2 Il progetto di Ricerca: sviluppo e metodologie utilizzate

Il progetto *Voices of Italian Poets* nasce nel 2017, come sviluppo e ampliamento di un primo nucleo di studio di Fonetica sperimentale, applicato a dodici letture della prosopopea di Giorgio Caproni *Congedo del viaggiatore cerimonioso*. Si presenta nella sua parte *online* (cfr. § 2.1) a partire dal gennaio 2018 (http://www.lfsag.unito.it/ricerca/VIP_index.html): anche se l'archivio consta di ulteriore materiale non disponibile in rete, che aumenta considerevolmente il numero di registrazioni totali, la piattaforma digitale consente un aggiornamento continuo del materiale. A partire dalla prima ricerca caproniana, punto di avvio di questa Ricerca, numerosi spunti sono emersi, al punto da richiedere un ampliamento del lavoro in un archivio che potesse permettere uno sguardo più ampio a livello quantitativo e offrire una panoramica più completa della lettura poetica, concentrata non su un solo autore ma, per lo meno, allargata a una selezione significativa di poeti del Novecento e viventi. Tuttavia, per compiere uno studio attento su un piano fonetico, anche se la

dicanti la disponibilità dei dati sonori nei principali *database* di riferimento (ICBSA e Rai Teche) o la cessione del materiale in licenza gpl per scopi di Ricerca.

[3] Laddove non sia in nostro possesso l'audio o sia presente in altre banche di dati, viene inserito il *link* di rimando alla piattaforma corrispondente o, in alcuni casi, viene inserita unicamente l'indicazione relativa all'archivio corrispondente, nella sezione *credits*.

[4] Sono inseriti *online* in formato .txt i materiali di annotazione relativi agli audio, come conversione dal formato .txtGrid. Ai fini di una visualizzazione di questi sul *software* Praat, è necessaria una ulteriore conversione.

vastità del materiale costituisce un arricchimento di informazioni e una molteplicità di approfondimenti, è necessario scegliere un ventaglio di dati mirato a osservazioni di tipo qualitativo e quantitativo. Difatti, in uno studio come questo, i dati sonori vengono sottoposti a un ascolto e a un processo di annotazione e visualizzazione mediante *software* specifici, come, ad esempio, Praat.

La connessione tra la lettura e il testo poetico, centro di una ricerca di questo tipo, viene dunque a essere osservata attraverso l'ascolto attento e l'osservazione di oscillogrammi e spettrogrammi, unitamente alle annotazioni, effettuate manualmente e su più livelli. A partire dal livello del verso (*Verse Tier*), sono generalmente aggiunti altri livelli, in grado di segmentare le letture secondo l'effettiva esecuzione (ad esempio, si considerino l'*Utterance Tier*, l'*Intonation Tier* e il *Tone Tier*[5]): è primariamente la relazione tra verso e intonazione a essere osservata, unitamente alle caratteristiche specifiche che essa assume nella sua realizzazione e alla concretizzazione prosodica della parte ritmica del testo poetico.

Al fine di osservare i diversi aspetti prosodici del parlato poetico, vengono compiuti studi ulteriori di diverso tipo, che permettono di estrapolare e calcolare informazioni ulteriori attraverso *scripts* e *routines* specifici, tratti, tra gli altri, dal progetto AMPER. La realizzazione di grafici comparativi e la volontà di eseguire una piccola parte di analisi statistica dei dati possono permettere di ricavare informazioni importanti per stendere una prima descrizione comparativa e dettagliata di una tendenza di lettura mutata nel tempo e identificarne le caratteristiche che la rendono possibile.

Per una descrizione esaustiva e adeguata al tipo di lavoro, non ancora effettuato in ambito scientifico sulla lettura della poesia italiana, si fa inoltre uso di una terminologia apposita, che eviti la confusione metodologica con altri approcci diversi da quello della Fonetica, a partire da quella di Colonna (2017). Tra i termini-chiave coniati e che più ricorrono nei nostri lavori si menzionino *verso-curva* (per intendere i casi in cui l'unità intonativa coincide con la linea intera del verso sulla pagina) e parola ritmica (per indicare le unità prosodiche che manifestino, per una specifica evidenza e forma prosodica, punti salienti all'interno della scansione ritmica della curva intonativa).

La descrizione del materiale, come anticipato, si propone come intergenerazionale e come lavoro comparativo in grado di evidenziare diverse realizzazioni di uno stesso testo a opera di parlanti diversi. Entrambe le osservazioni volgono però alla ricerca di elementi comuni e di distacco, consentendo in questo modo di stilare delle prime classificazioni sommarie e di approfondire nel dettaglio le variabili che le differenziano.

I principali stili di lettura, la variazione e la continuità che vive uno stesso testo attraverso voci differenti, l'osservazione su più livelli dell'apparato prosodico, senza escludere anche l'impianto retorico che lo caratterizza, in quanto "sovra-testo" ul-

[5] Sono utilizzate anche etichette specifiche, secondo il lavoro di Romano (2014), in grado di identificare dei *pattern* intonativi ricorrenti e le diverse tendenze melodiche in uso nella lettura della poesia italiana.

teriore, consentono inoltre di approfondire la delicata questione del ritmo nel verso libero a partire dalla lettura e le modalità di attuazione della struttura retorica del testo.

Prendere in considerazione singoli autori nelle loro letture per una rassegna diacronica consente, da un lato, di stilare una prima, seppur parziale, storia della lettura poetica contemporanea. Dall'altra parte, un'analisi che tenga conto di un archivio in cui è possibile ascoltare uno stesso testo in numerose interpretazioni permette di soffermarsi maggiormente sulle potenzialità sonore del testo letterario e prosodico, a partire dal dato acustico.

Una fonte come la registrazione sonora, sempre connessa a quella testuale, da cui non può prescindere un'analisi di questo tipo, costituisce una risorsa importante, finalmente utilizzabile e visibile anche a livello tecnologico, per un approfondimento concreto del verso libero nella sua forma musicale.

Questo tipo di Ricerca si propone di stilare una prima panoramica della lettura poetica e di approfondire alcuni aspetti specifici che la riguardano, non avendo invece l'obiettivo di rappresentare un lavoro di tipo normativo, che intenda dare un giudizio soggettivo al dato oggettivo preso in analisi.

4. *Alcuni dei primi risultati*

Sebbene la fase di annotazione e di studio del materiale sia attualmente ancora in corso, in questa sezione sarà possibile individuare alcuni dei primi risultati emersi e già presentati in altre sedi.

Ci soffermeremo principalmente su due autori del Novecento approfonditi in modo differente, secondo un approccio comparativo, da un lato, e più specifico, dall'altro, ovvero su alcune letture di Giorgio Caproni e Mario Luzi.

Se da un'osservazione di 12 letture della più nota prosopopea caproniana del *Congedo del viaggiatore cerimonioso* (attualmente presente nel nostro archivio in oltre 50 letture, di cui si prevede uno studio di tipo quantitativo) emergevano due prevalenti tendenze melodiche, una caratterizzata da omogeneità e l'altra invece da varietà interna, è opportuno specificare che questa scissione evidente appariva anche in coincidenza di alcune specifiche categorie di parlanti[6]. Più nel dettaglio, una maggiore omogeneità (che non chiameremo "monotonia" cfr. Beccaria, 1964 e Cohen, 1966) veniva evidenziata nella lettura di quattro poeti, opponendosi invece a una modalità intonativa più varia al suo interno nelle letture di quattro poeti e quattro professionisti della voce. Anche l'aspetto della segmentazione prosodica si individuava in tre categorie generali, che erano state nominate nei seguenti modi: "metrica", laddove la segmentazione delle unità prosodiche corrispondeva al verso (così come appare nella pagina); "sintattica", quando la lettura seguiva la sintassi e la

[6] In questo studio preliminare, che ha costituito il nucleo di avvio delle successive ricerche, erano state prese in analisi le interpretazioni del poeta stesso, di sei poeti contemporanei (tre uomini e tre donne), invitati a leggere appositamente il testo, due *speaker* radiofoniche e poetesse e infine tre attori (di cui uno era Achille Millo, dedicatario della stessa poesia e contemporaneo di Caproni).

sua punteggiatura, infrangendo quindi anche il *layout* del testo (e con esso anche "il bianco della pagina", col suo silenzio); "metrico-sintattica", nei casi di lettura che tenessero conto di una sintassi globale, ma strutturassero i respiri facendo riferimento, prevalentemente e laddove possibile, al verso col suo metro. L'osservazione anche delle pause aveva permesso di individuare un diverso uso del silenzio tra i locutori uomini e i locutori donna, così come anche tra le diverse categorie professionali. Anche una prima osservazione di figure retoriche del testo, come l'*enjambement*, viste in relazione alla lettura effettiva, avevano fornito informazioni di particolare interesse. Infine, emergeva una struttura retorica della lettura, con la sua struttura e i suoi andamenti, da un'analisi dettagliata dell'interpretazione originale di Giorgio Caproni: essa andava così a sovrapporsi allo schema retorico soggiacente al testo.

Per quanto riguarda invece le letture luziane, il tipo di studio svolto vedeva un'analisi concentrata su un nucleo ristretto di letture originali dell'autore, unitamente a una comparazione con una lettura attoriale di uno stesso testo. Se la lettura caproniana risaliva al 1983[7], la selezione luziana era invece, eccetto che per la registrazione di *A che pagina della storia* del 1985, degli ultimi anni del poeta, deceduto nel 2005[8]. Anche in questo caso è stata presa in considerazione, su più interpretazioni del medesimo autore, la sua tendenza a gestire il comportamento melodico e a organizzare la segmentazione prosodica, dimostratasi particolarmente interessante, se considerata alla luce anche delle teorie metriche dell'autore, molto sensibile alla questione del metro, nell'ottica di una tradizione nazionale. Altri aspetti rilevanti sono stati quelli dell'osservazione del *verso a gradino*, tipico di una gran parte della produzione luziana, della mancata produzione di *enjambement* e della ricorrenza di alcune fasce melodiche specifiche (in parte verificatasi anche in Caproni), emersa dall'analisi della curva di densità di probabilità di f_0.

Non approfondiremo in questa sede i dati presi in analisi: tuttavia, eventuali dimostrazioni applicative e ulteriori aspetti teorici, relativi alla metodologia e alla segmentazione o relativi all'analisi della voce di poeti e letture diverse possono essere approfonditi, rispettivamente, in Colonna, Romano (2019a) e in Colonna, Romano (2019b).

5. *Conclusioni*

Il progetto VIP, con la sua doppia anima di *corpus* di dati acustici e di progetto di Ricerca, mira a costituire nel tempo un riferimento per gli studiosi della dimensione musicale della poesia, e un progetto-pilota per studi di Fonetica di questo tipo, in Italia non ancora decollati. Diverse sono ancora le prospettive e le potenzialità che

[7] Si tratta di una registrazione conservata presso l'ICBSA Istituto per la Conservazione dei Beni Sonori e Audiovisivi, a Roma.

[8] La selezione delle registrazioni considerate rientra nelle letture dall'opera *Sotto specie umana* e appartiene all'archivio dell'Associazione Mendrisio Mario Luzi nel mondo. La lettura anteriore invece, tratta dall'opera *Il fuoco della controversia*, è tratta da un programma radiofonico presente in Rai Radio 6 Teca.

il progetto include e che ci si ripromette di potere sviluppare e ampliare nel tempo, incluse anche quelle relative all'utilizzo del materiale in ambito didattico e degli ipovedenti.

Voices of Italian Poets si propone come realtà che, seppure ancora giovane, ambisce alla interdisciplinarietà e alla crescita mediante il confronto e la condivisione dei propri dati e degli studi connessi. Seppure in questa sede non si siano presentate nel dettaglio le prime linee di Ricerca applicate ai dati, il tipo di analisi in corso si presenta, per quanto ancora in fase di sviluppo, particolarmente interessante e ci si ripromette di approfondirlo in altre sedi.

I materiali raccolti e ordinati nell'archivio offrono numerose declinazioni di studio e di utilizzo, così come diverse possono essere le diramazioni che da un lavoro di questo tipo possono emanarsi: finalità del progetto è dunque anche la loro documentazione all'interno della piattaforma che, in quanto tale, si propone di farsi connessione, non solo tra voce e testo scritto, ma anche tra ascoltatori e interpreti, tra passato e presente, nel segno della natura più profonda e originale della poesia e della conoscenza.

Bibliografia

Beccaria, G.L. (1964). *Ritmo e melodia nella prosa italiana*. Firenze: Olschki.

Beccaria, G.L. (1975). *L'autonomia del significante*. Torino: Einaudi.

Bertinetto, P.M. (1973). *Ritmo e modelli ritmici: analisi computazionale delle funzioni periodiche nella versificazione dantesca.* Torino: Rosenberg & Sellier.

Bertinetto, P.M. (1978). Strutture soprasegmentali e sistema metrico. In Cremante, R., Pazzaglia, M. (Ed.), *La metrica*. Milano: Il Mulino, I, 67-78.

Cohen, J. (1966). *Structures du language poétique*. Paris: Flammarion.

Colonna, V. (2017). Prosodie del «Congedo». Analisi fonetica comparativa di dodici letture. Tesi magistrale, Università degli Studi di Torino (a.a. 2016-2017).

Colonna, V., Romano, A. (2019a). Introduzione alla prosodia luziana. In *Luziana*, 3, 13-23.

Colonna, V., Romano, A. (2019b). Variazioni intonative del *Congedo*. Analisi comparativa di dodici letture. In de Castro Moutinho, L. (Ed.), *Variação Linguística nas Línguas Românicas* (Atti del convegno di Aveiro, Portogallo, 2-4 maggio 2018), ebook, 375-390.

Crystal, D. (1975). *The English Tone of Voice: Essays in Intonation, Prosody and Paralanguage Usage*. London: Edward Arnold.

Fónagy, I. (2000). Languages and Iconicity. In Violi, P. (Ed.), *Phonosymbolism and Poetic Language.* Turhout: Brepols.

Hartman, C.O. (1980). *Free Verse. An Essay on Prosody.* Princeton: Princeton Legacy Library.

Mistrorigo, A. (2018). *Phonodia. La voz de los poetas, uso crítico de sus grabaciones y entrevistas.* Venezia: Ca' Foscari.

Mac Arthur, M.J., Zellou, G. & Miller, L.M. (2018). Beyond Poet Voice: Sampling the (Non-) Performance Styles of 100 American Poets. In *Journal of Cultural Analytics*. April 18, 2018. DOI: 10.31235/osf.io/5vazx.

Pamies Bertrán, A. (2010). Quelques malentendus à propos du concept de rythme en linguistique. In Russo M. (Ed.), *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Roma: Aracne.

Romano, A. (2014). Etichette per l'analisi prosodica di file di parlato. http://www.lfsag.unito.it/ricerca/Etichette_prosodiche_2014-18.pdf.

Schirru, G. (2004). Costituenza metrica e lingua poetica italiana. In Albano Leoni, F., Cutugno, F., Pettorino, M. & Savy R. (Ed.), *Il parlato italiano* (Atti del convegno nazionale di Napoli 13-15 febbraio 2003). Napoli: D'Auria, testo F9.

Sitografia

Phonodia (UniVe) http://phonodia.unive.it/ (ultima consultazione 24/09/2018).

Rhythmicalizer (Freie Universität Berlin) https://www.geisteswissenschaften.fu-berlin.de/v/rhythmicalizer/index.html (ultima consultazione 24/04/2019).

DOI: 10.17469/O2106AISV000002

CARLO GERACI, ROLAND PFAU, PIETRO BRAIONE, CARLO CECCHETTO, JOSEP QUER

Hidden languages in a digital world: The case of sign language archives

SIGN-HUB is a European project involving collaborators from seven countries funded within the Horizon 2020 Research and Innovation program. The scope of the project is both socio-cultural and linguistic, as its aim is to document and preserve the culture, the history, and the languages of European Deaf communities. After a brief description of the various components of the project, we focus on the documentation of the life stories of Deaf people and the creation of a digital sign language archive.

Key words: sign language, digital archive, Deaf culture, documentation, preservation.

1. *Introduction*

Sign languages are the natural languages used by Deaf[1] people in everyday communication. They have been shown to display complex grammatical structures fully on a par with spoken languages, despite the fact that they employ a different modality for signal production and perception (Sandler, Lillo-Martin, 2006; Pfau, Steinbach & Woll, 2012). That is to say, while spoken languages are based on the auditory-vocal modality, sign languages make use of the visual-spatial modality. Most sign languages are minority languages in the sense that they are typically immersed in an environment with a dominant spoken language[2], and indeed they share some features with minority spoken languages: they are often marginalized or even discriminated, they are not taught at schools, and their users may constitute a subculture within the mainstream culture (see, e.g., Burns, 1996 for Irish Sign Language and Mougeon, Nadasdi, 1998 for a general perspective on minority language communities). However, they also differ from minority spoken languages in several ways: beside the modality of transmission, the pattern of language acquisition is unique, since typically only 5-10% of deaf children are born into Deaf families where sign language is present, and most Deaf signers are exposed to sign language beyond the very first years of life (see Quer, Steinbach, 2019).

[1] We follow the by now well-established convention of distinguishing 'deaf' as the physical condition of lack of hearing (with lowercase d) from 'Deaf' as the cultural and linguistic identity of signers (with capital D).
[2] A notable exception are shared sign languages, sometimes also called rural or village sign languages, a term that identifies sign languages that are used in small communities with an unusually high incidence of (often hereditary) deafness. In such contexts, deafness is often less stigmatized, and a high percentage of hearing community members is also fluent in the local sign language. For an overview of shared sign languages and their specific linguistic properties, see Nyst (2012) and De Vos, Pfau (2015).

One peculiarity of sign languages is that none of the Deaf communities of the world has independently developed a writing system for their sign language, the main reason being that Deaf people are generally educated in the dominant spoken language (i.e., they can at least read and write the dominant spoken language, albeit at varying levels of proficiency)[3]. The direct consequence of this fact is that sign language communities are prominently "oral" communities (Byrne, 2016), which implies that their culture – including artistic expressions like storytelling, poetry, and theatre – is transmitted "orally" (i.e. in a non-written, visual form) via sign language and has therefore, for the longest time, not been documented[4].

In fact, the lack of a writing system or other suitable ways to transmit cultural productions is one of the main reasons why historians of Deaf communities always have troubles in finding first-hand documents. Sign language as a means of communication for the Deaf is already mentioned in Plato's *Cratilus*, but the oldest film document of a signer probably comes from American Sign Language (ASL) and dates back only to 1913. In this film, George Veditz, the president of the National Association of the Deaf, talks about the *Preservation of the Sign Language* (film title) and famously expresses "As long as we have deaf people on earth, we will have signs". This and subsequent movies are now considered some of the most significant documents in Deaf history (for historical ASL materials, see Supalla, 2001 and Supalla, Clark, 2015)[5].

In this paper, we describe the preliminary results of the Sign-Hub project, the first systematic attempt (i) to document the history, the culture, the experiences, the identity, and the languages of various European Deaf communities; and (ii) to store all of this information and make it available via a digital on-line platform. That is, the project has two closely related goals, which, in a sense, are like two sides of a medal: on the one hand, it is devoted to the creation of sign language-related content; on the other hand, it involves the development of a digital infrastructure that a) allows for immediate access to sign language content, and b) guarantees long term preservation of the digital files.

In the remainder of the paper, we first offer a general overview of the Sign-Hub project (Section 2). In Section 3, we then offer a description of the components of the sign language archive. The solutions we are implementing for sharing, preserving, and protecting the digital files are described in Section 4. Section 5 concludes the paper and addresses some of the future challenges.

---

[3] There are various transcription systems that employ specific symbols for representing the sub-lexical components of signs. However, these systems are highly intricate and are only used for scientific purposes (for an overview, see Frishberg, Hoiting & Slobin, 2012).

[4] Of course, there are many books and papers – both of a scientific and a more popular nature – on sign languages, Deaf culture, Deaf communities, and Deaf education. But crucially, these publications are not composed in the language of the population they talk about. For a detailed description of the Deaf world, see Bauman (2008), Gertz, Boudreault (2016), Goodstein (2006), and Parasnis (1996). Quotes from Deaf subjects, for instance, are usually translated into the language of the publication. See for instance, Padden, Humphries (2005) and Leigh, Andrews & Harris (2018).

[5] The video can be viewed on Youtube.

2. *The SIGN-HUB project*

SIGN-HUB (2016–2020) is a European project funded within the Horizon 2020 Research and Innovation program[6]. It involves a network of ten universities and research centers from seven different countries, coordinated by Josep Quer from the Universitat Pompeu Fabra in Barcelona[7]:

- France: CNRS, Institut Jean-Nicod, Paris – coordinator: Carlo Geraci;
  University of Paris-Diderot – coordinator: Caterina Donati;
- Germany: Georg-August University Göttingen – coordinator: Markus Steinbach;
- Israel: University of Tel Aviv – coordinator: Naama Friedmann;
- Italy: Ca' Foscari University, Venice – coordinators: Chiara Branchini and Anna Cardinaletti;
  CINI, Milan – coordinator: Mauro Pezze;
  University of Milano-Bicocca – coordinator: Carlo Cecchetto;
- The Netherlands: University of Amsterdam – coordinator: Roland Pfau;
- Spain: Universitat Pompeu Fabra, Barcelona – coordinator: Josep Quer;
- Turkey: Bogaçizi University, Istanbul – coordinator: Meltem Kelepir.

Within the general frame of preserving the culture, the history, the identity, and the languages of European Deaf communities, SIGN-HUB has two main goals: the production of materials and contents related to Deaf communities and their sign languages and the development of the digital infrastructure needed to make the content accessible, and to protect and preserve it in the long-term.

2.1 Documenting sign languages and Deaf communities

There are four main outputs of this part of the project: (i) grammars of six European sign languages, (ii) a sign language atlas, (iii) clinical tests for the assessment of sign language abilities in special populations, and (iv) an archive of the life stories of elderly Deaf signers. Here, we briefly address the first three components, while the last one will be discussed in more detail in Section 3.

The SignGram Blueprint is an open access tool for language specialists and sign language linguists that provides detailed information on how to write a sign language grammar (Quer, Cecchetto, Donati, Geraci, Kelepir, Pfau & Steinbach, 2017)[8]. In itself it is not a grammar, but it contains (i) a checklist of relevant grammatical phenomena and processes from which the table of content of a grammar can be built; (ii) a detailed discussion of background information related to the linguistic features, phenomena, and constructions listed in the checklist; these are

---

[6] The SIGN-HUB grant agreement number is 693349.

[7] More information about the teams and the specific goals of each research unit can be found on the project website (https://www.unive.it/pag/33750/) and on the national SIGN-HUB websites.

[8] This publication is the outcome of another European project, SignGram (2011-2015), of which all the participants to SIGN-HUB (except Israel) were members (COST Action 1006: Unravelling the grammars of European Sign Languages; European Science Foundation). The SignGram Blueprint is freely accessible at: https://www.degruyter.com/viewbooktoc/product/467598.

described from the perspective of languages in the visual-gestural modality; and (iii) guidelines on how to identify and analyze linguistic phenomena (including suggestions for data collection) as well as relevant references. On the basis of these detailed guidelines, grammars of the following six sign languages will be provided within SIGN-HUB: Catalan Sign Language (LSC), German Sign Language (DGS), Italian Sign Language (LIS), Sign Language of the Netherlands (NGT), Spanish Sign Language (LSE), and Turkish Sign Language (TİD). One crucial aspect of these grammars is the fact that, in addition to the usual text descriptions, various types of visual materials (images and digital videos) will be included. Besides the intrinsic value of detailed language descriptions, this is the first step towards the long-term preservation of the individual sign languages. We expect the grammars of these languages to be used by language instructors and teachers as a tool for creating additional teaching materials for sign language courses. We also expect that the grammars will give increased visibility – and thus status – to each of these languages, and that local Deaf communities may capitalize on such a resource to advocate in favor of sign language recognition where it has not happened yet, and/or new policies in the school education of Deaf children (e.g., bilingual-bimodal educational approaches).

Along with in-depth language descriptions, SIGN-HUB also aims at describing typological variation across sign languages (e.g., Perniss, Pfau & Steinbach, 2007; Zeshan, Palfreyman, 2017). This goal will be achieved by developing an atlas dedicated to sign language structures, modelled on existing ones like WALS, SSWL, and APiCS[9]. Specifically, four questionnaires have been prepared (phonology & lexicon, morphology, syntax, pragmatics & socio-history). Each questionnaire is designed to include around fifty questions targeting a variety of aspects from each domain of sign language linguistics. The plan is to collect information from at least one-hundred sign languages from all over the world. The questionnaires will be filled in by experts in the particular sign language, and, if available, additional experts from different domains can be consulted for providing missing information for a particular language. We expect the sign language atlas to have a large and immediate impact on the study of typological variation across sign languages. Given the design of the questionnaires and their implementation into the on-line atlas, we expect the data collection of this part of the project to continue after 2020 with new questionnaires covering additional sign languages to be added in the future. As with the sign language grammars, in this respect, too, local Deaf communities can widely benefit from this tool. In smaller communities in particular, realizing that grammatical phenomena or features attested in their sign language are also found in other (related or unrelated) languages will likely reinforce (or even initiate) the process of language awareness, upon which Deaf identity is built.

Accurate detection of language disorders is a key factor for early intervention in clinical linguistics, both for adults with post-stroke or progressive aphasia and for

[9] More information about these projects can be found at the following websites: WALS (https://wals.info), SSWL (http://test.terraling.com/groups/7), and APiCS (https://apics-online.info).

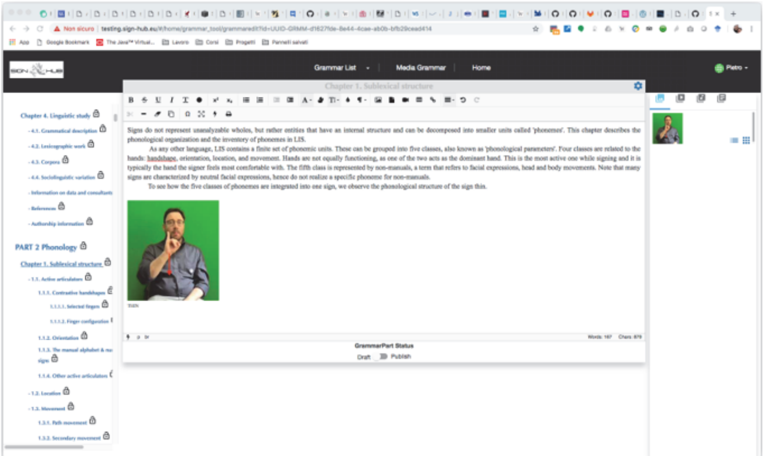
children with developmental language disorders. Significant progress has been made on spoken language assessment tools. However, nothing comparable has been put in place for sign languages (see Mann, Haug, 2014 for an overview of assessment tests in sign language), although valuable work has been done for some sign languages, i.e., ASL (Hauser, Paludneviciene, Riddle, Kurz, Emmorey & Contreras, 2016) and British Sign Language (for more info see DCAL Assessment Portal, https://dcal-portal.org). The third component of SIGN-HUB is therefore dedicated to creating a variety of assessment tests to detect lexical impairment and measure proficiency in various grammatical domains. One key aspect of the project is to standardize the tests across healthy populations of signers with various levels of sign language exposure. Indeed, the signing population is quite stratified and includes Deaf native signers (typically, deaf people born into signing families), who represent only a small percentage of deaf signers, signers with early exposure to sign language (deaf people born into hearing families who are exposed to sign language before school age), and late learners (deaf people who are exposed to sign language only during or after the school age). While hardly detectable in normal conversation (at least in some cases), the various groups of signers may have different levels of competence with respect to various, often subtle, properties of the grammar (for recent discussion, see Mayberry, Kluender, 2018). The tests will cover all the areas of SL grammar including phonology, morphology, lexicon, syntax and semantics. We expect the tests to be systematically used for early diagnosis and clear identification of language pathologies both in Deaf adults and children.

2.2 Digital infrastructure

This part of the project is dedicated to the creation of an on-line digital platform to host the materials and products described in Section 2.1. Without entering into the technical details of the coding, there are two parts of the platform with which people will be able to interact. One is the set of end-user interfaces/tools developed to implement each of the four research components of SIGN-HUB (grammars, atlas, assessment tests, digital archive); the other is the set of end-user interfaces that will allow display of and access to the final products of each research component. At this stage of the project, only the first set of tools has been developed, which are the ones used by SIGN-HUB members to upload and create their materials. The tools for the grammar writers, atlas questionnaires, and language assessment are briefly described here. Issues related to the video archive will be addressed in Section 3.3.

The grammar tool is designed for grammar writers to type text materials (including tables) directly onto the platform and/or to upload their materials (e.g., texts produced with external software, images, and videos). As illustrated in Figure 1, the workspace is divided into three main areas. On the left side, the grammar writer finds the table of content, which is based on the SignGram checklist, and which allows for easy navigation from one part of the grammar to another by clicking on the links. The central part contains a rich text editor, where text can either be typed or be pasted already formatted from other sources. The right side contains a quick access to the database of visual materials (images and videos) to be used in the grammar.

Figure 1 - *Outline of the Grammar Tool*



Figures 2 and 3 illustrate the workspace of the atlas and the assessment components. The basic concept is that the questions of a questionnaire (atlas) and the items of a test (assessment) are structured like presentation slides. The buttons on the right side of the workspace offer quick access to direct measures (e.g., multiple answers, mutually exclusive checkboxes, continuous scales, etc.), while the central area corresponds to the slide itself. An adjustable grid allows for adapting the positioning of the various objects, text boxes, clickable images, videos, etc. Finally, on the right side, visual materials from the SIGN-HUB database of images and videos can be quickly dragged and dropped on the slide.

Figure 2 - *Outline of the Atlas Tool*

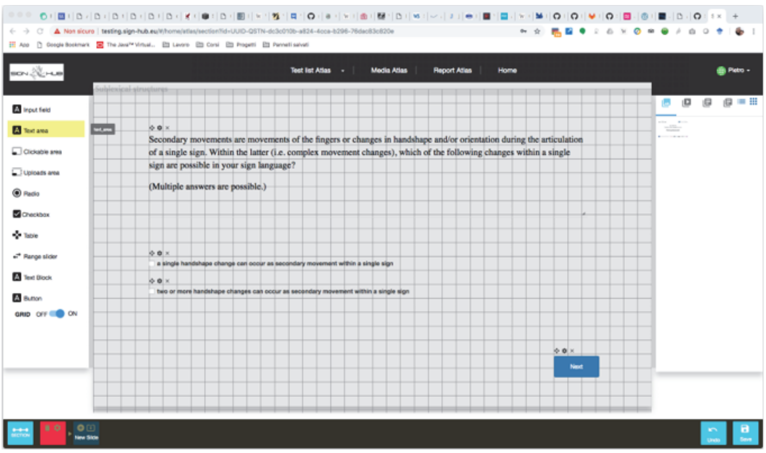
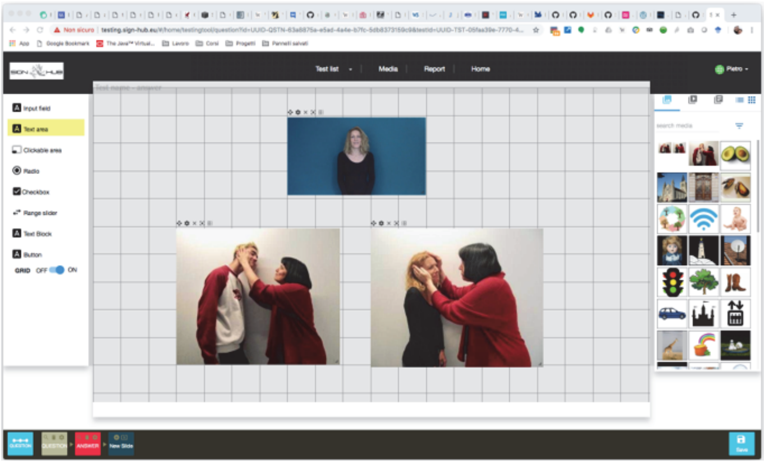


Figure 3 - *Outline of the Assessment Tool*



3. *Life stories of Deaf people*

3.1 Significance and urgency

As already mentioned in the introduction, Deaf communities have not developed writing systems for sign languages for everyday use. Therefore, Deaf communities are intrinsically bound to "oral" transmission of their cultures in the sense of passing on cultural values, traditions, and productions in a non-written form. One clear example is the role of narratives in the development of a Deaf identity in deaf children (Sutton-Spence, 2010)[10]. Although numerous publications document (aspects of) Deaf communities and their languages (e.g., Monaghan, Schmaling, Nakamura & Turner, 2003; Woll, Ladd, 2003; McCaskill, Lucas, Bayley & Hill, 2011), access to primary data has always been problematic until video recording tools have been made available at reasonable prices. Indeed, it was only after cameras and videotapes have been accessible to "non-professionals" that cultural events of Deaf people started being documented. There are three crucial facts which make starting a systematic documentation of Deaf history, culture and communities particularly urgent and compelling.

First, the first half of the 20th Century has witnessed some of the most devastating events of human history (the two World Wars, the Shoah, the atomic bomb). Of course, Deaf people have their own perspective on these (and other) events. For instance, most Deaf children were left in residential schools during World War II with little information about what was happening around them and why (e.g., bombing, starvation, sterilization programs, etc.), as explained in the documentary *1939–1945 Que Faisaient Les Sourds?* ('1939–1945 What did the Deaf do?') on the experiences of Deaf people in France during the second world war. Yet, given

[10] See also Rutherford (1993) and Rayman (1999); for more comprehensive overviews of aspects of Deaf Culture (in the United States), see Padden, Humphries (2005) and Leigh, Andrews & Harris (2018).

that the survivors are now elderly signers, we are left with a very short time-window for documenting their unique memories.

Second, the entire approach to deaf education has radically changed in recent years, in part due to the need of social integration between hearing and deaf children, in part due to the fact that (hearing) parents of deaf children now commonly choose for a cochlear implant (Spencer, Marschark, 2003; also see Humphries, Kushalnagar, Mathur, Napoli, Rathmann & Smith, 2019 for an overview of strategies to support parents with deaf children). As a result, (residential) deaf schools disappeared almost everywhere, and deaf children are nowadays mostly educated in mainstream schools with some additional provisions (e.g., interpreters, special educators, communication assistants, etc.). This apparently harmless shift has radically changed access to sign language. In the past, even in strictly oralist schools, where sign language was forbidden in class, deaf pupils often used sign language to communicate with each other outside the classroom. Now deaf children can be scattered across mainstream schools, with few (sometimes only a single one) in an entire school. If this happens, the common way for deaf children to have spontaneous access to sign language from peers is basically eliminated (but see Marschark, Tang & Knoors, 2014). In some countries, this change in education happened quite suddenly in the mid Seventies, inducing a significant gap between older and younger generations of signers. Third, even for the rare cases in which some memories have been collected on video, the paradigm changes imposed by the digital era make these recordings quickly obsolete and useless (Castells, 1996).

Taken together, the memories of elderly Deaf signers are a treasure that is at risk of being lost forever (also see Legg, 2016). In the next sections, we will illustrate the steps that have been taken within SIGN-HUB to preserve this treasure. We adopted two strategies: in some of the participating countries, we conducted semi-structured interviews; in others, we digitized documentation that was already available. We briefly describe our methodology of data collection/selection in the next section, and then we summarize the main findings in Section 3.3.

3.2 Materials

A crucial component of our documentation and preservation effort are the interviews with Deaf elderly signers conducted in five of the participating countries, Germany, Italy, the Netherlands, Turkey and Spain[11]. Our goal has been to interview at least 20 signers above the age of 65 per country. In the end, we conducted between 21 and 25 interviews per country, the age range of interviewees being between 66 and 93 years.

All interviews were conducted by Deaf research assistants fluent in the local sign language. They vary considerably in length, but all of them are semi-structured in that they follow a questionnaire that has been developed specifically for the interviews. The questionnaire is organized around various topics (e.g., family, Deaf experiences, historical events), and interviewers were encouraged to cover each of the

[11] Interviews were conducted in six different sign languages, as two sign languages were involved in Spain, i.e., LSC and LSE.

topics. Still, it was also made clear at the outset that the questionnaire was meant as a guideline. Interviewers were informed that they did not have to strictly follow the order of questions, and that the inclusion of other topics that came up during the interview was welcome. The most important consideration was to create a comfortable and safe atmosphere for the interviewee. During basically all interviews, it became clear that the elderly Deaf signers enjoyed sharing their life stories. In fact, they would often bring photos and other memorabilia to share with the interviewer. In total, 137 interviews were conducted of a total length of approximately 175 hours.

All interviews were filmed with two cameras: one on both the interviewer and the interviewee, and one on only the interviewee. Participants were informed about the purpose of the interviews and signed a consent form. Although the video recordings are meant to preserve memory and may at times include non-frontal shooting or the interviewee leaving the scene for a moment, the quality of the videos is such that they can be reused for other purposes. Besides their obvious historical and cultural value, the interviews are also of linguistic use in that they contribute to the documentation and preservation of SLs. The possibility of subjecting the productions of elderly signers to linguistic analysis will thus allow a direct comparison with the sign languages used by younger generations, as commonly done in more traditional corpus-based sociolinguistic studies where signers' age is used as a predictor of language variability (Lucas, Bayley & Valli, 2001; Geraci, Battaglia, Cardinaletti, Cecchetto, Donati, Giudice & Mereghetti, 2011; Fenlon, Schembri, Johnston & Cormier, 2015). Such comparative sociolinguistic research has been initiated within SIGN-HUB by annotating fragments of the videos with the computer-based annotation tool ELAN (Crasborn, Sloetjes, 2010). Finally, varying amounts of data per country have also been subtitled in the local spoken language and English in order to be shared with international non-signing audiences (for instance, in the context of a documentary movie that has been created).

Besides the interviews, we will also include in the digital archive materials that exist outside the SIGN-HUB project. Certain aspects of the life of Deaf people in the 20th century are relatively well documented for the Deaf communities of France and Israel. These existing materials will also constitute an important part of our archive.

As for France, the CNRS video library already contains an English subtitled documentary from 2000 entitled "Deaf witnesses, silent witnesses", directed by Brigitte Lemaine and Stéphane Gatti. The documentary, which is publicly available[12], describes the consequences on the Deaf community of the racial hygiene law imposed by Hitler in 1933, which lead to the sterilization of many people with disabilities including Deaf people; it also addresses obligatory abortions and euthanasia in the context of the so-called T4 program. In addition to this, the Académie de la Langue des Signes Française, a Deaf association based in Paris, has a rich archive of videotapes documenting a variety of activities and events that happened in the

[12] The streaming of the documentary is available at the following address: https://videotheque.cnrs.fr/doc=905.

second half of the 20th century. Some of these videotapes are unique copies already in a deteriorated state. The president of the association, Ronit Leven, selected three movies that are most important to preserve, and these have already been digitized (for the procedure see Section 4). All three are amateur documentaries shot by people working at the Académie de la Langue des Signes Française:

– *1939–1945 Que Faisaient Les Sourds?* ('1939–1945 What did the Deaf do?')
  The movie reports on an event organized by the Académie on March 19, 2004, which dealt with the situation of Deaf people during the Second World War. Footage of the live event is intertwined with short clips on WWII and interviews with both Deaf and hearing signers.
– *La Vie Des Sourds Pieds Noirs Et Juifs D'Algérie* ('Life of the Deaf *Pieds Noirs* and Jews of Algeria')
  This movie reports on a similar event (date unknown), this time about the war in Algeria (1954–1962).
– *Quel Avenir Pour Les Personnes Agées Sourdes?* ('What is the future of the elderly Deaf?')
  This is a collection of interviews with elderly Deaf people in France conducted in the early 2000s.

The digitized documentaries are now in the process of being subtitled in French. Subsequently, they will be voiced-over in French, and, in a third step, English subtitling will also be provided. A noteworthy aspect that makes these movies a unique and crucial contribution to the documentation of Deaf history is the fact that they have been realized either entirely by Deaf people or under the direct supervision and direction of Deaf people in an effort to preserve their own memories. They therefore represent an impressive symbol of Deaf identity and of a minority community.

Turning now to Israel, our original intention had been to digitize and include in the digital archive two types of documents[13]: (i) narratives from Deaf Holocaust survivors stored at the Yad Vashem World Holocaust Remembrance Center in Jerusalem, and (ii) life stories of Deaf elderly signers that had been recorded prior to our project. In the end, however, Yad Vashem decided against handing over their narratives to our open access digital archive. The life stories come from the Deaf archive of the University of Haifa, an impressive collection compiled by the late Irit Meir over the past 20 years. There are approximately 20 stories by elderly signers, but for budget reasons, only 4–5 of these will be included in the digital archive. The signers in the selected recordings are representative of different backgrounds in Israel, and the content of their stories is varied (e.g., Holocaust, immigration, experiences at Deaf school). Together, the stories have a length of roughly one hour, and they will be subtitled in Hebrew and English.

3.3 A glimpse at the life stories

Browsing the available materials – the interviews conducted in the context of the project as well as the pre-existing materials – it immediately becomes obvious that

[13] We are indebted to Rose Stamp for providing this information.

they include a wealth of invaluable information and anecdotes. As for the life stories, most of the interviewees readily shared both confronting and funny experiences. Although the analysis of the interviews, narratives, and digitized materials is still on-going, several recurrent topics are emerging across Deaf communities.

First, many signers talk about experiences during war time, including events that they, as Deaf subjects, may have experienced differently. As mentioned above, the Israel materials include the story of a Holocaust survivor. Second, many interviews include renditions of school experiences. This may be delightful anecdotes of the life at boarding schools for the deaf, but there are also shocking stories about neglect and abuse. Third, almost all signers share experiences of suppression and discrimination, e.g., at school or at the work place. This also includes discrimination against the use of sign language, which leads to the common experience of being excluded. Fortunately, at the same time, the signers also share stories about Deaf empowerment and increasing independence of Deaf people. Fourth, identity issues figure prominently in the narratives. These may have to do with the role of a Deaf person within a hearing family, but issues may also arise when a Deaf person identifies with multiple minorities (e.g., Deaf immigrants or Deaf homosexual people). Finally, some fragments are interesting from a linguistic perspective, as they reflect metalinguistic awareness, e.g., concerning lexical change or the origin of name signs.

The multifarious content of the narratives will be reflected in two outcomes of the SIGN-HUB project: a 40-minute documentary movie and an edited volume. The guiding topic of the movie will be 'independence (or lack thereof)'. The edited volume is expected to contain 13 chapters on cultural, socio-historical, and linguistic aspects of Deaf elderly signers. Most SIGN-HUB countries will contribute; in addition, a number of scholars external to SIGN-HUB (from Belgium, Germany, the Netherlands, and the US) have agreed to contribute a chapter related to the topic of the volume.

In addition, the three French documentaries, all complete interviews (including those that will not make it into the documentary), and the complete narratives from Israel will be part of the SIGN-HUB digital archive.

4. *The digital archive*

The easiness with which people can take videos from digital devices (computers, phones, tablets, etc.) and the storage systems that, as private users, virtually everybody can have access to (e.g., hard-drives, personal cloud space, etc.) may lead people to thinking that digital files are always accessible and imperishable. Of course, this is not the case either because physical space on individual devices is limited or because several factors may undermine access to digital files in the long term (e.g., evolving formats, upgrading software, removal of support by software developers, etc.). Notice that, hardware storage resources like DVDs, CD-ROMs are quickly becoming obsolete both for immediate access and long-term preservation uses. Furthermore, these supports are not ideal once the estimated amount/size of the video materials that SIGN-HUB will produce is considered. Luckily, the problem of digital space for academic projects has been tackled at a higher level. In fact, a few platforms at the European level are now offering digital space to host the products of academic research projects, like DARIAH (https://www.dariah.eu) and CLARIN (https://www.clarin.eu), to mention just two of them.

In this section, we describe the main points of the SIGN-HUB data management plan, with a particular focus on the procedure adopted to digitize, store and make accessible the video materials produced within SIGN-HUB.

4.1 Digitizing old videos and collecting new materials

One of the most complex challenges that the teams working within SIGN-HUB has been faced with is to cope with the need of having high quality videos, the limited storage resources for digital materials, the varieties of standards and devices used in each research unit and the status of already collected materials. Overall there are two approaches: a top-down and a bottom-up approach. In the former approach, a uniform project-internal policy is used for all units imposing the use of similar (if not identical) hardware materials for filming, similar environmental conditions during the recordings, exporting in a pre-defined format, etc. In the latter approach, almost complete flexibility is given to individual units (e.g., mild requirements on the formatting could be given).

Within SIGN-HUB, we adopted a mixed approach depending on the requirements of specific parts of the project. For instance, the materials to be used as part of the assessment tests have to be of the best quality possible and must be as uniform as possible both within a research unit (i.e., the various tests that each units builds must have similar quality), and across units (i.e., the same standard should apply across the research units to make data comparable). On the other hand, for the interviews, the only requirement is that files are saved/exported in MPEG 4 format (H.264, 30 fps). The same requirement applies to the video materials already collected from previous projects. The estimated size of all video materials expected for the project is above 6 TB. Table 1 reports the amount of video materials divided by each country.

Table 1 - *Estimated size of video materials*

Sign language	Estimated total size of produced video files
LIS	3 TB
LSC & LSE	600 GB
NGT	1 TB
TİD	1 TB
LSF	70 GB
DGS, ISL	unknown
Total	>6 TB

One particular challenge for the project was the digitization of the three videotapes shared with the project by the Académie de la Langue des Signes Française. Indeed, the technology for converting tapes is already obsolete and very few centers have it. The Research Center of the Italian public broadcasting television, Centro Ricerche RAI, offered its facility for the conversion[14].

The analog to digital conversion process involved three machines, a video cassette recorder for playback of the video (JVC BR-6400TR), a scan converter providing time base correction (SONY DSC-1024 G) and an analog-to-digital converter (DataVideo Y.U.V to DV DAC-2). Adobe Premiere CS6 was then used for video capturing and editing. The files are then exported in RAW format and in mp4 compressed format.

4.2 Immediate access vs. long-term storage

The ultimate goal of SIGN-HUB is the preservation of culture, identity, memories and languages of European Deaf communities, by giving the largest possible audience access to sign language grammars, atlas, documentaries and to have diagnostic tools to detect language disorders. The idea of implementing a centralized platform for a direct access to these contents seems to be the easiest way to achieve that goal. The activities described in Sections 2 and 3 are a big step towards that goal; however, on the technological side preserving digital files and giving access to them are two almost completely separate goals. Simplifying the issue, the expectation of generalized access to content can be exemplified by a well-organized website with a search engine that provides the requested content directly on the web browser, while the expectation of long-term preservation is that the quality of video remains intact across technological evolution of formats, devices, software, etc. The former expectation is generally met by allowing video files to be always accessible by streaming via internet. The latter expectation is met once data integrity and file fixity are maintained and metadata are stored in a standardized format. Once archived, the data are not immediately accessible and the process of retrieving data from the archive may take a considerable amount of time (which depends on the size of the file, the specific technique used to store the data, etc.).

Institutional providers that offer both services are quite rare. For instance, DARIAH only allows for long-term preservation, but not for streaming files, while only some centers of the CLARIN consortium offer both services, but the requirements for streaming are quite stringent. SIGN-HUB will use two French-based institutional infrastructures: HUMA-NUM (https://www.huma-num.fr) and ORTOLANG (https://www.ortolang.fr). HUMA-NUM is a large digital infrastructure that can be used for storage for all human sciences at the National and European level. ORTOLANG is a parallel network infrastructure which includes a repository of language data (corpora, lexicons, dictionaries, etc.) and readily available, well-documented tools for its processing specific for language and linguistic projects.

HUMA-NUM will host in its repository all files for long-term preservation (i.e., all data that need to be archived and not immediately available for streaming).

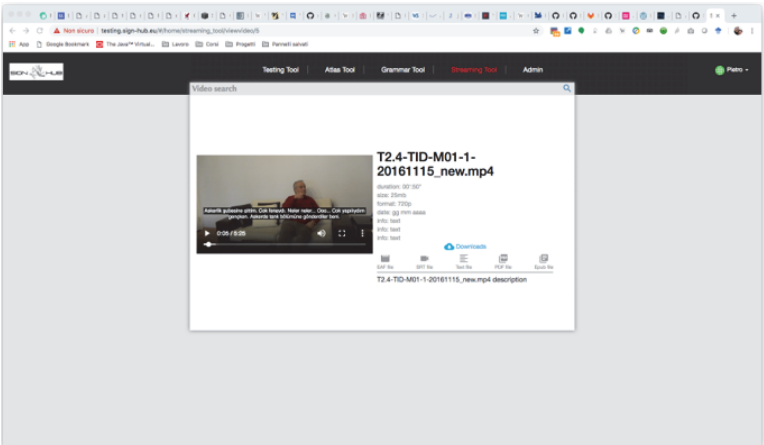
[14] We are very grateful to Andrea Del Principe for the help he provided.

ORTOLANG will host all video data that will be played in streaming, including e.g., video examples of the grammars, of the atlas questionnaires and final products, testing items for assessment, documentaries, etc. ORTOLANG will also host, on a virtual machine, all non-video materials, including the software that will be used to run the SIGN-HUB platform itself, the webpages, the databases of the assessment tests, the atlas questionnaires and the sign language grammars.

One clear advantage of using institutional repositories offering services at the European level, rather than private third-party services, is that they are by law compliant with security measures in terms of research data and sensitive data protection regulations. This is a delicate aspect of SIGN-HUB that cannot be fully addressed in this paper. However, to give a very brief outline of the challenges the project was faced with, consider that all data coming from healthy Deaf signers must be protected with the maximum level of security established by the European regulations. The reason is that, even when the purpose of these data is not medical at all, they come from a minority population with sensory disability. In this respect, institutional repositories offer a safety net in terms of data security.

The current layout of the streaming tool is illustrated in Figure 4. After selecting the video from the archive, users will have access to basic functions like full-screen size, subtitles and other options including downloading ELAN annotated files if available.

Figure 4 - *Layout of the streaming tool*



5. *Conclusions and perspectives*

In this paper, we presented the main goals of SIGN-HUB, an EU funded project whose main goal is to preserve the culture, the identity, the memories and the languages of European Deaf communities. In its core, SIGN-HUB contains a digital video-archive of Deaf communities, which are linguistic minorities using sign languages.

There are several directions in which SIGN-HUB can evolve after the end of the project. As far as the digital archive and the memories of (European) Deaf communities are concerned, future technological developments will further reduce costs both for long-term storage and streaming (simply think about the apparently unlimited space offered for free by private video developers like YouTube, Vimeo, etc.). This opportunity will guarantee sustainability of the current project and preservation of the digital materials created during SIGN-HUB. In parallel, we hope to achieve quite a large impact on other signing communities both within and outside Europe, so that more sign language grammars will be available, the atlas will include more languages and more interviews of elderly signers are made available. In particular, we hope that non-urban signing communities will have access to our platform, on the one hand to benefit from our infrastructure; on the other hand, to extend the domain of sign language research. Another direction would be that of expanding the domain of the archive by including recordings of artistic performances like theater and poetry performances, which constitute an important component of Deaf culture in Western societies.

Bibliography

Bauman, H.-D.L. (Ed.) (2008). *Open your eyes: Deaf studies talking.* Minneapolis: University of Minnesota Press.

Byrne, A. (2016). Sign language literature. In Gertz, G., Boudreault, P. (Eds.), *The SAGE Deaf studies encyclopedia*. London: SAGE Publishing, 832-835.

Burns, S.E. (1998). Irish Sign Language: Ireland's second minority language. In Lucas, C. (Ed.), *Pinky extension and eye gaze: Language use in Deaf communities*. Washington, DC: Gallaudet University Press, 233-273.

Castells, M. (1996). *The Rise of the Network Society, Vol. 1*. Oxford: Wiley-Blackwell.

Crasborn, O., Sloetjes, H. (2010). Using ELAN for annotating sign language corpora in a team setting. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, LREC 2010.*

De Vos, C., Pfau, R. (2015). Sign language typology: The contribution of rural sign languages. In *Annual Review of Linguistics*, 1, 265-288.

Fenlon, J., Schembri, A., Johnston, T. & Cormier, K. (2015). Documentary and corpus approaches to sign language research. In Orfanidou, E., Woll, B. & Morgan, G. (Eds.), *The Blackwell guide to research methods in sign language studies.* Oxford: Wiley-Blackwell, 156-172.

Frishberg, N., Hoiting, N. & Slobin, D.I. (2012). Transcription. In Pfau, R., Steinbach, M. & Woll, B. (Eds.), *Sign language. An international handbook.* Berlin: De Gruyter Mouton, 1045-1075.

Geraci, C., Battaglia, K., Cardinaletti, A., Cecchetto, C., Donati, C., Giudice, S. & Mereghetti, E. (2011). The LIS Corpus project: A discussion of sociolinguistic variation in the lexicon. In *Sign Language Studies*, 11(4), 528-574.

Gertz, G., Boudreault, P. (Eds.) (2016). *The SAGE Deaf studies encyclopedia*. London: SAGE Publishing.

Goodstein, H. (Ed.) (2006). *The Deaf Way II Reader*. Washington, DC: Gallaudet University Press.

Hauser, P.C., Paludneviciene, R., Riddle, W., Kurz, K.B., Emmorey, K. & Contreras, J. (2016). American Sign Language comprehension test: A tool for sign language researchers. In *The Journal of Deaf Studies and Deaf Education*, 21(1), 64-69.

Humphries, T., Kushalnagar, P., Mathur, G., Napoli, D.J., Rathmann, C. & Smith, S. (2019). Support for parents of deaf children: Common questions and informed, evidence-based answers. In *International Journal of Pediatric Otorhinolaryngology*, 118, 134-142.

Legg, J. (2016). Exploring the promise of digital deaf histories. In *Sign Language Studies*, 17(1), 42-58.

Leigh, I.W., Andrews, J.F. & Harris, R.L. (Eds.) (2018). *Deaf culture*. San Diego, CA: Plural Publishing.

Lucas, C., Bayley, R. & Valli, C. (2001). *Sociolinguistic variation in American Sign Language*. Washington, DC: Gallaudet University Press.

Marschark, M., Tang, G. & Knoors, H. (Eds.) (2014). *Bilingualism and bilingual deaf education*. Oxford: Oxford University Press.

Mann, W., Haug, T. (2014). Mapping out guidelines for the development and use of sign language assessments: Some critical issues, comments and suggestions. In Quinto-Pozos, D. (Ed.), *Multilingual aspects of signed language communication and disorder*. Bristol: Multilingual Matters, 123-139.

Mayberry, R.I., Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. In *Bilingualism: Language and Cognition*, 21(5), 886-905.

McCaskill, C., Lucas, C., Bayley, R. & Hill, J. (2011). *The hidden treasure of black ASL – its history and structure*. Washington, DC: Gallaudet University Press.

Monaghan, L., Schmaling, C., Nakamura, K. & Turner, G.H. (Eds.) (2003). *Many ways to be deaf: International variation in deaf communities*. Washington, DC: Gallaudet University Press.

Mougeon, R., Nadasdi, T. (1998). Sociolinguistic discontinuity in minority language communities. In *Language*, 74(1), 40-55.

Nyst, V. (2012). Shared sign languages. In Pfau, R., Steinbach, M. & Woll, B. (Eds.), *Sign language: An international handbook*. Berlin: De Gruyter Mouton, 552-574.

Padden, C.A., Humphries, T. (2005). *Inside deaf culture*. Cambridge, MA: Harvard University Press.

Parasnis, I. (Ed.) (1996). *Cultural and language diversity and the Deaf experience*. Cambridge: Cambridge University Press.

Perniss, P., Pfau, R., Steinbach, M. (2007). Can't you see the difference? Sources of variation in sign language structure. In Perniss, P., Pfau, R. & Steinbach, M. (Eds.), *Visible variation: Comparative studies on sign language structure*. Berlin: Mouton de Gruyter, 1-34.

Quer, J., Cecchetto, C., Donati, C., Geraci, C., Kelepir, M., Pfau, R. & Steinbach, M. (Eds.) (2017). *SignGram blueprint: A guide to sign language grammar writing*. Berlin: De Gruyter Mouton.

Quer, J., Steinbach, M. (2019). Handling sign language data: The impact of modality. In *Frontiers in Psychology,* 10:483. *doi: 10.3389/fpsyg.2019.00483.*

Rayman, J. (1999). Storytelling in the visual mode: A comparison of ASL and English. In Winston, E. (Ed.), *Storytelling & conversation: discourse in deaf communities*. Washington, DC: Gallaudet University Press, 59-82.

Rutherford, S. (1993). *A study of American Deaf folklore*. Burtonsville, MD: Linstok Press.

Sandler, W., Lillo-Martin, D. (2006). *Sign languages and linguistic universals*. Cambridge: Cambridge University Press.

Spencer, P.E., Marschark, M. (2003). Cochlear implants: issues and implications. In Marschark, M., Spencer, P.E. (Eds.), *Oxford handbook of deaf studies, language, and education*. Oxford: Oxford University Press, 434-448.

Supalla, T. (2001). Making historical sign language materials accessible: A prototype database of early ASL. In *Sign Language & Linguistics*, 4, 285-297.

Supalla, T., Clark, P. (2015). *Sign language archaeology*. Washington, DC: Gallaudet University Press.

Sutton-Spence, R. (2010). The role of sign language narratives in developing identity for deaf children. In *Journal of Folklore Research*, 47(3), 265-305.

Woll, B., Ladd, P. (2003). Deaf communities. In Marschark, M., Spencer, P.E. (Eds.), *Oxford handbook of deaf studies, language, and education*. Oxford: Oxford University Press, 151-163.

Zeshan, U., Palfreyman, N. (2017). Sign language typology. In Aikhenvald, A.Y., Dixon, R.M.W. (Eds.), *The Cambridge handbook of linguistic typology*. Cambridge: Cambridge University Press, 178-216.

Acknowledgments

This contribution has been made possible thanks to the SIGN-HUB project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 693349. Josep Quer's contribution is also made possible by the Spanish Ministry of Economy, Industry and Competitiveness and FEDER Funds (FFI2015-68594-P), and to the Government of the Generalitat de Catalunya (2017 SGR 1478).

DOI: 10.17469/O2106AISV000003

ANTONIO ROMANO, VALENTINA DE IACOVO

La base di dati "Tramontane": dati di parlato su lingue, dialetti, etnoletti e interletti del laboratorio di fonetica sperimentale "Arturo Genre"

In this contribution we discuss the consistency and the range of study of an oral archive based on the reading of the well-known Aesop's fable "The North wind and the Sun" in different languages and dialects. The database currently hosts more than 500 sound files and from 2017 it is available online as a platform or an interactive map with a new setting (http://www.lfsag.unito.it/ark/trm_index.html).

Key words: oral archive, dialects of Italy, endangered languages, language database.

1. *Storia delle tramontane*

In vari settori della linguistica e della dialettologia è diffusa la consuetudine di raccogliere campioni di parlato sulla base di uno stesso testo. La favola di Esopo "La tramontana e il sole" (usata per oltre un secolo dall'Associazione Fonetica Internazionale/International Phonetic Association AFI/IPA) si è oggi diffusa per illustrare molte lingue del mondo. Sebbene spesso in competizione con altri testi[1], la tradizione d'illustrare le caratteristiche di pronuncia di varie lingue (e, nel caso dell'IPA, la versatilità del sistema di trascrizione proposto), si è infatti affermata nel corso del Novecento e contribuisce alla descrizione della variazione in diversi spazi linguistici documentata in un certo numero di archivi[2].

Tra le fonti più autorevoli e/o di un certo impatto in Italia, ricordiamo il campione di produzioni regionali offerto da Tagliavini (1966) e il DOP - Dizionario di ortografia e pronunzia di Migliorini, Tagliavini e Fiorelli (1969)[3]. Come ricorda Romano (2016), ne abbiamo diverse attestazioni anche per varietà d'italiano e dialetti d'Italia, a partire dall'"Italiano" di A. Camilli, del 1911, oppure da "Italian - stage pronunciation" presente nella prima edizione dei Principles dell'IPA, del 1912, o, ancora, dall'"Italian (Pure Roman)" di S. Rivetta, del 1914 (autore nello stesso anno anche di quella del "Common dialect of Rome", del "Common dialect of Naples" e del "Common dialect of Turin"). Il dialetto di Servigliano (AP) figura

[1] Soprattutto in ambito romanzo è diffuso il ricorso al testo tradotto della "Parabola del figliol prodigo" (ma v., tra gli altri, Allières, 2000).
[2] Vedi http://www.lfsag.unito.it/ark/index.html (in costruzione).
[3] Dal 2008, vanta anche un'edizione on-line (http://www.dizionario.rai.it/) nella quale è disponibile il testo di varie versioni de "La tramontana e il sole" in lingue diverse. In un paio di casi, come in quello dell'italiano, i testi sono corredati da file sonori.

illustrato come uno dei "Dialetti marchigiani" che aveva intenzione di tratteggiare A. Camilli (a partire dal 1913)[4]. La tradizione continua con R. Invrea nel 1936, che propone "La pronuncia settentrionale dell'italiano" (in una versione intitolata "il vento e il sole") e con lo specimen di "Italian – Roman" (in realtà fiorentino) che appare nei Principles di IPA del 1949 (e nella ristampa del 1966). È poi la volta di L. Canepari, che nel 1970 propone una nuova illustrazione dell'"Italian pronunciation", migliorando la notazione di alcuni fenomeni e, infine, quella discutibile di Rogers e D'Arcangeli del 2004. Al di là della sua diffusione in pubblicazioni dell'IPA, il riferimento a centinaia di versioni de "La tramontana e il sole" ricostruite col metodo della fonetica naturale per lingue e dialetti diversi è poi disponibile in tutti gli ultimi lavori di L. Canepari (da Canepari, 1999 a Canepari, 2004a, 2004b e oltre). Oltre alle *Illustration* che continuano a essere pubblicate sui numeri del JIPA (Journal of the IPA) si può far riferimento anche alle decine di lingue illustrate in HIPA (1999). Una lista completa delle lingue illustrate fino al 2014 è disponibile in Olson (cfr. bibliografia). Recentemente l'affidabilità del testo è stata messa in discussione sul piano della sua adeguatezza a rappresentare un valido metodo di descrizione del sistema fonologico della lingua illustrata (Deterding, 2006, Jesus, Valente & Hall, 2015). Tuttavia, altri autori (Mairano, Zovato, 2018) hanno dimostrato che il bilanciamento fonologico può non essere l'unico parametro di valutazione e lo stesso archivio che qui si presenta ne mostra l'utilità su altri piani (qualità vocale, indici di costruzione testuale)[5].

2. *Le origini del progetto*

L'idea di allestire un archivio di registrazioni presso il LFSAG non è recente ma risale a più di una decina di anni fa quando, per poter testare alcune metriche ritmiche su campioni di parlato, si era scelto di utilizzare il testo della storia della tramontana e del sole e di farla quindi leggere a parlanti islandesi (Mairano, 2006) e, successivamente, a parlanti di lingue del mondo (Mairano, 2011). Nel 2017 viene svolto quindi un primo censimento che si conclude con la catalogazione di circa 300 file sonori[6], rinominati questa volta seguendo una nomenclatura fissa che tenesse conto della lingua, del punto d'inchiesta (o dell'origine geografica), del sesso e dell'eventuale bilinguismo del locutore. Interrogandosi su come sfruttare al meglio questi dati in un'ottica di conservazione e valorizzazione dell'archivio sonoro, si decide a

[4] Ancora nei *Fondamenti di Grafia Fonetica* del 1933, D. Jones e A. Camilli riproducono secondo il sistema dell'Associazione Fonetica Internazionale due versioni de "la tramontana e il sole" in romanesco e nel dialetto di Predappio (FO).
[5] Deterding (2006: 190-191) dedica un paragrafo ai problemi di misurazione del ritmo partendo da valutazioni fonologiche. Mairano (2011) e Romano, Mairano & Calabrò (2011) ne hanno mostrato invece l'utilità proprio attraverso la misurazione delle metriche ritmiche in considerazione delle reali modalità di enunciazione.
[6] Una prima descrizione della base dati è consultabile in De Iacovo, Romano (2017).

questo punto di organizzare una sezione apposita[7] sul sito del LFSAG[8] che permetta la consultazione dei dati attraverso una doppia modalità di fruizione. Arrivato nella pagina iniziale (Fig. 1), l'utente può scegliere la modalità di consultazione dei dati: una mappa geografica interattiva o un database. Nel primo caso (Fig. 2), la mappa può essere ingrandita così da scegliere un punto d'inchiesta tra quelli presenti (in arancione) che, cliccato, porta all'apertura di una piccola finestra *pop-up* nella quale vengono riportate le informazioni geografiche e linguistiche del dato, la barra di ascolto e l'eventuale trascrizione ortografica. L'accesso tramite lista (Fig. 3) permette invece di visualizzare l'elenco completo dei dati raccolti attraverso due macro-categorie, le lingue del mondo e i dialetti d'Italia. In ciascuna di queste i campioni sono organizzati per dominio linguistico, area (e suddivisioni dialettali eventuali) e località, con l'indicazione aggiuntiva di una valutazione sommaria sulla qualità del contenuto[9] e delle informazioni autoriali (crediti o riferimenti bibliografici da menzionare nel caso se ne faccia un uso in opere pubblicate). Ciascun file ha una durata media di 35-40 secondi. I dati, in costante aggiornamento, vengono organizzati attraverso un database SQLite[10] dal quale, in un secondo tempo, attraverso uno script in python, viene generata la pagina html. Essi vengono registrati in formato wav a 44100 Hz, si trovano online ricampionati a 16000 Hz e compressi in formato mp3 a 128 kbps e sono concessi sotto Licenza Creative Commons NC. Essi sono anonimizzati rispetto ai parlanti che li hanno prodotti[11] e sono a disposizione di chiunque ne faccia richiesta a condizione che il loro utilizzo avvenga per motivi esclusivamente culturali, di studio o di documentazione, senza fini di lucro e nel totale rispetto della normativa, impegnandosi a non effettuarne ulteriori copie e a citare il LFSAG come fonte depositaria dei documenti.

---

[7] La sezione "Archivi sonori" è in realtà dedicata ai diversi progetti (recenti e non) per i quali è stata predisposta la creazione di una base dati. Si veda, tra gli altri, il progetto "Le voci di VINCA" che raccoglie brani di parlato elicitato per immagini relative al progetto VINCA (Varietà di Italiano di Nativi – Corpus Appaiato) (Romano, De Iacovo, 2017), il progetto "Galileo G14-126" sul dialetto coratino delle comunità migranti dell'Isère in collaborazione con l'Università di Grenoble (Romano, De Iacovo, Bucci, Ronco & Carpitelli, 2016) o ancora la raccolta di proverbi griki di Calimera (LE).

[8] Dal sito del LFSAG (http://www.lfsag.unito.it) si può accedere alla sezione dedicata agli archivi sonori (http://www.lfsag.unito.it/ark/index.html) e quindi all'archivio sulla lettura del testo della tramontana e il sole (http://www.lfsag.unito.it/ark/trm_ index.html).

[9] La valutazione della qualità dell'audio è stata effettuata tenendo conto delle condizioni di registrazione (mediando su tre variabili: cabina silente = *** ÷ ambiente rumoroso = *; locutore disinvolto = *** ÷ impacciato = *; versione linguisticamente insindacabile = *** ÷ versione approvata, ma con interferenze, condizionamenti extralinguistici etc. = *). I file presentano mediamente una qualità medio-alta dovuta alla massima attenzione in fase di registrazione da parte degli intervistatori. Laddove il dato presenta invece una minor qualità, essa è da attribuire ad ambienti di registrazione difficilmente insonorizzabili (cfr. la versione munteno del romeno) o ad apparecchi ingrati (cfr. la versione di Paternò).

[10] La tabella funziona come una vera e propria banca dati in cui sono inserite eventuali informazioni non riportate nel html (come la data di avvenuta registrazione o l'eventuale bilinguismo del locutore).

[11] Il LFSAG ne dispone in seguito alle disposizioni di Ateneo che hanno definito una liberatoria nei confronti dei diritti di riservatezza dei donatori, ai sensi dell'art.13 D. Lgs. 30/06/2003 n. 196.

3. *I dati*

Attualmente la base dati contiene 528 dati etichettati, di cui 316 campioni di lingue del mondo e 212 dialetti d'Italia. Tra le lingue del mondo, meritano attenzione alcune lingue della famiglia bamileke (bangou, fefe, ghomalaa, medumba, ngeembo-on, ngomba, nguemba), niger-congo (igbo, kirundi, kituba, lingala, vili, wolof, twi, ditammari, baoulé) o niger-sahariana (gorane) e alcune varietà dialettali del serbo o l'apichat per il catalano. Oltre alla maggior parte di lingue presenti in Europa (italiano, inglese, francese, tedesco, spagnolo, portoghese, euskara, catalano, romeno, lussemburghese, ungherese, sloveno, ceco, polacco, olandese, albanese, bosniaco, serbo, macedone, croato, greco, lettone, lituano, islandese, finlandese, moldavo, russo, bielorusso, ucraino), troviamo diverse lingue mediorientali (arabo, turco, persiano dari e pašto, ebraico) e asiatiche come il cinese (anche dialetti locali registrati grazie agli studenti del progetto Turandot), il thailandese (cfr. Romano *et al.*, 2011), il tagalog, il singalese, il vietnamita, il coreano, il panjabi, l'hindi e il giapponese. Non mancano varietà del francese (belga, camerunense, congolese, canadese), dell'inglese (indiano, americano, australiano, sudafricano), dello spagnolo (Argentina, Colombia, Cile, Perù, Messico e Venezuela) e del portoghese d'oltremare (Brasile, Timor Est, Capoverde e Angola). Diversi sono inoltre i locutori bi/trilingue che hanno letto il testo in più varietà (ad esempio in spagnolo/galiziano/portoghese). Sono infine presenti alcune registrazioni di occitano parlato a Tolosa e di istrioto e istroveneto parlato a Dignano d'Istria (v. nn. in basso).

Per quanto riguarda invece la sezione sui dialetti parlati in Italia, va sottolineata la presenza di dati su alcune minoranze linguistiche (come l'arbëresh parlato in Calabria, il francoprovenzale arpitano e piemontese, il walser parlato in Valle d'Aosta e Piemonte, l'occitano parlato in Piemonte o il griko parlato nella Grecìa Salentina). Sono presenti anche alcune varietà come quella ligure tabarkina parlata a Carloforte (CA) e quella catalana parlata ad Alghero.

Alcune selezioni di dati sono già state oggetto di studi e approfondimento in diversi ambiti. Ad esempio, i campioni italiani, islandesi e di alcune altre lingue (come le lingue del sud-est asiatico) erano stati impiegati precedentemente per studi sulla caratterizzazione ritmica e intonativa del parlato in varie parlate[12]. Diversi studi sono stati condotti da tesisti[13], dottorandi o ricercatori; in altri casi invece sono state

---

[12] In particolare, Romano (2010).

[13] Una parte considerevole dei dati provengono da lavori di tesi come quelli sul cinese, l'islandese, il portoghese o l'occitano del Piemonte. Tra i principali contributi ricordiamo quelli delle Tesi di Laurea Magistrale (Opportune Mouti, 2005, Phonétique et tonétique du Ditammari: une langue voltaïque au contact du Français; Riccardo Mura, 2007, Analisi fonotonetica del portoghese atlantico e africano; Paolo Mairano, 2007, Rhythm assessment: a Case Study of Icelandic, Gregory Silaev, 2014, Analisi quantitativa della qualità della voce in parlanti bilingui russo-italiano; Gelsomina Giordano, 2015, Influenze del substrato occitano nella prosodia del francese parlato a Tolosa: uno studio pilota; Lorenzo Pirina, 2015, La lingua sarda. Studio storico-linguistico e analisi dei testi) e delle Dissertazioni di Laurea triennale (Maria Grazia Belci, 2006, Un'indagine sociolinguistica a Dignano d'Istria; Laura E. Capovilla, 2007, Lingue isosillabiche e isoaccentuali: misure strumentali su finlandese e tedesco; Marilena Dalmasso, 2009, Strutture tonali del kirundi: un contributo di analisi acustica; Navarole

condotte soltanto delle inchieste sul campo (come nel caso dell'occitano parlato a Tolosa o le varianti friulane). Gran parte delle registrazioni sono inoltre avvenute nella cabina silente del LFSAG, mentre in altri casi i dati sono stati registrati in ambienti esterni. Uno degli aspetti più spinosi riguarda la trascrizione propriamente ortografica dei testi, fornita soltanto per quelle lingue con una tradizione scritta consolidata. Nel caso di alcuni dialetti italiani, una trascrizione provvisoria è stata fornita sulla base dei criteri riassunti in Matranga (2007) e tenendo presenti le riflessioni proposte da Iannàccaro, Dell'Aquila (2008).

4. *Conclusioni e prospettive future*

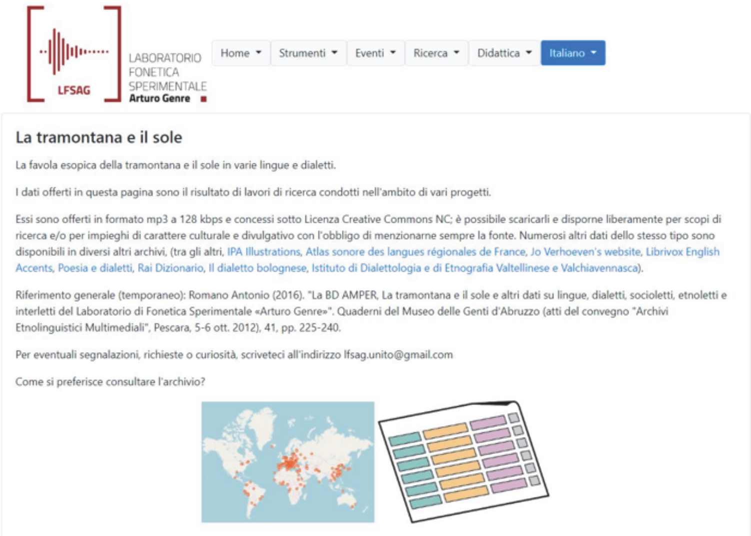
La creazione di un archivio sonoro rappresenta un progetto ambizioso se si pensa alle svariate applicazioni che questo permette di sviluppare. Da un lato esso rappresenta un solido corpus di dati sonori (ma non solo) sfruttabile dalla comunità scientifica ma anche patrimonio culturale poiché riflette la lingua in un determinato momento storico, attribuendole così un valore socioculturale. Dall'altro lato esso rappresenta uno strumento di divulgazione scientifica fondamentale per mantenere viva la memoria linguistica di quelle lingue e quei dialetti che verosimilmente non avranno più voce tra una decina d'anni. Il compito di chi si occupa della manutenzione di un archivio sonoro è quindi duplice: egli deve tener d'occhio la situazione geolinguistica che vive per poter testimoniare, attraverso il dato, l'evoluzione linguistica che avviene. Fermo restando che la base di dati è da considerarsi non esaustiva né rappresentativa di una lingua o di un dialetto, ma come singolo dato individuale (idiolettale), l'occasione sembra proficua per suggerire alcune possibilità ulteriori di sviluppo. In particolare, se da un lato i dati sono ora perfettamente compatibili e integrabili con quelli di altre raccolte, le opportunità di collaborazione e condivisione di risorse locali lasciano intravedere nuove prospettive d'uso. A questo proposito, recentemente il nostro progetto è stato coinvolto in uno simile riguardante i dialetti di Francia (a cura di Philippe Boula de Mareüil), il che ha permesso di dare maggior

---

Nguepedja Djouognwou, 2010, Analisi fonetica e tonetica del dialetto bangou del Bamiléké; Stojan Kamcev, 2013, Analisi acustica e fonologica del Macedone; Virginia Lamberti, 2014, I cambiamenti qualitativi della voce da una lingua all'altra: analisi quantitativa e studio spettrografico di caratteristiche fonatorie in realizzazioni di parlanti sinofoni in cinese e italiano; Matteo Bertoni, 2015, Liens linguistiques entre le corse et les parlers de l'Italie du nord; Alessandra Anglani, 2016, Peculiarità linguistiche e analisi dei trattamenti fonetici nel dialetto di Carovigno; Giacomo Gregorio, 2017, Resoconto d'inchiesta sull'occitano in alta Valle Varaita; Sara R. Decembrino, 2016, Il dialetto tra odio e amore. Lo studio linguistico della parlata di Manfredonia; Marco Ceppaglia, 2016, Il dialetto di Martina Franca (TA): analisi fonetica descrittiva; Viviana Cangelosi, 2016, Variabilità di alcuni fenomeni linguistici in Sicilia: un'indagine nei dialetti pollinese e palermitano; Deserti Erica, 2017, Un contributo sullo studio fonetico del dialetto di Quindici (AV); Giada Mantione, 2017, Analisi di alcuni fenomeni fonetici della Sicilia centrale: il caso di Canicattì; Luisa Giannino, 2017, Dialetti a confronto: Aspetti del mutamento del dialetto di Margherita di Savoia e Trinitapoli; Gaia M. Broodoo, 2017, Il morisyen: collocazione sociolinguistica e aspetti variazionali; Raouia El Mhadi, 2018, Analisi acustica e linguistica del sistema intonativo dell'arabo marocchino).

visibilità all'archivio. Diverse sono infine le migliorie da compiere: molte lingue e dialetti non sono stati ancora registrati e diverse trascrizioni dei testi sono da rivedere o uniformare. Infine, in una prospettiva più ampia, la sezione di questo archivio potrebbe confluire in una rete (inter)nazionale che coinvolga più progetti[14] (come ad esempio il consorzio CLARIN) così da uniformare l'accesso ai dati e accrescere l'accessibilità ai risultati della ricerca scientifica italiana.

Figura 1 - *Schermata della pagina iniziale della sezione "La tramontana e il sole"*



[14] Anche se sembra difficile pensare a un repository comune, i dati sono compatibili con quelli di: Ph. Boula de Mareüil, A. Rilliard, F. Vernier, Atlas sonore des langues régionales de France (https://atlas.limsi.fr/), J. Verhoeven, Phonetic Archive of the University of London (https://www.phonetics.expert/north-wind-and-the-sun), AA.VV., Librivox Dialect and Accent Collection Vol. 1 (https://archive.org/details/ dialect_accent_0909_librivox), D. Vitali, R. Cavicchi, F. Pizzirani, Il vento e il sole nei dialetti emiliani e romagnoli (http://www.bulgnais.com/ventoesole.html). Ovviamente, i nostri dati sono perfettamente compatibili anche con le registrazioni associate alle illustrazioni dell'IPA che, seppur accessibili soltanto ai soci dell'associazione (Sound Samples, v. https://www.internationalphoneticassociation.org/), sono quelle che hanno ispirato la raccolta.

Figura 2 - *Schermata della mappa interattiva della sezione "La tramontana e il sole"*

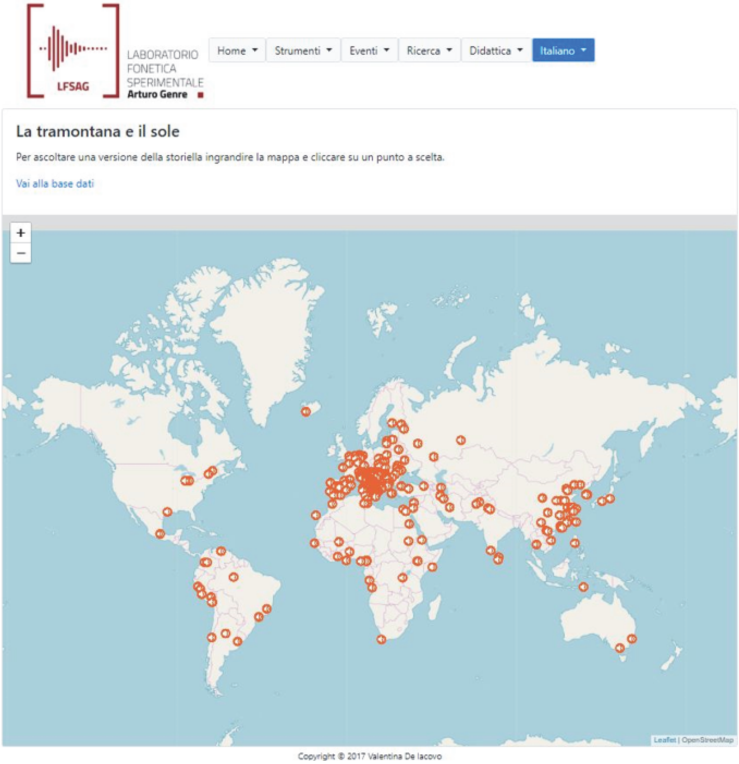
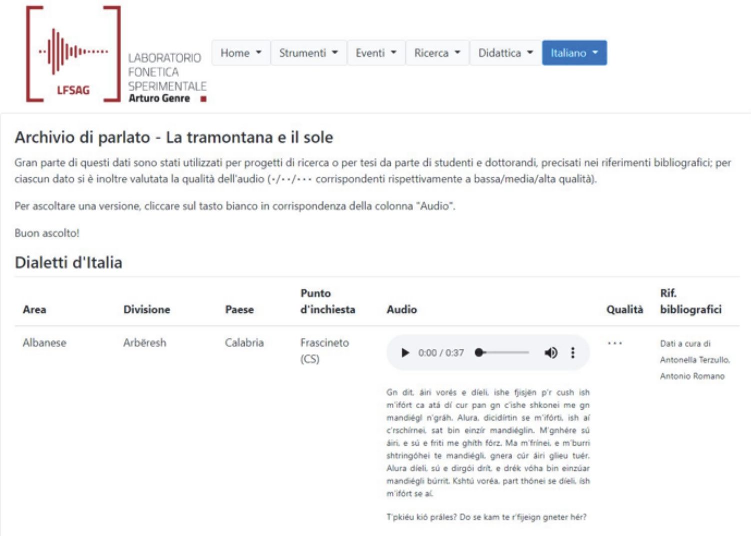


Figura 3 - *Schermata della base dati della sezione "La tramontana e il sole" per la sezione "Dialetti d'Italia"*



Bibliografia

Allières, J. (2000). *Les langues de l'Europe*. Paris: PUF.

Boula De Mareüil, Ph., De Iacovo, V., Romano, A., Vernier, F. (2019). Un atlante sonoro delle lingue di Francia e d'Italia: focus sulle parlate liguri. In Toso, F. (Eds.), *Il patrimonio linguistico storico della Liguria. Raccolta di studi*, III Serie - Dispensa, 163-174.

Canepari, L. (1999). *Manuale di Pronuncia Italiana (MaPI)*. Bologna: Zanichelli.

Canepari, L. (2004a). *Manuale di Pronuncia (ManPr)*. Monaco: Lincom Europa.

Canepari, L. (2004b). *Manuale di Fonetica (ManFo)*. Monaco: Lincom Europa.

De Iacovo, V., Romano, A. (2017). Tramontane, Archivio online di più di 300 campioni sonori di parlato letto/narrativo in centinaia di lingue e dialetti (http://www.lfsag.unito.it/ark/trm_index.html). (VD ha curato l'allestimento dei dati nel sito del LFSAG).

De Iacovo, V., Romano, A. (2019). Intonation of Sicilian among Southern Italo-romance dialects. Paper presented at *ICPHS 19*, University of Melbourne, August 5-9, 2019.

Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. In *Journal of the International Phonetic Association*, 36(2), 187-196.

HIPA (1999). *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Iannàccaro, G., Dell'Aquila, V. (2008). Per una tipologia dei sistemi di scrittura spontanei in area romanza. In *Estudis Romanics*, 30, 311-331.

Jesus, L., Valente, A. & Hall, A. (2015). Is the Portuguese version of the passage 'The North Wind and the Sun' phonetically balanced? In *Journal of the International Phonetic Association*, 45(1), 1-11.

Mairano, P., Zovato, E. (2018). Phonetically-balanced vs phonetically-rich: clarifying common misunderstandings and evaluating texts. Oral presentation at the XIV AISV Conference, 25-27 January 2018, Free University of Bozen (Italy).

Matranga, V. (2007). *Trascrivere. La rappresentazione del parlato nell'esperienza dell'Atlante Linguistico della Sicilia (con contributi di G. Paternostro e R. Sottile)*. Palermo: Centro di Studi Filologici e Linguistici Siciliani.

Migliorini, B., Tagliavini, C. & Fiorelli, P. (1969). *Dizionario di ortografia e pronunzia*. Roma-Torino: RAI ERI (ed. *online* http://www.dizionario.rai.it/, 2008).

Olson, K.S., Illustrations of the IPA: https://scholars.sil.org/kenneth_s_olson/ipa_ illustrations (Accessed 20/04/2019).

Pellegrini, G. (1977). *Carta dei dialetti d'Italia*. Pisa: Pacini.

Romano, A. (2010). Speech Rhythm and Timing: Structural Properties and Acoustic Correlates. In Schmid, S., Schwarzenbach, M. & Studer, D. (Eds.), *La dimensione temporale del parlato* (Atti di AISV2009, Università di Zurigo, Kollegiengebäude, 4-6 Febbraio 2009). Torriana (RN): EDK, 45-75.

Romano, A. (2016). La BD AMPER, La tramontana e il sole e altri dati su lingue, dialetti, socioletti, etnoletti e interletti del Laboratorio di Fonetica Sperimentale «Arturo Genre». In *Quaderni del Museo delle Genti d'Abruzzo*, 41, 225-240.

Romano, A. (2018). I contributi di Amerindo Camilli all'IPA e alla descrizione della pronuncia dell'italiano e delle sue varietà. In *Bollettino dell'Atlante Linguistico Italiano*, 42, 199-203.

Romano A., De Iacovo, V. (2017). Le voci di VINCA: riferimenti generali nelle produzioni di un campione di parlanti nativi. In Corino, E., Onesti, C. (Eds.), *Italiano d'apprendenti. Studi a partire dai corpora VALICO E VINCA*. Perugia: Guerra Edizioni, 131-155.

Romano, A., De Iacovo, V., Bucci, J., Ronco, G. & Carpitelli, E. (2016) Il dialetto di Corato (BA) in città e nelle comunità migranti (Isère, Francia). In *Bollettino dell'Atlante Linguistico Italiano*, III Serie - Dispensa, 163-174.

Romano, A., Mairano, P. & Calabrò, L. (2011). Measures of Speech Rhythm in East Asian Tonal Languages. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 17-21 August 2011, 1714-1717.

Romano, A., Mouti, O. (2006). Analisi preliminare delle strutture tonali del Ditammari (Benin). In Savy, R., Crocco, C. (Eds.), *Proceedings of the 2nd AISV Congress*, Salerno, 30-2 December 2005, 346-362.

Romano, A., Rivoira, M., Cugno, F., Ronco, G., De Iacovo, V. & Colonna, V. (forthcoming). Basi di dati dialettali e archivi vocali di lingue locali e minoritarie. In Brunetti, D., Robotti, D. & Salvalaggio, E. (Eds.), *I documenti sonori. Voce, suono, musica in archivi e raccolte*, no. Monografico di Archivi e biblioteche in Piemonte, 5.

Tagliavini, C. (1966). *La corretta pronuncia italiana: corso discografico di fonetica e ortoepia*. Bologna: Capitol (Dischi C.E.B.).

Ringraziamenti

Un ringraziamento particolare a tutti gli studenti, docenti e cultori locali che si sono spesi nelle registrazioni del testo.

DOI: 10.17469/O2106AISV000004

MANUELA SOLDI

L’archivio sonoro del Festivaletteratura di Mantova

The paper recalls the phases of preservation and promotion of the Festivaletteratura archive, in particular of the audio section. Festivaletteratura is one of the most famous and long-running literary festivals in Italy, born in 1997 in Mantua. In 2009 the Organizing Commitee opened its archive to the public and started the promotion of its heritage. Since 2016, the Archive has published online over two thousand audio recordings from 1998 to 2018.

Key words: Audio Archives, Digital Archives, Literary Festivals, Mantua, Literary Archives.

1. *Introduzione*[1]

In questo articolo si dà conto di un ormai decennale lavoro di riordino e valorizzazione della sezione sonora dell’Archivio del Festivaletteratura di Mantova, che negli ultimi anni è stato al centro della prioritaria attenzione dello staff dell’archivio e degli organizzatori della manifestazione al fine di permettere a un pubblico sempre più vasto l’accesso ai propri contenuti, attraverso la digitalizzazione, la descrizione e la pubblicazione on-line. Si darà inoltre conto dell’attuazione di strategie di valorizzazione che si sono concretizzate in azioni dentro e fuori il contesto del festival. Poter effettuare questa riflessione in occasione del *XV Convegno Nazionale AISV. Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale* (14-16 febbraio 2019, Dipartimento di Scienze della Formazione, Scienze umane e della Comunicazione interculturale, Arezzo) rappresenta un’ulteriore opportunità per diffondere gli effetti del lavoro svolto in questi anni e per intercettare un pubblico di studiosi in grado di osservare secondo canoni nuovi questo patrimonio di voci dalle caratteristiche omogenee e ben definite. Sarà tuttavia necessario, prima di concentrarsi sulla sezione sonora, descrivere il contesto di produzione e conservazione della documentazione in esso contenuta.

2. *Il soggetto produttore Festivaletteratura*

Per prima cosa è dunque necessario individuare quali siano le peculiarità del soggetto produttore di questo fondo e quelle della sua attività principale, ovvero l’organiz-

[1] Sia per la realizzazione del poster che per quanto riguarda la stesura dell’articolo, è necessario sottolineare il ruolo imprescindibile dello staff dell’archivio (Giorgia Nanni e Michele Triboli), dei consulenti grafici di Festivaletteratura (Pietro Corraini e Maria Chiara Zacchi), del consulente archivistico Danilo Craveia, della Segreteria Organizzativa (in particolar modo di Alessandro Della Casa) e del Comitato Organizzatore di Festivaletteratura.

zazione del Festivaletteratura di Mantova. Questa manifestazione celebra nel 2019 la sua 23a edizione, ed è oggi uno dei più longevi festival culturali italiani. Si svolge in cinque giorni (da mercoledì a domenica) nella prima o nella seconda settimana di settembre. Attualmente qualche centinaio di eventi (la maggior parte a pagamento, a cui si affianca un consistente numero di attività gratuite) si svolgono contemporaneamente in un elevato numero di *location* disseminate sul territorio della cittadina lombarda, spesso valorizzando luoghi difficilmente accessibili[2], animati da ospiti di caratura nazionale ed internazionale[3].

Nelle ultime edizioni le presenze di pubblico alla manifestazione si sono attestate intorno alle 120 mila annue. La composizione di queste presenze – per i dati in possesso dell'organizzazione – mostra un pubblico prevalentemente adulto e a maggioranza femminile. L'attenzione del festival è da sempre rivolta anche ai più giovani, sia attraverso una specifica programmazione rivolta a bambini e ragazzi all'interno di ogni edizione, sia attraverso l'elaborazione nel tempo di specifici progetti di promozione alla lettura che trovano spazio anche al di fuori dei cinque canonici giorni di settembre[4].

Il programma del festival è da sempre estremamente variegato, non si impone temi specifici ma cerca di rispecchiare le tendenze della discussione culturale e del panorama editoriale contemporaneo, proponendo al pubblico eventi che li rileggono grazie al contributo dei principali autori di narrativa e saggistica, in un dialogo informale che rappresenta fin dagli esordi la cifra della manifestazione.

La selezione di ospiti e temi non è affidata a un direttore artistico ma a un gruppo di lavoro allargato che annovera al proprio interno persone dagli interessi e dalla formazione molto differente, che a loro volta accolgono stimoli e segnalazioni di diversa provenienza, garantendo la multiformità e la varietà del programma.

2.1 Nascita e sviluppo di Festivaletteratura

Festivaletteratura è stato fondato da un comitato di volontari costituito da otto cittadini mantovani[5], tutt'oggi a capo della manifestazione, coadiuvato da una segreteria organizzativa che nel tempo ha accresciuto, compatibilmente con lo sviluppo del festival, la propria entità e il proprio ruolo[6].

La ragione di questa struttura risiede nella genesi della manifestazione, che ha le proprie radici nella Mantova nei primi anni Novanta, e in particolare nella stesura di

---

[2] Il numero è variabile, e nelle ultime edizioni è oscillato tra i trecentocinquanta e i quattrocento, mentre inizialmente erano un centinaio.

[3] Mediamente negli ultimi anni il numero di ospiti è stato di oltre trecento per edizione, dei quali circa settanta stranieri.

[4] Si ricorda il progetto *Qui comincia la lettura*, che tra 2006 e 2010 ha disseminato a cadenza annuale un romanzo nella provincia di Mantova chiedendo ai lettori di organizzare eventi che lo riguardassero, e il progetto di promozione della lettura tra i ragazzi *Read On*, sostenuto dall'Unione Europea, tutt'ora in corso.

[5] Il Comitato organizzatore è composto da: Laura Baccaglione, Carla Bernini, Annarosa Buttarelli, Francesco Caprini, Marzia Malerba, Luca Nicolini (presidente), Paolo Polettini, Gianni Tonelli.

[6] Da due componenti nel 1997 si è giunti a un gruppo di lavoro che comprende undici persone.

un *Piano strategico per la massimizzazione delle risorse culturali* (presentato nel settembre 1994), promosso dall'Osservatorio culturale della Regione Lombardia, che sceglie la società inglese Comedia per gestire lo studio pilota, sostenuto dal Comune e dalla Regione. Nell'ambito di questi lavori si forma un gruppo composto di funzionari di enti pubblici e di operatori culturali privati che affianca Comedia, che, fungendo da "cassa di risonanza", dovrebbe avere il compito di tradurre in modalità operative i risultati della ricerca. È in questo contesto che si individuano i problemi della gestione culturale della città e si identificano i potenziali punti di forza sulla base dei quali è possibile sviluppare nuove progettualità, tra le quali quella di fare di Mantova una "città del libro". È in questo contesto, ma in un momento successivo, che alcuni dei partecipanti ai lavori conoscono e visitano il festival gallese di Hay-on-Wye, avanzando l'ipotesi che questo format – se è lecito prendere in prestito il termine dal mondo televisivo – potesse essere riadattato, penetrando nelle piazze e nei palazzi cittadini e rispondendo alle esigenze mantovane.

Il progetto è in seguito realizzato al di fuori del contesto istituzionale, attraverso l'autonoma iniziativa di alcuni partecipanti ai lavori gestiti da Comedia, e l'istituzione di un comitato da parte loro (Polettini, 2012: 474). Il nuovo organismo fin da subito fa appello alla città per la realizzazione del festival. La risposta arriva sia in termini di fondi privati, che si affiancano a quelli provenienti dagli enti locali, sostenitori dell'evento, sia attraverso la nascita dell'Associazione Filofestival, che raduna ancor oggi i volontari del festival[7]. Prende così corpo la prima edizione, che ha luogo dal 10 al 14 settembre 1997 e attira subito l'attenzione di pubblico (circa 15 mila presenze), critica e addetti ai lavori[8].

La stampa sottolinea stupita la presenza di un folto pubblico pagante (Arbasino, 1999) e l'abbattimento delle barriere tra gli autori e il loro pubblico. Negli anni successivi il festival si svilupperà, ampliando notevolmente il numero degli eventi proposti (inizialmente un centinaio) e degli ospiti, proponendo nuove formule e attirando un numero crescente di spettatori, ma terrà fede all'impianto iniziale che, oltre alle caratteristiche già elencate, prevede il coinvolgimento degli ospiti in attività progettate ad hoc: non mere presentazioni dei libri ma creazione di contenuti nuovi, spesso scaturiti dal confronto inedito con altri autori.

Marzia Malerba, intervistata da Stefano Salis, interpreta quell'esplosivo successo come la risposta ad un bisogno latente, avvertito non solo da lei e dai suoi colleghi del comitato ma da molti altri, più di quanti loro stessi sospettassero: «Erano anni in cui dominava prepotentemente la tv. Noi abbiamo aggregato intorno al libro una comunità che fino a quel momento non riusciva a trovare un punto dove poter convergere» (Salis, 2017). Diffondendo questa formula, Festivaletteratura contribui-

---

[7] I volontari di Festivaletteratura hanno un'età minima di sedici anni e sono selezionati tramite una call che permette ai candidati di esprimere delle preferenze per la mansione che andranno a ricoprire. Durante le ultime edizioni il numero medio annuo dei volontari si è aggirato intorno ai settecento.

[8] Tra i presenti alla prima edizione: Roberto Calasso, Mario Luzi, Hahif Kureishi, Ian McEwan, Salman Rushdie, Emilio Tadini, Gianmaria Testa, Abraham B. Yehoshua.

sce all'affermazione del modello festival in Italia, ormai vero e proprio fenomeno di costume, come vedremo successivamente.

3. *L'Archivio di Festivaletteratura*

Nel 2009 il Comitato organizzatore di Festivaletteratura ha avviato una nuova avventura: l'allestimento di un archivio accessibile al pubblico. Una novità nel panorama dei festival letterari nazionali, che prevede una grande banca dati online dedicata alla letteratura contemporanea internazionale e al tempo stesso un luogo fisico all'interno della città per la fruizione di questo importante nucleo documentario, per rivivere la stessa atmosfera anche al di fuori del festival e apprezzarne nuovamente l'offerta culturale lungo il corso dell'anno.

3.1 Necessità e idealità: rendere accessibile il proprio archivio

Naturalmente sia il Comitato che l'Associazione Filofestival, in quanto soggetti giuridici, disponevano già di un archivio interno necessario al normale svolgimento dell'attività amministrativa e all'organizzazione artistica e logistica della manifestazione. Ma la rapida crescita di quest'ultima e il conseguente ampliamento dell'organigramma, sommati all'iniziale incertezza sul futuro del festival, a frequenti cambi di sede e alla natura volontaria di molte collaborazioni, non avevano favorito una consapevole ed ordinata sedimentazione della documentazione. A ciò si deve sommare il particolare frangente temporale di questa crescita, cioè il decennio tra la fine degli anni Novanta del Novecento e i primi anni Duemila, quando il digitale si fa strada velocemente nella vita personale e lavorativa di ognuno, inevitabilmente connesso a fenomeni di proliferazione documentaria incontrollata e ad una fideistica, quanto immotivata alla prova dei fatti, fiducia nella stabilità dei nuovi supporti per la conservazione a lungo termine (Zanni Rosiello, 2009; Vitali, 2004). La concomitanza di questi fattori alla fine degli anni Duemila rende urgente all'interno di Festivaletteratura un intervento volto a mettere in sicurezza la memoria di quanto era stato fino a quel momento, anche al fine di garantire l'efficienza e la rapida risposta alle sollecitazioni esterne di una struttura organizzativa sempre più complessa. A questi fattori pratici si aggiungeva certamente l'ideale coscienza, già forte all'interno dell'organizzazione, che il patrimonio documentario accumulato potesse rivestire un interesse anche oltre le mura degli uffici di Festivaletteratura, per tutti coloro che si occupano ad esempio di letteratura contemporanea o che studiano il fenomeno dei festival. La novità delle azioni messe in atto, oltre a una formazione archivistica, aveva sollecitato fin dagli esordi in alcuni componenti del comitato un'attenzione alla preservazione delle fonti.

I primi passi vengono mossi con l'avvio del progetto *Un archivio per l'innovazione. I giovani e i documenti di Festivaletteratura* sostenuto dall'ANCI, che lo inserisce nella Rete dei Festival aperti ai giovani, del Ministero della Gioventù e del Comune di Mantova. Il primo annuncio è dato durante l'edizione 2009 del festival (Festivaletteratura, 2009: 80), quando vengono proposti alcuni eventi che riflettono sul concetto di archivio al quale prendono parte diversi istituti. Durante i primi mesi del 2010 un corso di formazione ospitato presso l'Archivio di Stato di Mantova[9] coinvolge undici giovani aspiranti archivisti provenienti da tutta Italia, poi impegnati nell'allestimento dell'archivio, aperto al pubblico a settembre dello stesso anno, con un'operazione che secondo il comitato si muove tra "radicamento" e "scoperta" (Festivaletteratura, 2010: 2)[10]. Il concetto di "scoperta", connesso prevalentemente alla conoscenza del proprio patrimonio, che è diventato una base per una nuova produzione culturale, ha successivamente allargato il proprio significato. Includere all'interno del gruppo di lavoro professionalità archivistiche ha infatti notevolmente incrementato le iniziative e gli eventi dedicati a questo mondo nell'ambito del programma del festival[11].

3.2 Il patrimonio

Il cambio di paradigma nell'attività del Comitato organizzatore dovuto all'apertura dell'archivio ha contemplato anche la trasformazione dello stesso da promotore di un evento effimero a soggetto conservatore di un patrimonio. Un mutamento che ha avuto un impatto sull'attività ordinaria, fino a quel momento principalmente vocata ai cinque giorni di settembre e oggi sempre più impegnata a interagire con il contesto culturale nazionale e internazionale anche nel corso dell'anno e a riflettere sulle modalità di riattivazione del sedimento documentario di quei cinque giorni, che merita una descrizione più approfondita.

L'Archivio conserva innanzitutto la documentazione prodotta da due soggetti giuridici distinti: il Comitato organizzatore del Festival internazionale della letteratura in Mantova e l'Associazione Filofestival. Si tratta in entrambi i casi di fondi aperti dato che gli enti sono tutt'ora attivi. L'incremento è continuo, specie nel corso della manifestazione quando vi concorrono un largo numero di volontari. Una

[9] Tra il corpo docente operatori attivi in alcune tra le maggiori realtà archivistiche italiane, che hanno accompagnato i ragazzi in un percorso tra le varie tipologie di archivio e di documentazione. Si ricordano tra i soggetti coinvolti: CSAC – Centro studi archivio del contemporaneo dell'Università degli Studi di Parma, Centro di Documentazione RCS Mediagroup, Museo Alessi, Fondazione Zegna, Centro di ricerca sulla tradizione manoscritta di autori moderni e contemporanei dell'Università di Pavia, Archivio Ernesto de Martino, ICBSA-Istituto centrale per i beni sonori e audiovisivi, la Biblioteca Cantonale di Lugano, Fondazione Mondandori.

[10] In quest'occasione i partecipanti al corso prendono parte, nell'ambito dell'edizione 2010 di Festivaletteratura, a una riflessione sul rapporto tra archivi e letteratura con Carlo Lucarelli e Serena Vitale (Festivaletteratura, 2010).

[11] Oltre a eventi più classici dedicati a particolari archivi, si ricordano progetti di avvicinamento agli archivi in collaborazione con l'Archivio di Stato di Mantova, destinati a mostrare al grande pubblico come il patrimonio degli archivi riguardi ognuno di noi molto da vicino (*Genealogie*, 2013, sulla costruzione dell'albero genealogico attraverso la documentazione anagrafica; *I soldati negli archivi*, 2014, sulla ricostruzione delle storie dei soldati al fronte durante la prima guerra mondiale attraverso i ruoli matricolari; *La storia della mia casa*, 2015, sulla ricostruzione della storia delle abitazioni attraverso la documentazione catastale; *L'immagine della città,* sulla raccolta di immagini fotografiche della città contemporanea, 2016). Questi progetti sono stati presentati nel 2017 alla prima Conferenza italiana di Public History (Ravenna, 5-9 luglio).

peculiarità che nel tempo ha innescato riflessioni sulle modalità di gestione di una documentazione della quale i soggetti produttori principali non sono disponibili nelle fasi di ordinamento.

È conservata all'interno dell'archivio la documentazione amministrativa e relativa alla direzione creativa (circa cinquecento unità), oltre che la biblioteca di riferimento che comprende le edizioni di Festivaletteratura[12] – testi di autori intervenuti o proposti al festival[13] – a cui si affiancano i cospicui materiali inerenti la comunicazione. Quest'ultima è in parte afferente alle attività della segreteria come la vasta rassegna stampa, i gadget e i materiali promozionali, in parte alla redazione di volontari che fin dalle prime edizioni ha documentato il festival: oltre 190 mila fotografie e circa duemila ottocento video, un numero elevatissimo di testi comparsi sulle pagine del sito di Festivaletteratura e oggi di post sulle pagine social ufficiali del festival[14]. Sono inoltre presenti una collezione grafica, composta per lo più dai doni degli artisti e illustratori intervenuti al festival, e un importante patrimonio sonoro. Per questi motivi nel 2014 il patrimonio dell'archivio di Festivaletteratura è stato dichiarato dalla Sovrintendenza archivistica della Lombardia di notevole interesse culturale e numerosi partner hanno sostenuto il progetto di descrizione e valorizzazione[15].

Fin dal 2010 è incominciata un'attività di descrizione inventariale del fondo[16]. Dato che l'attività dei soggetti produttori, entrambi privati, non aveva sedimentato la documentazione secondo modalità condivise in un titolario, l'albero dell'archivio è stato ricostruito studiando la prassi quotidiana del lavoro degli uffici. Ne è risultato che la gran parte della documentazione è riconducibile all'organizzazione delle singole edizioni e/o degli eventi e dunque a una struttura "eventocentrica".

[12] I programmi delle edizioni, i dizionarietti bio-bibliografici *Centoautori* (editi dal 1998 al 2012), i periodici *Colibrì* e *Piazzalberti*, le antologie dedicate agli scrittori emergenti *Scritture giovani* e molte altre pubblicazioni inerenti a progetti specifici.

[13] In realtà la natura stessa dell'attività di scambio tra organizzatori, ospiti e moderatori degli eventi non favorisce la sedimentazione di un patrimonio librario, in quando i testi tendono a circolare. Inoltre una vasta parte della biblioteca raccolta a partire dal 1997 fu donata a metà degli anni Duemila, quando ancora non si pensava a un'autonoma e strutturata valorizzazione del patrimonio, a una biblioteca di quartiere.

[14] Se nel 1997 il compito di documentare il festival fu affidato a due fotografi volontari, nel tempo si è costituta una vera e propria redazione, che oggi conta circa ottanta elementi in attività per lo più durante i cinque giorni della manifestazione. Dal 2002 ai fotografi si sono affiancati gli operatori video. La provenienza dei membri della redazione è varia: molti dei più giovani sono studenti, ma prendono parte ai lavori anche professionisti della comunicazione (videomaker, social media manager, fotografi) che si mettono a disposizione come volontari. La redazione fruisce solitamente di una formazione interna all'avvio dei lavori e tende a condividere con l'organizzazione strumenti di lavoro personali dei suoi membri. Chi svolge mansioni di responsabile prende servizio alcuni giorni prima dell'avvio della manifestazione per organizzare il lavoro e gli spazi.

[15] Tra di essi Regione Lombardia, Fondazione Cariplo e Fondazione Cariverona.

[16] La descrizione è stata effettuata dal 2010 al 2018 mediante la piattaforma XDams-Regesta, mentre dal 2019 è stata effettuata la trasmigrazione alla piattaforma Archiui di Promemoria. Nel giugno 2018 è stato pubblicato il nuovo portale dell'Archivio di Festivaletteratura, gestito tramite la seconda piattaforma. Entrambe tengono conto dei principali standard di descrizione, in primis ISAD e ISAAR.

È del resto l'organizzazione dell'evento il punto focale dell'attività del comitato: è stato perciò inevitabile, all'avvio dei lavori, dedicarsi all'attenta descrizione dei programmi dei festival, dove ogni evento è trattato come un oggetto tangibile, che aggrega diversi elementi a loro volta protagonisti di schede dedicate (un luogo, degli ospiti, un sostenitore, una rassegna di riferimento) e produce contenuti descritti attraverso tag tematici che permettono un più facile accesso al pubblico. Se infatti l'approccio alla singola edizione, quella corrente, è auspicabilmente inteso dall'organizzazione come totalizzante, al fine di operare una costruzione consapevole del proprio percorso di spettatore, diversa è parsa la necessità dell'utente dell'archivio di interrogare una mole di dati più cospicua e complessa.

Solo una volta effettuato un primo censimento, la revisione dei dati da includere negli indici descrittivi (di persone, enti, luoghi, temi e rassegne) e degli eventi prodotti, si è avviato anche il lavoro sul patrimonio, proprio a partire dalla sezione sonora, ritenuta la più completa e appetibile per l'utenza.

La notifica di rilevante interesse storico, insieme ai lavori di catalogazione, hanno stimolato anche la pubblicazione di un'OPAC che permettesse al pubblico la fruizione in remoto dei materiali e degli inventari in fase di stesura. La prima versione del sito è stata pubblicata a settembre 2015, in concomitanza con la diciannovesima edizione del festival, in seguito viene annunciato l'avvio della pubblicazione online della sezione sonora dell'archivio[17].

3.3 La durevolezza dell'effimero: il patrimonio sonoro, ordinamento e descrizione

La nascita della sezione sonora dell'Archivio di Festivaletteratura risponde principalmente a un'esigenza organizzativa. Dopo la prima edizione è infatti apparso evidente che i tempi e le condizioni logistiche non permettevano agli organizzatori di essere presenti agli eventi e di valutarne la qualità e l'aderenza alle aspettative. Si decise pertanto di prevedere, dall'edizione successiva (1998), la registrazione integrale di una selezione di eventi, affidandola ai tecnici audio presenti sui luoghi. Negli anni seguenti la quota di registrazioni degli eventi aumenta sensibilmente: dal terzo circa iniziale si passa alla quasi totalità negli anni più recenti, dopo l'apertura dell'archivio[18]. Si giunge in questo modo a costituire, in ventitre edizioni, un patrimonio di oltre tremila registrazioni, di qualità ed estensione variabile[19].

L'archivio sonoro riflette in pieno la multiformità del festival, che ha prodotto oltre seimila eventi, esplorando i campi del sapere: dalla storia alle scienze, dalla musica alle nuove tecnologie, dalla narrativa all'ambiente, con un particolare riguardo per le giovani generazioni.

---

[17] La sua realizzazione è frutto del lavoro congiunto di Festivaletteratura con la società romana Regesta.exe, che cura la piattaforma XDams, Studio Indaco (Mantova), che segue i siti internet connessi alla manifestazione, Pietro Corraini e Maria Chiara Zacchi per il progetto grafico.

[18] Non vengono registrati eventi laboratoriali per bambini e ragazzi, eventi musicali e teatrali.

[19] I fattori che influenzano la qualità dell'audio sono principalmente il rumore ambientale (molti eventi si tengono in luoghi aperti) e la dimestichezza degli ospiti con i microfoni. La durata media delle registrazioni ricalca quella integrale dell'evento è di poco più di un'ora.

Il confronto con la sezione audiovisiva risulta particolarmente interessante. Da un lato le registrazioni sonore sono realizzate da professionisti con apparecchiature professionali, sebbene la rilevazione dei contenuti dell'evento non nasca con finalità di pubblicazione, prassi instaurata solo recentemente. Dall'altro la documentazione audiovisiva[20], realizzata da un nucleo di volontari, è concepita con l'esplicito fine di pubblicazione da un gruppo di lavoro non necessariamente costituito da professionisti e con strumentazione di buona qualità ma non sempre professionale. Osservazioni che valgono sia per le riprese che per i montaggi che, infine, per la preparazione del crescente numero di interviste video effettuate negli ultimi anni.

All'avvio dei lavori per l'allestimento dell'Archivio, la documentazione sonora rinvenuta si presentava su diversi supporti. I materiali attinenti alle edizioni 1998-2002 sono conservati su audiocassette, per lo più in doppia copia[21]. Ogni nastro presenta un'etichetta adesiva riportante solitamente il nome dell'autore, il titolo e il numero progressivo che distingue l'evento all'interno del programma della manifestazione. Dal 2003 in poi le registrazioni sono state consegnate dai fonici in formato per lo più .mp3 e conservate dalla segreteria prima su cd e poi sul server dell'ufficio. La documentazione nativa digitale presenta problematiche tipiche in questi casi: ridondanza di copie e nomenclatura non parlante, mentre solo in rarissimi casi si sono rilevati problemi di lettura dei files. Fino a quel momento non si era posta una grande attenzione al riordino e al recupero di queste fonti, oggetto di ascolti estemporanei per verificare l'andamento del festival o per esigenze di comunicazione. Dal 2010 in poi il progetto dell'archivio ha determinato una nuova presa di coscienza, che ha previsto l'emanazione di linee guida ai fonici per quanto riguarda la rilevazione, il formato digitale (è stato scelto il .wav, più adatto alla conservazione) e la riconsegna dei materiali.

Si è poi proceduto, con la consulenza del personale dell'ICBSA, a predisporre una rudimentale struttura per la digitalizzazione dei nastri relativi alle prime edizioni. Grazie al sostegno di Regione Lombardia tra 2012 e 2013 ne sono stati digitalizzati circa centoventi, mentre i restanti attendono un nuovo intervento.

Nel frattempo, si è messa a punto una scheda che ha definito come unità descrittiva la registrazione di un singolo evento, a prescindere dal suo supporto (analogico, digitale, o in alcuni casi entrambi quando la registrazione è stata digitalizzata). Le registrazioni sono state ordinate in serie cronologiche per edizione. Tali edizioni non racchiudono soltanto gli eventi avvenuti durante i cinque giorni di settembre,

---

[20] Per un'idea della produzione audiovisiva di Festivaletteratura, si veda il canale Youtube ufficiale della manifestazione, dove sono pubblicati i montaggi effettuati dalla redazione. Se nei primi anni si trattava prevalentemente di filmati di 3 minuti circa che condensavano i contenuti di un singolo evento, oggi si dà la precedenza all'intervista e ai filmati dal taglio tematico. Riprese di singoli eventi sono comunque tutt'ora effettuate ad uso prevalentemente interno, non si tratta però di una copertura totale del festival, affidata alla registrazione sonora, meno impegnativa dal punto di vista delle risorse da mettere in campo. La selezione degli eventi da filmare è solitamente effettuata in base alla novità della presenza di un'ospite o dal particolare peso dell'elemento visivo che ne caratterizza alcuni, tale da rendere meno efficace la registrazione sonora.

[21] Le audiocassette conservate presso l'Archivio di Festivaletteratura sono circa trecento cinquanta.

ma anche quelli organizzati durante l'anno nel contesto della promozione del festival e dell'attività dell'Associazione Filofestival[22]. I primi test di catalogazione sono stati effettuati sulle registrazioni relative alle edizioni più recenti, sedimentatesi in maniera più ordinata; in seguito il lavoro di descrizione si è occupato dei materiali risalenti alle prime edizioni, che ponevano maggiori problemi. Una volta avviata la descrizione, si è posto come urgente il tema della valorizzazione, che ha determinato in primo luogo la fruizione interna del patrimonio sonoro e un'ulteriore acquisizione di consapevolezza da parte della struttura della propria peculiarità.

3.4 Non solo un format

Quando si parla di festival non si non si parla solo di un format: l'esame della sezione sonora dell'archivio ha reso infatti possibile osservare il consolidarsi, di edizione in edizione, di un insieme di atteggiamenti ed espressioni di veicolazione della cultura tipici di questo contesto, che si allontanano dall'accademia per relazionarsi in modo più aperto e diretto con il pubblico, senza per questo abbassare troppo il livello complessivo della comunicazione. Gli eventi del festival costituiscono dunque un patrimonio di voci di intellettuali che raccontano la cultura e la società degli ultimi vent'anni, spesso anticipando temi successivamente divenuti cruciali, in altri casi portando tesi che oggi possiedono un notevole valore storico per la ricostruzione del dibattito intorno alle più varie questioni. Voci autorevoli che parlano in un contesto informale e avvertito come protetto, talvolta confrontandosi su problematiche che esulano dai loro abituali campi d'azione.

Un modello di divulgazione che ha trovato recentemente una definizione anche nel corso di formazione *Il conversatore letterario* (Mantova, 23-24 febbraio 2019), dedicato all'approfondimento delle modalità dell'intervista letteraria e della gestione del dialogo tra autore e pubblico (*Il conversatore*, 2019)[23]. I lavori preliminari all'allestimento della didattica, hanno dimostrato ancora una volta la fondamentale importanza di fare auto-memoria della propria attività grazie alla presenza di un archivio accessibile e ordinato. L'analisi degli interventi precedentemente tenuti al festival da professionisti del settore si è infatti rivelata in questo caso un potente strumento didattico.

3.5 Dar voce alle voci, la valorizzazione all'interno del festival e nel mondo digitale

Le modalità di fruizione e valorizzazione dell'archivio e della sua sezione sonora in particolare sono state oggetto di costante riflessione all'interno dell'organizzazione,

[22] La cronologia dell'anno del festival si avvia in autunno con l'apertura del tesseramento dell'Associazione Filofestival. Nei mesi successivi diversi appuntamenti mantengono vivo il legame con pubblico e volontari, culminando nella presentazione in città della lista degli ospiti dell'edizione successiva (a fine giugno) e con la distribuzione delle prime copie del programma in Piazza Alberti (luogo dove si tenne la prima assemblea pubblica che diede vita al festival), oggi contestuale alla pubblicazione online dello stesso (a fine luglio).

[23] Tra i docenti: Simonetta Bitasi, Laura Cangemi, Peter Florence, Fabio Geda, Gaia Manzini, Gabriele Romagnoli, Laura Torelli, Alessandro Zaccuri.

che della riscoperta questo patrimonio sempre più vasto e significativo ha fatto un punto di partenza per nuove elaborazioni culturali, un percorso ormai quasi decennale che pone sfide continue per mantenerlo vivo.

Nel 2016, l'incipiente ventennale del festival ha fornito l'occasione per avviare la pubblicazione online di una vasta selezione della sezione sonora dell'archivio, contravvenendo al progetto iniziale che prevedeva l'esposizione dei soli inventari senza alcun allegato. Il comitato ha deciso di festeggiare in questo modo il traguardo raggiunto: il nuovo progetto prevedeva infatti il rilascio in diverse fasi delle registrazioni riguardanti una singola edizione, dalle più recenti scendendo via via fino al 1998[24]. Da allora in poi, il database viene costantemente aggiornato con le registrazioni relative all'edizione più recente, pubblicate non appena terminata la descrizione archivistica, solitamente nei primi mesi dell'anno successivo.

È stata contemporaneamente avviata una campagna di verifica e recupero dei permessi di pubblicazione, poiché in molti casi le registrazioni erano precedenti all'idea di realizzare un archivio online. È stata poi valutata la qualità delle stesse, e si è optato per la pubblicazione senza un preliminare lavoro di editing, dato che nella maggior parte dei casi i contenuti sono perfettamente intellegibili.

Fin dall'avvio della campagna, la fruizione è prevista anche tramite il portale Soundcloud, raggiungendo così fasce di pubblico distanti da quelle del festival ma potenzialmente interessate alla fruizione di materiale sonoro.

Sono online in streaming on demand attualmente oltre duemila e trecento file, che aumenteranno ulteriormente attraverso la diffusione di quelli relativi all'edizione 2019. Dal 2017 è stato avviata anche la fruizione in modalità podcast, che prevede la pubblicazione periodica di registrazioni selezionate. L'archivio sonoro può dunque essere fruito attraverso canali diversi. Se si sceglie l'accesso da OPAC è possibile interrogare il database da diversi punti di vista: in primis quello tematico ma anche, per esempio, guardando sulla ricchezza linguistica, un punto di partenza meno scontato ma possibile essendo presenti registrazioni in diverse lingue estere[25], che rendono appetibile il nucleo di registrazioni anche dal punto di vista della didattica delle lingue e permettono l'accesso di utenti stranieri.

A prescindere dalla pubblicazione online, la sezione sonora è stata in realtà valorizzata fin dall'apertura dell'archivio. La prima modalità individuata è stata quella dell'installazione sonora, che nell'ambito dei cinque giorni del festival, quando ancora il pubblico non poteva fruire liberamente della gran parte del fondo attraverso i propri dispositivi, metteva a disposizione una selezione di registrazioni con un taglio tematico ben preciso, solitamente connesso ad aspetti della contempora-

[24] La pubblicazione è stata avviata nel mese di maggio 2016 ed è continuata fino al settembre 2017.

[25] Se da alcune edizioni un numero limitato di eventi in lingua straniera è stato offerto al pubblico senza alcuna intermediazione, fin dal 1997 la presenza di ospiti stranieri prevede una traduzione consecutiva. Questo significa che nella registrazione la voce dell'ospite straniero e quella dell'interprete non si sovrappongono mai, risultando entrambe perfettamente fruibili. Le lingue più rappresentate dopo l'italiano sono l'inglese (circa settecento eventi), il francese (circa duecento eventi) lo spagnolo (circa cento eventi), il tedesco (circa settanta eventi).

neità che trovavano riscontro nel programma. Le installazioni, fruibili liberamente durante l'intero festival, trovavano solitamente dei testimonial in ospiti della manifestazione esperti della materia[26]. Recentemente la modalità dell'installazione è stata ripresa in *Archivio Ambiente* (2018), con un taglio leggermente diverso. Il progetto riprendeva, in occasione dell'imminente World Forum on Urban Forests, tenuto a Mantova poco dopo, alcuni eventi promossi nell'ambito della rassegna *Consapevolezza verde*, da anni parte del programma del festival, suddividendoli in quattro sottocategorie tematiche. Con le registrazioni ormai pubblicate online, non appariva particolarmente accattivante l'idea di metterle a disposizione integralmente, esattamente come si trovano in rete, anche per la grande mole di materiale dedicato al tema disponibile in archivio. Si è optato dunque per la realizzazione di una "foresta" dove gli utenti potevano fruire di estratti audio, scelti da esperti chiamati a commentare le registrazioni.

Questa nuova installazione risente di altre esperienze effettuate nel lustro precedente, che hanno insistito sull'interazione tra patrimonio dell'archivio e ospiti del festival. La principale è stata *Giocando al Fantafestival*, rassegna semiseria destinata a festeggiare il ventennale del festival, avviata durante l'edizione 2015 e continuata nei mesi successivi, dove alcuni autori che frequentano il festival fin dalle prime edizioni, sulla base della loro memoria personale e della consultazione del neonato OPAC, allestivano la loro ideale giornata del festival, inanellando gli eventi e gli autori secondo loro più significativi delle precedenti edizioni[27]. La stessa richiesta, sia attraverso le pagine social del festival che il programma, è stata avanzata al pubblico, invitato a presentare le proprie *playlist*, per stimolare i primi accessi all'OPAC[28].

Più recentemente il coinvolgimento degli autori è continuato con la rassegna *Nelle puntate precedenti* (2017) che ha previsto un approccio meno libero al patrimonio, orientando a temi particolari, connessi a quelli trattati da altre rassegne in programma durante quell'edizione, la scelta degli eventi che gli autori sono chiamati a discutere, fornendo così spunti di ulteriore approfondimento al pubblico.

La valorizzazione dell'archivio sonoro non ha però battuto solo la strada dell'evento effimero nell'ambito del festival, cercando anche altri canali.

In primo luogo ha fornito le basi per alcuni volumi dedicati a progetti che avevano trovano concretezza nell'ambito del festival. Il primo è stato *Diario Flaiano* (2011), risultato di *Biblioteca Flaiano,* che ha previsto il riallestimento a Mantova, durante i giorni del festival, in collaborazione con i sistemi bibliotecari della provincia di Mantova e la Biblioteca Cantonale di Lugano, della biblioteca di studio di Ennio Flaiano (Festivaletteratura 2010). Il fondo è stato oggetto di alcune "visite

[26] Si ricordano: *Voci dal mondo arabo* (2011), che riprende gli eventi di autori arabi in un anno in cui grande attenzione è tributata alla cosiddetta "primavera araba"; *Dai nostri inviati* (2012), che si concentra in particolare sulle voci dei reporter; *Latinoamericana* (2013), dedicata agli scrittori latinoamericani in un anno nel quale il festival effettuava un focus su di loro.

[27] Sono coinvolti: Massimo Cirri, Matteo Corradini, Lella Costa, Bruno Gambarotta, Beppe Finessi, Bianca Pitzorno, Luca Molinari, Beppe Severgnini.

[28] Dall'avvio della pubblicazione delle registrazioni sonore (maggio 2016) al convegno AISV gli ascolti sono stati circa 35 mila.

guidate" da parte di esperti ed appassionati della figura di Flaiano (Anna Bolzoni, Bruno Gambarotta, Hans Tuzzi, Marcello Veneziani). Proprio le registrazioni di queste visite guidate, editate con la collaborazione degli autori, sono andate a costituire i quattro capitoli del volume. Più recentemente l'archivio ha fornito i materiali per la pubblicazione che ha coronato il progetto *Vocabolario europeo* (2008-2017), che prevedeva la donazione al festival, da parte di alcuni ospiti, di un lemma a loro particolarmente caro nella loro lingua per un ideale vocabolario europeo. L'archivio sonoro ha sopperito in particolare la mancanza di testi scritti forniti dagli autori per i primi anni della rassegna, permettendo di realizzare un volume che la testimoniasse nella sua completezza.

Un'altra modalità sperimentata è stata l'interazione tra testo teatrale e registrazione audio in una serie di spettacoli teatrali realizzati con la regia di Luciano Minerva e Silvano Piccardi, omaggi ad alcuni ospiti del passato come Anna Politkovskaja e Tiziano Terzani (*Anna, il coraggio di dire* e *Terzani. Parole contro la guerra*, Festivaletteratura 2012), al protagonista della Retrospettiva di Festivaletteratura 2013 David Grossman (*Percorsi nel festival: su David Grossman*). Di interazione tra audio e fotografia si è trattato invece nel caso della mostra *Scatti d'autore* in *Memorandum. Festival di fotografia storica* 2014 (Biella-Torino, 2014), dove ai ritratti di autori scattati dai volontari del festival si accompagnavano trascrizioni di estratti dalle registrazioni sonore.

Molto soddisfacente anche la valorizzazione attraverso il mezzo radiofonico, attraverso il ciclo *Voce alle voci* (2016) prodotto e curato dall'emittente radiofonica locale Radiobase che da sempre seguiva la manifestazione sia durante il suo svolgimento sia attraverso un percorso di avvicinamento nei giorni precedenti. In occasione del ventennale del festival alle attività usuali si è aggiunta questa produzione, che in dieci episodi tematici esplorava l'archivio sonoro mettendolo in relazione con tematiche trattate nell'edizione in corso e inserendo gli ascolti di alcuni estratti all'interno delle puntate, poi rimaste disponibili sul sito dell'emittente.

4. *Conclusioni*

Trarre delle conclusioni su un lavoro in itinere è difficile. Il progetto archivistico descritto in questo articolo presentava certamente ai suoi albori, nel 2009, caratteristiche pionieristiche e per certi versi la valorizzazione dell'archivio rappresenta ancora oggi una prassi non consolidata in toto per manifestazioni di questo tipo. Lo ha dimostrato anche il tavolo di lavoro dedicato agli archivi dei festival nell'ambito di *Archivissima*, primo festival degli archivi italiano (Torino, Biblioteca civica centrale, 7 giugno 2018)[29]. Tra festival letterari italiani presenti, solo tre erano dotati di un archivio strutturato: oltre a Festivaletteratura, il Festival della Mente di

[29] Hanno partecipato: Manuela Soldi per Festivaletteratura, Alessandro Lombardo per il Festival della Psicologia, Benedetta Marietti per il Festival della Mente di Sarzana, Francesca Bellotto per Dialoghi sull'uomo di Pistoia, Davide Camarrone per il Festival delle letterature migranti di Palermo, Gianmaria Pilo per La Grande Invasione di Ivrea.

Sarzana e il Festival Dialoghi sull'uomo di Pistoia[30], che nascono però in ambito pubblico e condividono fin dalle origini le buone prassi archivistiche della pubblica amministrazione. È chiaro che, dato il proliferare nell'Italia degli ultimi vent'anni di manifestazioni che assumono più o meno indistintamente la denominazione di festival, pur essendo caratterizzate da longevità, promotori e livelli di qualità molto diversi[31], è difficile avere un quadro chiaro e aggiornato del fenomeno, pur studiato e definito in maniera scientifica (Guerzoni, 2008: 79-80)[32], figurarsi del trattamento che viene riservato alla documentazione prodotta nell'ambito dei festival stessi, nonostante il problema abbia già attirato l'attenzione di alcuni studiosi e in particolare di Guerzoni che sottolinea come dal 2011 i festival italiani abbiano iniziato a diffondere online una grande quantità di video e fotografie prima realizzati con finalità archivistiche e poi utilizzati in maniera più consapevole per la propria comunicazione.

Esistono tuttavia festival accostabili per entità del programma e longevità a quello mantovano, che danno oggi la possibilità al pubblico di accedere a informazioni e dati relativi alle precedenti edizioni. Il Festival della mente di Sarzana e Dialoghi sull'uomo, ad esempio, mettono sistematicamente a disposizione le registrazioni sonore e audiovisive. Altri sostanzialmente mantengono in vita le pagine web allestite per la loro comunicazione, contenenti la descrizione del programma, delle location e le biografie degli ospiti, dove sono talvolta accessibili anche selezioni fotografiche e video. A soluzioni di questo tipo sono pervenuti per l'appunto ad esempio Festival della filosofia di Modena, Carpi e Sassuolo e il festival Pordenone legge. Altri ancora si limitano a raccogliere in una pagina le principali informazioni relative alle precedenti edizioni, come l'International festival of Journalism di Perugia. Non risultano però ad oggi accessibili sui siti di questi festival funzionalità di ricerca avanzata che diano la possibilità all'utente di interrogare la descrizione di questi importanti patrimoni.

Queste operazioni hanno in comune con quella del festival mantovano – che si trova impegnato nel riordino di un patrimonio quantitativamente più cospicuo dato il maggior numero di anni di attività – la necessità di creare una struttura logica che guarda al festival come a un patrimonio immateriale frutto di una comunità più o meno allargata, ma tangibile attraverso le sedimentazioni documentarie. Un approccio che può forse aggirare il problema della mancanza delle "attività culturali" all'interno della definizione di bene culturale presente nel nostro Codice dei beni culturali[33] (Guerzoni, 2012: 1-2), di fatto superata con la notifica di interesse culturale all'Archivio di Festivaletteratura avvenuta nel 2014 da parte della Sovrintendenza archivistica della Lombardia.

---

[30] Il Festival della Mente di Sarzana si tiene nella cittadina toscana dal 2004 tra la fine di agosto e i primi di settembre e nel tempo è passato da circa venti a sessanta eventi. Dialoghi sull'uomo si tiene invece a Pistoia, a maggio, da dieci anni e promuove tra i venti e venticinque eventi per edizione.

[31] A supporto di questa proliferazione, si ricorda anche la pubblicazione di alcune guide volte a favorire l'orientamento del pubblico in questo panorama, come quelle dell'editore Morellini.

[32] Guido Guerzoni in particolare avvia un'analisi scientifica del fenomeno italiano nell'ambito delle sue ricerche Effetto festival, dal 2008 in poi.

[33] D. Lsg. 22 gennaio 2004, n. 42, Art. 10.

Il progetto archivistico del festival di Mantova differisce tecnicamente da quelli a cui abbiamo fin qui accennato perché non si limita a descrivere l'evento collegando ad esso le sue prove documentarie, ma considera e descrive come entità a sé stanti tali documenti, allargando a dismisura le possibilità di descrizione e di ricerca per diverse tipologie di pubblico.

L'archivio di Festivaletteratura, così come gli altri archivi e in particolare quelli di enti ancora attivi, è un'entità in continuo movimento, che si arricchisce di nuovi significati e pone nuove sfide ogni giorno, in primis quella di una conservazione a tempo indeterminato nell'era della smaterializzazione, nonché quella più prosaica ma egualmente strategica del reperimento delle risorse necessarie a garantirla.

Quello afferente a Festivaletteratura è un nucleo documentario che si pone all'attenzione di diverse discipline, anche per la particolare natura dei soggetti produttori: da un lato un'associazione di volontariato, dall'altra un archivio letterario che non basa la sua esistenza sulla parola scritta, ma su quella detta, poiché il produttore non si dedica all'attività letteraria in senso stretto ma la promuove per lo più attraverso l'organizzazione di eventi effimeri.

Il riordino e la catalogazione in atto hanno certamente determinato all'interno della struttura organizzativa un'attenzione sempre maggiore a governare i flussi di lavoro e – per il sonoro in particolare – la consapevolezza profonda del valore delle "parole dette" che costituiscono l'archivio.

Un lavoro continuo di descrizione che lo renda sempre più accessibile a dei fruitori diversi dal soggetto produttore potrà chiarire ulteriormente possibili percorsi di ricerca e identificare nuove potenzialità.

Bibliografia

Arbasino, A. (1999). Incontra lo scrittore e paga il biglietto. In *La Repubblica*, 11 settembre 1999.

Archivissima. Festival degli archivi (2018). *Programma*.

Baldini, P., Carbè, E. & Weston, P.G. (2017). Se i bit non bastano: pratiche di conservazione del contesto di origine per gli archivi letterari nativi digitali. In *Bibliothecae.it*, 6, 1, 155-177.

Antonelli A., Motolese M. (Eds.) (2017). *Vocabolario Europero. Le parole (d)agli autori 2008-2017*. Mantova: Corraini.

Archivio Festivaletteratura (Ed.) (2011). *Diario Flaiano*. Mantova: Corraini.

Craveia, D., Soldi, M. (2014). Scatti d'autore. In *Memorandum - festival di fotografia storica 2014 (catalogo delle mostre)*. Biella: E20 Progetti, 176-189.

Damiani, C. (2019). La memoria rappresentata: dalla descrizione inventariale agli archivi narranti. In *Officina della storia*, 9 gennaio. https://www.officinadellastoria.eu/it/2019/01/09/la-memoria-rappresentata-dalla-descrizione-inventariale-agli-archivi-narranti/ Accessed 20.04.19.

Feliciati, P., Allegrezza, S., Belisario E., Vassallo, V., Carbé, E., Baldini, P. & Weston, P.G. (2018). Archivi digitali d'autore: ipotesi di lavoro. In Spampinato, D.

(Ed.), *Patrimoni culturali nell'era digitale. Memorie, culture umanistiche e tecnologia. Books of abstracts*. Settimo Convegno Annuale AIUCD 2018, Bari, 31 gennaio - 2 febbraio 2018, 256-259.

Festivaletteratura (2009). *Programma*.

Festivaletteratura (2010). *Programma*.

Festivaletteratura (2016). *Programma*.

Frezza, G. (Ed.) (2008). *L'arca futura. Archivi mediali digitali, audiovisivi, web.* Roma: Meltemi.

Guerzoni, G. (2008). *Effetto festival: L'impatto economico dei festival di approfondimento culturale.* La Spezia: Fondazione Carispezia.

Guerzoni, G. (2017). *Effettofestival 3. L'impatto di comunicazione dei festival di approfondimento culturale. Il caso Pistoia-Dialoghi sull'uomo 2017.* Pistoia: Fondazione Cassa di risparmio di Pistoia e Pescia.

Guerzoni, G. (Ed.) (2012). *Effetto festival 2012.* Firenze: Fondazione Florens.

Il conversatore (Ed.) (2019). *Il conversatore letterario. Programma del corso, da un'idea di Simonetta Bitasi e Gabriele Romagnoli*, Mantova, 23-24 febbraio 2019. https://s3.eu-central-1.amazonaws.com/festlettweb/Programma-conversatore-letterario.pdf Accessed 11.09.19.

Lancini, M., Buday E. (2013). *Effettofestival adolescenti. Volontariato e impatto formativo dei festival di approfondimento culturale.* La Spezia: Fondazione Carispezia Strumenti.

Pasti, D. (1997). Posti in piedi per sentire lo scrittore. In *La Repubblica,* 12 settembre 1997.

Piano strategico (1994). Piano strategico per la massimizzazione delle risorse culturali di Mantova. Relazione finale presentata al Comune di Mantova.

Polettini, P. (1996). Festival of Literature, Hay on Wye. Gran Bretagna, 12/18 aprile - 24/29 maggio 1996. Relazione dello stage presentata al Comune di Mantova.

Polettini, P. (1999). Su Festivaletteratura e i suoi misteri. In *Civiltà mantovana*, 109, novembre 1999.

Polettini, P. (2012). Il sistema cultura. La leggerezza innovativa dell'evento: il caso Festivaletteratura. In Leoni, G. (Ed.), *Storia di Mantova. Uomini, ambiente, economia, società, istituzioni. Volume terzo.* Mantova: Tre Lune, 472-477.

Salis, S. (2017). «L'importante è farsi meravigliare». In *Il Sole 24 Ore*, 04 settembre 2017.

Vitali, S. (2004). *Passato digitale. Le fonti dello storico nell'era del computer.* Milano: Bruno Mondadori.

Zanni Rosiello, I. (2009). *Gli archivi nella società contemporanea*. Bologna: Il Mulino.

DOI: 10.17469/O2106AISV000005

ROSANNA SORNICOLA, GIOVANNI ABETE, ELISA D'ARGENIO,
CESARINA VECCHIA

Raccontare un archivio di fonti orali: il progetto *Voci, parole e testi della Campania*[1]

In this paper, we share the experiences gained by our research team in the digitization and dissemination of an archive of dialectal speech. After presenting the structure and the contents of our archive, we discuss some strategies we employed to disclose the archive to an audience of non-specialists. Such strategies include a website with multimedia narratives based on the contents of the archive itself, and the creation of small museums which further enhance such contents and which are located in the same areas from which the data were collected. Thanks to these different dissemination strategies, our work meets the need for giving back to the communities under investigation and it offers a model for the enhancement of oral sources as intangible cultural heritage.

Key words: oral sources, digitization, dissemination strategies, Campanian dialects.

1. *Il progetto* Voci, parole e testi della Campania

In questo contributo si discutono le esperienze maturate dal nostro gruppo di ricerca nell'ambito delle attività svolte per la comunicazione e la valorizzazione di un archivio di parlato dialettale, con particolare riferimento ai risultati di un recente progetto sulla digitalizzazione e la messa in rete di materiali d'archivio.

Il progetto *Voci, parole e testi della Campania. Un archivio sociolinguistico ed etnografico digitale per la promozione culturale del territorio*[2], diretto da Rosanna Sornicola, è nato su iniziativa di tre gruppi di ricerca del Dipartimento di Studi Umanistici dell'Università di Napoli "Federico II"[3], con lo scopo di raccogliere e conservare testimonianze parlate e scritte della vita sociale, culturale ed economica della Campania e della sua storia, segnata sin da epoca antica dal multilinguismo e dalla diversità etnica e culturale: voci di persone che raccontano in maniera diretta le loro esistenze, registrate nell'immediatezza e spontaneità dei contesti in cui sono

[1] Sebbene questo articolo sia frutto del lavoro congiunto dei quattro autori, tuttavia, ai soli fini accademici, i paragrafi sono da attribuirsi nella maniera seguente: 1 e 6 a tutti gli autori, 2 e 3 a Elisa D'Argenio, 4 a Giovanni Abete, 5 a Cesarina Vecchia.

[2] www.archivicampani.unina.it. Il progetto è stato finanziato dal Programma Operativo Regionale (POR) Campania FESR 2007-2013 nell'ambito degli interventi dell'Obiettivo Operativo 1.10 "La cultura come risorsa - Sviluppo di tecnologie per la digitalizzazione e messa in rete di archivi e biblioteche".

[3] I tre gruppi di ricerca sono coordinati rispettivamente da Rosanna Sornicola (Linguistica generale), Nicola De Blasi (Storia della Lingua italiana) e Raffaele Giglio (Letteratura italiana).

state emesse[4]; parole che fanno parte dell'immenso patrimonio culturale immateriale costituito dalle lingue e dai dialetti; testi scritti, come documenti, lettere, diari, la cui produzione ha scandito la grande storia e la piccola storia di ogni giorno della Campania. Al centro del progetto è stata dunque posta la dimensione linguistica dei beni culturali, non solo per il grande interesse scientifico e culturale che i fatti linguistici rivestono di per sé, ma soprattutto per il ruolo che possono svolgere come chiave di comprensione degli aspetti più profondi della vita e della storia di un territorio.

Le attività del progetto sono state incentrate intorno alla realizzazione di due obiettivi principali. Il primo obiettivo ha riguardato la messa in rete di materiali di archivi a carattere prevalentemente linguistico secondo standard di catalogazione internazionali e nazionali[5], con particolare riguardo alle nuove tecniche di archiviazione e catalogazione per i beni culturali immateriali (v. § 2). Il secondo obiettivo era collegato a quella che nella teoria archivistica italiana è convenzionalmente descritta come la terza fase di vita di un archivio. Dopo la fase di formazione e di deposito dell'archivio,

> nella fase storica, dedicata alla conservazione e alla realizzazione di strumenti archivistici descrittivi, indispensabili per la ricerca, l'archivio diventa prevalentemente, anche se non esclusivamente, oggetto di indagine storiografica e di valorizzazione (Bonfiglio-Dosio, 2010: 30).

Perseguendo il fine di promuovere verso l'esterno la conoscenza dei materiali raccolti, il progetto ha dunque cercato di definire un nuovo modo di rendere "vivo" un archivio in rete, valorizzandone i contenuti attraverso lo sviluppo di modalità di fruizione diversificate, per lo studioso esperto e per un pubblico più ampio. A tal riguardo, infatti, come sarà illustrato più dettagliatamente nel § 4, sono state create delle "narrazioni" culturali che consentono di raccontare l'archivio ad un pubblico di non specialisti, mettendo in luce, anche attraverso la creazione di ulteriori contenuti multimediali, l'interesse dei materiali d'archivio sia sul piano linguistico, sia sul piano storico-culturale. Sono stati inoltre allestiti dei luoghi di accesso fisico all'archivio in alcuni punti del territorio campano, nella forma di piccoli musei multimediali, che abbiamo chiamato *mnemoteche*, rivolti ad un pubblico generalizzato, di cui si tratterà nel § 5. Tali attività sono connesse allo scopo più generale di raccordare la realizzazione di archivi con strategie che promuovano la conoscenza del patrimonio linguistico e culturale di territori meno noti e incentivino lo sviluppo turistico dei luoghi "minori" rappresentati dai piccoli comuni. Le citate modalità di fruizione "mediata" dell'archivio sono poi affiancate da una modalità di fruizione "diretta" dell'archivio in rete, destinata principalmente agli specialisti (dialettologi, etnografi, storici, etc.)[6].

---

[4] Sulla rappresentazione della voce come bene culturale si veda De Dominicis (2002a; 2002b).

[5] Sull'importanza di rendere facilmente accessibili alla comunità scientifica i corpora vocali e sulle modalità di conservazione dei materiali sonori su supporti digitali si veda Romito, Lio (2007).

[6] Per consultare l'archivio in rete (www.archivicampani.unina.it/archivi_campani_dev) è necessario disporre delle credenziali di accesso. Le credenziali possono essere richieste inviando una e-mail all'indirizzo vptcampania@gmail.com con l'indicazione dei dati anagrafici del richiedente e le ragioni di interesse.

2. *L'*Archivio dei Dialetti Campani *on-line*

Per il progetto *Voci, parole e testi della Campania* il nostro gruppo di ricerca ha lavorato sulla messa in rete di materiali provenienti da un archivio di parlato dialettale già esistente, l'*Archivio dei Dialetti Campani* (*ADiCa*). Formatosi a partire dagli anni Novanta del secolo scorso e tuttora in ampliamento, l'*ADiCa* documenta la ricca diversità linguistica della Campania, dalle aree costiere all'entroterra, attraverso la raccolta sul campo di testimonianze orali, in particolare di parlato spontaneo dialettale (Como, Milano, 1999; Sornicola, 1999; Sornicola, Calamai, 2014). Le inchieste, condotte da ricercatori e studenti, si sono concentrate finora in particolare sull'area flegrea, sulla città di Napoli e i suoi quartieri periferici, sull'area vesuviana e sull'Irpinia. Accanto alle interviste audio e video, l'*ADiCa* accoglie inoltre riprese video e fotografie a persone, luoghi e oggetti, legati non soltanto al momento dell'inchiesta vera e propria, ma più in generale alla storia, all'economia, alla vita sociale e culturale delle comunità linguistiche indagate.

La messa in rete ha riguardato una selezione di materiali dell'*ADiCa* raccolti nell'entroterra della Campania, in Irpinia, negli anni più recenti. La struttura dell'archivio digitalizzato, a cui d'ora in avanti faremo riferimento, e il relativo sito che lo ospita, sono stati progettati da Valentina Retaro (Retaro, 2015) e realizzati in collaborazione con il Politecnico di Milano, partner del nostro progetto per gli aspetti informatici.

Sulla base di studi preliminari della bibliografia specializzata, in particolare sul trattamento archivistico delle fonti orali, sono stati innanzitutto definiti gli standard di descrizione archivistica di riferimento per l'elaborazione della strutturazione dei dati dell'archivio digitalizzato (Retaro, 2015: 13-14)[7]. Per l'elaborazione delle schede catalografiche sono state inoltre prese in considerazione le riflessioni e le proposte metodologiche già sviluppate da altri gruppi di ricerca italiani impegnati in progetti analoghi[8]. L'elaborazione ha naturalmente tenuto in debito conto la natura dell'*ADiCa* quale archivio complesso, ossia un archivio che include materiale di diversa tipologia e conservato su supporti differenti (Dentoni-Litta, 1999: 137).

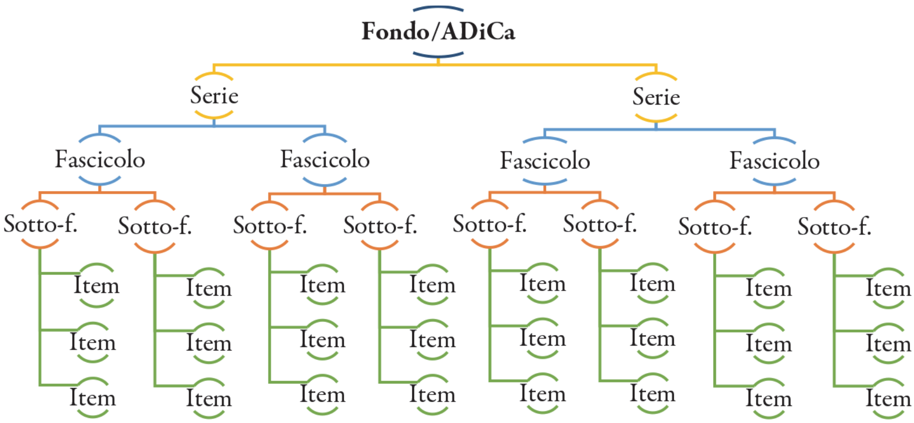
L'archivio on-line è strutturato in cinque livelli gerarchizzati secondo il criterio "dal generale al particolare", a ciascuno dei quali corrisponde una scheda catalografica. Come schematizzato in Figura 1, al livello più alto si pone il Fondo, inteso come sinonimo di "archivio" (Valenti, 2000: 107-109), ai livelli intermedi le Serie, i Fascicoli e i Sotto-fascicoli, e al livello più basso le Unità documentarie (*items*).

---

[7] Si vedano soprattutto gli standard ISAD (General International Standard Archival Description), ISAAR (International Standard Archival Autorithy Records for Corporate Bodies, Persons and Families) e le linee guida indicate dallo IASA (International Association of Sound and Audiovisual Archives).

[8] In particolare, si è fatto riferimento al progetto "Grammo-foni. Le soffitte della voce (Gra.fo)" della Scuola Normale Superiore di Pisa e dell'Università di Siena (Calamai, 2011a; 2011b; Calamai, Bertinetto, 2012; 2014; Canazza, Calamai, Bertinetto & De Dominicis, 2012) e al progetto "I granai della memoria" dell'Università degli Studi di Scienze Gastronomiche del Piemonte (Grimaldi, Porporato, 2011).

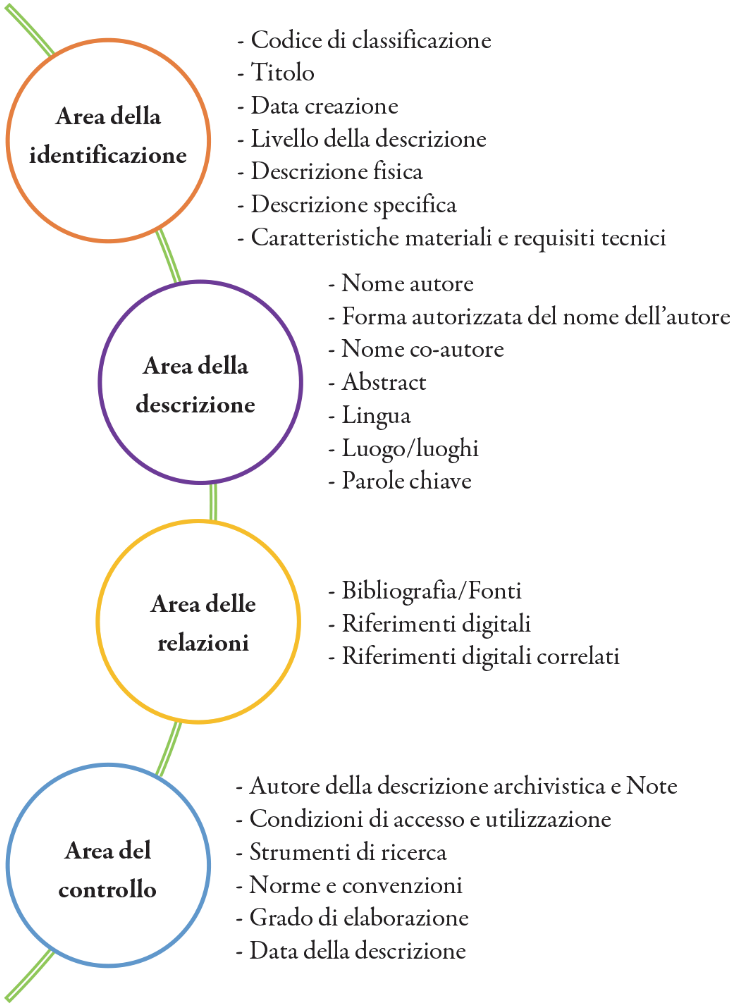
Figura 1 - *Struttura ad albero dell'archivio on-line*



Ad una Serie afferiscono i materiali prodotti all'interno di un singolo progetto di ricerca; ciò consente di rispettare il principio di provenienza e di preservare il vincolo archivistico tra le interviste e le altre tipologie documentarie ad esse collegate. I materiali sono poi distribuiti in Fascicoli secondo un criterio associativo di tipo tematico o legato ad una specifica attività. A loro volta i Fascicoli sono distinti in Sotto-fascicoli nei quali i materiali sono raggruppati sulla base di un'associazione più restrittiva. L'Unità documentaria contiene infine il singolo documento, al quale è associato un contenuto multimediale (audio, video, foto). Ad esempio, nella Serie "Transumanza e pastorizia", i documenti raccolti sul campo da Giovanni Abete per il progetto di ricerca "La cultura linguistica e materiale dei territori dell'Irpinia: la transumanza e la pastorizia in Alta Irpinia", sono distribuiti in quattro fascicoli denominati "Storia dei luoghi", "Ambiente e paesaggio", "Fonti Orali" e "Etnografia e Folclore". In quest'ultimo Fascicolo, il Sotto-Fascicolo "Attività produttive" raccoglie unità documentarie di tipo fotografico e audio-visivo relative alle attività tradizionali tipiche dell'Alta Irpinia, delle quali si documentano pratiche, strumenti, usi e abitudini, con particolare riguardo alla produzione dei formaggi tipici di questo territorio.

Le schede catalografiche si compongono di quattro macro-aree (area dell'identificazione, area della descrizione, area delle relazioni e area del controllo), a loro volta suddivise in una serie di campi comuni a tutti i livelli, dal Fondo all'Unità documentaria, a cui si aggiungono campi specifici per alcuni livelli. Si veda, a titolo esemplificativo, la Figura 2 che schematizza la struttura della scheda catalografica del livello dell'Unità documentaria.

Figura 2 - *Struttura della scheda catalografica del livello Unità documentaria*



3. *L'intervista nell'*ADiCa*: dall'etnotesto alle unità documentarie*

Sebbene, come si è visto, l'*ADiCa* si componga di diverse tipologie documentarie, il fulcro dell'archivio è rappresentato dalle interviste dialettologiche, condotte principalmente con le tecniche dell'intervista libera e della conversazione guidata (Sornicola, 1999; Como, Milano, 1999; Como, 2006; Abete, 2011). I materiali così raccolti rientrano tra quelli definibili come *etnotesti*, ossia

> testi orali che rappresentano un'espressione autonoma della cultura di una comunità linguistica: testi liberi (ricordi autobiografici, storie di vita, testimonianze di usi e tradizioni, descrizioni di oggetti e di tecniche ergologiche [...]); ma anche testi fissati dalla tradizione (proverbi, indovinelli, preghiere, ecc.), o testi solo parzialmente formalizzati (leggende, storie, ecc.) (Canobbio, 2004: 306).

Gli etnotesti, oltre a costituire una manifestazione spontanea del bagaglio di esperienze e conoscenze dei singoli individui che li hanno prodotti, sono in grado di fornire una rappresentazione della dimensione comunitaria a cui i singoli appartengono, in quanto

> descrivono dal di dentro l'identità culturale collettiva, ricostruendo verbalmente in maniera autentica [...] i costumi di vita e di lavoro della comunità e del piccolo gruppo, le sue tecniche e usanze, le sue occasioni rituali e il patrimonio di sapere e di credenze (Berruto, 1996: 168).

L'etnotesto ha dunque molteplici valenze: è oggetto di analisi linguistica e dialettologica e, «ai fini della ricostruzione e della documentazione di microsistemi culturali e identitari, l'etnotesto appare utile anche quando non venga espresso nel codice dialettale» (Paternostro, Sottile, 2010: 609); è una fonte storica per la ricostruzione delle dinamiche sociali e culturali vissute da un territorio e un mezzo per valorizzarne le peculiarità linguistico-culturali.

La centralità dell'etnotesto e dell'informatore si traducono nell'architettura dell'archivio nella presenza di uno specifico Fascicolo denominato "Fonti orali", in cui sono raccolti tutti i documenti relativi agli informatori intervistati. Nel Sottofascicolo dedicato al singolo informatore, l'intervista riceve poi una duplice catalogazione: da un lato, catalogata nella sua interezza, costituisce un'unità documentaria (a cui non sono però associati contenuti multimediali); dall'altro, in qualità di documento complesso[9], essa è scomposta, sulla base di diversi criteri (temi trattati, tecniche di elicitazione, tipologia testuale, etc.), in un certo numero di etnotesti che costituiscono ulteriori unità documentarie (alle quali è sempre allegato il contenuto multimediale audio o video). Inoltre, la visione dell'intervista «come co-costruzione dell'intervistato e dell'intervistatore (...) nella quale i due partecipanti cercano di rendersi mutualmente intelligibili saperi, pratiche, rappresentazioni» (D'Agostino, Ruffino, 2005: 45), è resa esplicita nella soluzione adottata per l'identificazione dell'autore del testo orale: se l'autore del documento-intervista e dei suoi estratti è naturalmente l'informatore che l'ha prodotto, il ruolo attivo del ricercatore nella realizzazione dell'intervista è valorizzato come co-autore.

Dal punto di vista delle tipologie di etnotesti archiviati, l'*ADiCa* si presenta particolarmente ricco, includendo sia testi liberi, sia testi formalizzati e semi-formalizzati. Un esempio di etnotesto parzialmente formalizzato è rappresentato da un "cunto", un racconto popolare, che, poiché non scomponibile in unità tematiche minori

[9] Dal punto di vista archivistico, si definiscono oggetti complessi gli «oggetti singoli logicamente e/o fisicamente compositi, le cui parti, concettualmente e/o fisicamente separabili, differiscono per un qualche aspetto ritenuto importante ai fini della classificazione» (Corti, 2003: 34). L'intervista dialettologica rientra a pieno titolo in questa tipologia, poiché può contenere al suo interno documenti diversi tra loro per i temi trattati, per le tecniche di elicitazione utilizzate (es. intervista libera vs questionario), per la tipologia testuale (es. filastrocche, fiabe, canti, etc.). Tale complessità interna giustifica la scomposizione dell'intervista in sotto-unità documentarie (v. sopra nel testo). Sul problema della definizione delle unità documentarie "di base" in un archivio digitale di fonti orali e sulle questioni metodologiche connesse cfr. Calamai, Biliotti & Bertinetto (2014); Bertinetto, Calamai (2016).

è stato archiviato nella sua interezza, con il relativo video allegato. L'informatore che lo ha prodotto vive a Montella (AV), è un appassionato di racconti popolari e da anni tramanda ai figli e ai nipoti le storie che ha ascoltato sin da bambino. Quella raccolta nell'intervista tratta di una disputa per la proprietà di un melo posto al confine tra un convento di frati Francescani e un convento di frati Domenicani. Il litigio si risolverà con una sfida consistente in un dialogo a gesti ("alla mupégna") tra il rozzo e ignorante torzone (il "picuózzo") del convento francescano e l'erudito padre guardiano dei Domenicani, la quale darà luogo ad una serie di comici equivoci nell'interpretazione dei rispettivi segni. Se ne dà di seguito la trascrizione della parte iniziale[10]:

> [na ˈvɔta ndʒi ˈstia nu ˈpjettso re ˈtɛrra ‖ ma ˈɛra abbasˈtantsa ˈgrwosso stu ˈpjettso re ˈtɛrra ‖ (.) e<e> e ndʒi ˈstiano rui kuɱˈvɛnti re ˈmɔnaʃi̥ ‖ (.) ˈuno nna+ a nno ˈlɑːto ‖ e nn ˈato a n ˈato ˈlɑto ‖ rind a ddui ˈtsinni divɛrsi ‖ (.) peˈrɔ ˈkwisti si puˈtiano veˈre komm a+ a lu fˈfummo re la kanˈnela ‖ (.) nun tsi puˈtiano veˈre ˈuno pe nn ˈɑto ‖ pekˈke ‖ no koɱˈvɛnto (.) ˈɛra+ re frantʃesˈkɑːni ‖ e nn ˈato ˈɛra re domeniˈkɑːni ‖ ˈkwiri ke ˈvanno visˈtuti ˈjaːŋgi ‖ (.)]

> Una volta c'era un pezzo di terra, ma era abbastanza grande questo pezzo di terra, e c'erano due conventi di monaci, uno a un lato, e un altro a un altro lato, in due lati diversi. Però, questi si potevano sopportare come il fumo della candela, non si potevano sopportare l'uno con l'altro, perché un convento era di Francescani e un altro era di Domenicani, quelli che vanno vestiti bianchi.

Agli antichi mestieri che sopravvivono ancora oggi in forme residuali si riferisce invece il secondo etnotesto che qui si presenta, "Vacche alla Puglia e pecore alla Terra di Lavoro". Il tema trattato riguarda la pastorizia transumante, un'attività tradizionale che ha avuto una grande importanza nella storia economica e culturale di molte aree d'Europa (v. Marino, 1991). Nell'etnotesto, estratto come unità tematica da un'audio-intervista più lunga svoltasi a Bagnoli Irpino (AV), l'informatore spiega i motivi, legati alle differenti caratteristiche del terreno, per i quali gli allevatori di bovini preferiscono svernare nei pascoli della Puglia, mentre gli allevatori di ovini prediligono i pascoli della cosiddetta "Terra di Lavoro":

> [peˈrɔ li ˈβaʃkuli ˈvɛːrə pə rə pˈpɛgur e ppə rə vˈvakke ‖ rə vvakk a la βuːʎʎ e rə pˈpɛgur a ttɛr laˈoː ‖ riˈʒimə nu ‖ (.) [...] la dˈdzɔna ˈvɔstə<ə> ru veˈsuvjə ‖ ɛ ttɛrr aʃˈʃuttə ‖ ˈtɛrra vesuˈvjaːna ‖ ɛ sˈsabbja ‖ (.) ˈpɔtə ˈkjɔvə ˈkwandə ˈvɔːlə ‖ ˈkwandə<ə> peˈrɔ<ɔ> stann aʃˈʃutti ‖ (.) iɱˈveʒə la dˈdzɔnə<ə> per ˈdirə salerniˈðaːnə ‖ (.) ˈdzɔnə laː ˈbuʎʎə<ə> la ˈbuʎʎə+ la ˈtɛrra<a> ˈgwannɪ ˈɟɔvə sə ˈf̬aʃə na ˈɣreːðɐ ‖ ˈf̬aʃə+ sə ˈmbastə ‖ iɱˈveʒə la ˈpɛgura va truˈwannə la ddzɔn aʃˈʃuttə k̬u lu ˈbɛrə ‖ se nɔ ˈɛssənə ˈtuttʊ tsɔp ‖ (.)]

> Però i pascoli veri per le pecore e per le vacche… le vacche in Puglia e le pecore alla Terra di Lavoro, diciamo noi [...] la vostra zona, il Vesuvio, è terra asciutta, terra

[10] Con il puntino tra parentesi tonde indichiamo le pause silenti (cioè silenzi di durata superiore ai 300 ms), con la doppia barra verticale le pause prosodiche (segnalate da particolari andamenti melodici ma senza interruzione di fonazione), e tra parentesi uncinate le pause piene; il segno '+' indica un'interruzione improvvisa.

vesuviana, è sabbia. Può piovere quanto vuole… però stanno asciutti (gli animali). Invece la zona, per dire, del salernitano o la zona della Puglia… la Puglia, la terra quando piove diventa una creta, si impasta, invece la pecora deve appoggiare il piede sull'asciutto (lett. vuole la zona asciutta con il piede) altrimenti si azzoppano tutte.

4. *Raccontare l'archivio*

Come argomentato nel § 1, l'obiettivo del nostro progetto è stato duplice: da un lato, si intendeva creare un archivio on-line di fonti orali che rispettasse i più moderni standard di archiviazione nel campo dei beni culturali immateriali; dall'altro, si avvertiva l'esigenza di raccontare questo archivio ad un pubblico ampio, fornendo le coordinate storiche, culturali e linguistiche necessarie ad apprezzare appieno il valore dei materiali archiviati. In relazione a questo secondo obiettivo, utilizzando il software *1001Storia*[11], abbiamo realizzato delle narrazioni multimediali che raccontano alcuni macro-temi rappresentati nell'archivio. In questa sede prenderemo in esame la narrazione "Tratturi e antiche vie della terra di mezzo", che descrive la pratica della transumanza in Irpinia[12]. La scelta di un tema così particolare è motivata nel paragrafo introduttivo della narrazione multimediale:

> Questa pratica, oggi quasi dimenticata, ha condizionato per secoli la vita del territorio campano permeando molti aspetti dell'economia e della cultura locali. Le voci dei pastori dell'Irpinia, documentate nell'archivio *ADiCa*, offrono una testimonianza diretta di questi aspetti e raccontano, attraverso un punto di vista originale, la storia culturale della regione[13].

Le narrazioni sono organizzate su più livelli gerarchici, ognuno dei quali ha un suo menu di navigazione: il livello della narrazione principale, un livello di approfondimenti, e infine una sezione "Documenti" che presenta alcuni materiali di archivio strettamente legati al tema della narrazione e costituisce pertanto l'anello di congiunzione con l'archivio on-line.

In Figura 3 si riporta uno *screenshot* dell'interfaccia utente della narrazione, impostata sul menu di primo livello.

---

[11] Si tratta di un software per lo *storytelling* che è stato creato dal laboratorio *HOC-LAB* del Politecnico di Milano, partner del nostro progetto per gli aspetti informatici. Per maggiori informazioni su questo strumento e sul suo uso applicato ai beni culturali sia materiali che immateriali si vedano Paolini, Di Blas (2014); Di Blas (2016).

[12] www.archivicampani.unina.it/narrazioni/adica_tratturi.html.

[13] La transumanza fornisce anche elementi importanti per l'interpretazione delle dinamiche dialettali che interessano il territorio irpino. Su questo argomento si vedano Abete (2016); Retaro, Abete (2018); Vecchia (2018a; 2018b).

Figura 3 - *L'interfaccia della narrazione multimediale*

Ogni livello è diviso in paragrafi, ciascuno dei quali ha una durata di 2-3 minuti e costituisce una micro-unità tematica che può anche essere fruita autonomamente[14]. Il racconto si dipana attraverso una voce narrante, registrata da uno *speaker* professionista e accompagnata da una carrellata di immagini. Ogni immagine è commentata da una breve didascalia sicché, a complemento del testo della voce narrante, immagini e didascalie costituiscono un piano parallelo di comunicazione delle informazioni. Le immagini utilizzate sono in prevalenza foto originali, che abbiamo raccolto direttamente sul campo, e che a loro volta sono catalogate nel nostro archivio. Da questa stessa schermata è possibile accedere agli approfondimenti e alla sezione "Documenti"; inoltre, la funzione "testo" consente di visualizzare in forma scritta il testo della narrazione.

Per dare un'idea dei contenuti veicolati in un singolo paragrafo, si riporta di seguito il testo del paragrafo 7, "Tratturi della terra di mezzo", che presenta i principali percorsi tratturali del territorio irpino, ricollegandosi ai materiali d'archivio che trattano di questo tema, e dei quali abbiamo fornito un esempio nel § 3 (v. in particolare gli ultimi due periodi):

> Il Pescasseroli-Candela non era certo l'unico tratturo che attraversasse il territorio irpino. Una rete di tratturi minori metteva in collegamento i pascoli montani dell'Irpinia con quelli del Tavoliere pugliese a est, della piana salernitana a sud, e della Terra di Lavoro ad ovest. I collegamenti con i pascoli orientali erano assicurati da due percorsi principali che tagliavano l'Irpinia in senso longitudinale: uno passava più a nord e attraversava l'altopiano del Formicoso, da Guardia Lombardi a Lacedonia, e di qui giungeva a Candela e al Tavoliere; l'altro era in parte parallelo a

[14] Come per ogni moderno ipertesto digitale, le nostre narrazioni possono essere fruite seguendone la struttura lineare prestabilita dall'autore o, più comunemente, l'utente può selezionare un percorso personalizzato di fruizione, sulla base dei propri interessi e del tempo a disposizione. La brevità di ciascuna unità tematica è funzionale a questa duplice possibilità e conforme agli standard della comunicazione in rete (v. Di Blas, 2016: 4-5).

> questo e passava più a sud, seguiva il corso dell'Ofanto fino a Monteverde, e da qui arrivava a Melfi, in Basilicata, dove si innestava sul tratturo regio Melfi-Castellaneta. Questi percorsi si congiungevano nella parte occidentale a due importanti vie per il Salernitano, rispettivamente attraverso la sella di Conza e il passo delle croci d'Acerno. Altri tratturi conducevano infine ad ovest e mettevano in collegamento i pascoli dei monti Picentini con le pianure della Terra di Lavoro, in particolare del nolano e dell'aversano, attraverso il vallo di Lauro e il vallo di Baiano. Sembra in effetti che i pastori irpini dei monti Picentini preferissero quest'area come pascolo invernale per le pecore, mentre i pascoli della Puglia erano ritenuti più adatti ai bovini.

Il testo di questo paragrafo combina insieme le informazioni raccolte sul campo dalla voce dei pastori irpini con le ricostruzioni degli storici che si sono occupati dei percorsi della transumanza (Benaiteau, 1986; Brancaccio, 1991; Di Guglielmo, 2002). Questo intreccio tra dato orale contemporaneo e dato storico è una costante della narrazione e consente di dare la giusta profondità alle vicende e alle attività documentate dalle fonti orali. Il tema dei tratturi irpini si intreccia, infatti, in maniera affascinante, con quello della viabilità antica, delle vie consolari romane, come l'Appia e la Traiana, giacché i moderni tratturi ricalcano in parte queste antiche vie. Nei diversi paragrafi della narrazione si parla quindi della transumanza nel mondo antico, in particolare romano; poi, della fase aragonese, con la sistemazione dei grandi tratturi regi e la nascita di una vera e propria industria della pastorizia transumante; quindi, della transumanza in Irpinia e dei principali tratturi che, da tempi remoti, attraversano questo territorio; infine, degli ultimi pastori irpini che, ancora oggi, praticano la pastorizia alla maniera transumante.

Quando un tema è documentato direttamente nei nostri materiali di archivio, un paragrafo della sezione "Documenti" ne tratta espressamente. Ad esempio, il paragrafo "Gli ultimi pastori transumanti" invita gli utenti a consultare i materiali di archivio relativi a questo tema così importante:

> La transumanza sopravvive ancora oggi in Irpinia, seppure in contesti ristretti. A praticarla sono soprattutto gli allevatori di bovini podolici, in particolare di Montella e di Bagnoli Irpino. Le caratteristiche dei bovini podolici, poco adatti alla stalla e bisognosi di pascoli molto estesi, hanno creato le condizioni perché questa razza venisse allevata ancora alla maniera transumante. Ogni anno, al sopraggiungere dell'inverno, gli allevatori di Montella e Bagnoli Irpino si mettono in viaggio con le loro mandrie, per raggiungere i pascoli di pianura del Tavoliere pugliese. L'archivio *ADiCa* racconta questo viaggio con una ricca collezione di foto e filmati, che consentono di seguire da vicino questa pratica affascinante.

La voce narrante è accompagnata in questo caso da foto originali che documentano la transumanza dei bovini podolici verso la Puglia, attraverso un'Irpinia completamente innevata (v. Figura 4). Furono realizzate nel novembre del 2013 da Giovanni Abete, insieme con diversi documenti audio-video, anch'essi catalogati nel nostro archivio. Come si può vedere, dunque, la sezione "Documenti" non presenta i materiali di archivio in maniera diretta, ma ne fornisce sempre un racconto "mediato", invitando però, al tempo stesso, l'utente a visitare l'archivio on-line e ad accedere ai documenti effettivi. Questa operazione di mediazione è ritenuta necessaria affinché le testimonianze orali raccolte nell'archivio, per loro natura frammentarie e di non facile lettura per i non addetti ai lavori, siano inserite in una trama organica e coerente e possano essere trasmesse in una forma più accessibile.

Figura 4 - *Esempi di foto presenti nella narrazione multimediale*

Inoltre, lo strumento della narrazione consente di divulgare tutta una serie di informazioni contestuali, di tipo storico, linguistico, culturale, etc. che sono di fondamentale importanza per la comprensione del valore documentario delle fonti orali, ma che difficilmente potrebbero comparire nei testi dei nostri informatori. Senza contare che proprio questi contenuti "extra" potrebbero attrarre un fruitore di cultura media e invogliarlo ad interessarsi ai documenti di archivio (Matranga, 2018: 598). Raccontare un archivio di fonti orali significa anche questo: consentire che l'archivio resti "vivo", assicurando non solo la conservazione del bene documentario ma anche la sua diffusione e valorizzazione.

5. *Le* mnemoteche

Un'ulteriore modalità di valorizzazione dei materiali raccolti riguarda la creazione di spazi fisici volti alla fruizione diretta dei contenuti dell'archivio. Sono stati infatti creati alcuni piccoli musei della memoria di tipo multimediale, le *mnemoteche*, in quattro diversi centri della Campania e in particolare nei comuni di Ariano Irpino, Greci, San Mango sul Calore e Sessa Aurunca. Come le narrazioni, anche le *mnemoteche* sono state pensate per presentare ad un pubblico di non specialisti i documenti raccolti nell'archivio, e in primo luogo le fonti orali, offrendo all'utente tutte le informazioni necessarie per poter comprendere e apprezzare appieno il valore culturale e identitario espresso da ciascuna testimonianza.

Il nostro gruppo di ricerca ha curato la progettazione e l'allestimento della *mnemoteca* di Ariano Irpino[15], interamente dedicata alla storia culturale, sociale e linguistica dell'Irpinia. All'interno della *mnemoteca* i documenti audio-visivi e fotografici

[15] La *mnemoteca* ha sede presso l'area museale del centro di ricerca di Biologia e genetica molecolare (Biogem).

dell'archivio sono stati organizzati in tre diversi percorsi caratterizzati da tre differenti modalità di fruizione: una mostra fotografica, postazioni tematiche interattive e una sala con proiezioni di documentari e filmati.

Il primo percorso è costituito da un'esposizione di fotografie dedicata all'Irpinia, ai suoi abitanti e ai luoghi più caratteristici di questo territorio. La mostra si apre con un'ampia selezione di foto dei diversi borghi medievali: piccoli centri storici raccolti intorno ai castelli di epoca longobarda e normanna, tutti caratterizzati da un fitto intreccio di strade, vicoli acciottolati e gradinate sui quali si affacciano le case in pietra locale e ricchi palazzi signorili. Le altre immagini mostrano invece i diversi paesaggi dell'Irpinia nella loro straordinaria varietà: le vedute sulle valli dell'Ofanto e del Calore riprese dalle alture e dai belvedere dei centri posti a corona dei Monti Picentini; i fiumi e i torrenti che percorrono l'intero territorio in diverse direzioni; i boschi di castagni e faggi nelle aree montuose; i filari di viti e ulivi e gli ampi campi di frumento caratteristici delle zone collinari e degli altopiani dell'Irpinia orientale.

Un'ulteriore sezione del percorso fotografico ritrae uomini e donne impegnati nei lavori tipici del territorio o nei momenti di vita familiare e comunitaria. Sono infatti rappresentati contadini, castagnai e pastori còlti nelle diverse attività che caratterizzano questi mestieri di antica tradizione. Nella raccolta sono incluse anche foto corali con ritratti di famiglia, matrimoni, celebrazioni collettive, feste religiose e civili che documentano le differenti forme di vita associativa all'interno delle varie comunità irpine.

Figura 5 - *Foto storica della battitura delle castagne*[16]

[16] La fotografia mostra una scena d'insieme dell'antica pratica di battitura delle castagne. In questa fase le castagne precedentemente essiccate sono private delle buccia per ottenere le "castagne bianche", uno dei prodotti tipici locali. La battitura delle castagne, oggi praticata con specifiche macchine, in passato si realizzava tramite una tecnica manuale. Come vediamo nella fotografia, la battitura prevedeva l'uso di un sacco piuttosto lungo, all'interno del quale venivano depositati circa dieci chili di castagne. Due addetti, prese le estremità del sacco, lo facevano roteare sui due lati battendo su un ceppo posto al centro. In questo modo la buccia e la pellicola interna che ricopre la parte amidacea, ormai secche, si staccavano lasciando libero il frutto, ovvero la parte bianca.

Il percorso fotografico accoglie anche diverse foto storiche (v. Figura 5), molte delle quali provenienti da collezioni private, che ci sono state donate dagli informatori o dalle associazioni locali. Altre fotografie, invece, sono state realizzate da noi sul campo durante le diverse campagne d'interviste[17]. In questo percorso l'intreccio suggestivo tra immagini antiche e moderne, evidente sul piano formale, vuole comunicare al visitatore la lunga durata di alcune tradizioni locali che, attraversando i secoli e la grande storia, rappresentano ancora oggi una parte importante della vita delle persone e delle comunità irpine.

Il secondo percorso della *mnemoteca* di Ariano è quello multimediale. Questa sezione comprende quattro postazioni tematiche interattive in cui vengono proposti materiali di vario tipo (audio, video, foto) accompagnati da testi divulgativi e da interviste a specialisti di diversi settori e in particolare a linguisti, storici, archeologi, storici dell'arte e della letteratura antica e moderna. I testi e le interviste ad esperti fanno da cornice ai materiali d'archivio e hanno la funzione di supportare e guidare l'utente per una corretta comprensione del contesto storico, sociale e culturale in cui si inseriscono i diversi documenti d'archivio e in particolare gli etnotesti. Ogni postazione interattiva è stata associata ad un tema:

1. *Una storia multietnica e multilingue: lingue e dialetti in Irpinia*. Come tutta l'Italia meridionale, anche l'Irpinia ha rappresentato sin dall'antichità un territorio di passaggio per genti di *ethnos*, lingua e cultura differenti. L'Irpinia è stata infatti scenario di movimenti di popolazioni di diversa provenienza le cui tracce sono in parte ancora visibili non solo nelle testimonianze tangibili, ma anche nelle parlate attuali. Nelle diverse fasi storiche e preistoriche si sono succedute sul territorio attraverso varie ondate migratorie genti di lingua osca[18], latina e germanica, dando vita, come normale nelle situazioni di contatto linguistico, a periodi più o meno lunghi di bilinguismo i cui effetti, soprattutto rispetto al contatto latino-germanico, sono oggi evidenti nel lessico dei dialetti locali. La postazione interattiva offre all'utente alcune informazioni di base di tipo storico, linguistico e filologico, sulle antiche lingue parlate sul territorio irpino, sia rispetto alle famiglie linguistiche di appartenenza, sia riguardo alle importanti testimonianze epigrafiche e letterarie che ci sono pervenute. Per quanto riguarda il quadro linguistico attuale, invece, si è cercato di documentare alcuni dei principali tratti dialettologici delle varietà irpine e la distribuzione dei fenomeni più rappresentativi sul territorio. L'area irpina, come tipico delle zone periferiche, presenta interessanti fenomeni fonetici, morfologici e sintattici di tipo conservativo, nonché interessanti forme di variazione diatopica tra i diversi punti. Questo territorio appare infatti fortemente articolato sul piano dialettologico, con differenze anche significative tra i centri occidentali, prossimi alla provincia di Napoli, e quelli invece situati nelle parte centro-orientale dell'Irpinia: que-

[17] Sia le foto storiche sia quelle raccolte da noi sul campo sono state incluse e catalogate nell'archivio.

[18] Proprio dall'Irpinia, e nello specifico da Avella, proviene una delle più importanti testimonianze epigrafiche per lo studio dell'osco, il *Cippus Abellanus*, un trattato stipulato da magistrati sanniti per la demarcazione dei confini tra le città di Abella e Nola.

sti ultimi infatti presentano caratteri propri e diversi elementi di discontinuità rispetto alle varietà di tipo napoletano. Sulla scorta delle inchieste e del lavoro di ricerca svolto dal nostro gruppo in Irpinia[19], per la postazione interattiva dedicata al multilinguismo sono stati realizzati vari testi di tipo divulgativo e alcune carte linguistiche a supporto dei documenti audio e audio-visivi contenenti parlato dialettale: questi contenuti testuali hanno lo scopo di mostrare anche a un pubblico di non esperti i tratti linguistici locali e la variazione dialettale che contraddistingue l'area. In questa postazione, infine, è stato dato spazio anche al racconto e alla descrizione della comunità arbëreshe di Greci, sede dell'unica minoranza linguistica della Campania. Rispetto alle altre varietà linguistiche dell'Arbëria nell'Italia meridionale, quella di Greci si presenta particolarmente conservativa e risulta ancora oggi vitale soprattutto tra i parlanti di età adulta e anziana[20].

2. *I segni della storia*. L'Irpinia conserva numerose testimonianze del passaggio dei diversi popoli che nel corso del tempo hanno conquistato e occupato il suo territorio. Attraverso l'apporto di storici e di archeologi, alcuni dei quali con esperienze di scavo e di ricerca proprio in Irpinia, si è cercato di documentare e presentare ad un pubblico di non specialisti i resti degli antichi insediamenti romani, longobardi e normanni che si possono ammirare oggi sull'intero territorio irpino. Del ricco patrimonio archeologico di epoca romana fanno parte gli scavi degli antichi *municipia* di *Aeclanum*, *Compsa*, *Aquilonia* e *Abellinum*, l'ampia collezione di manufatti esposti nel museo archeologico di Avellino, nonché i frammenti di monumenti funebri, di colonne e di iscrizioni marmoree inglobati come materiali di riuso in edifici di epoca più tarda. Per l'epoca medievale invece si evidenziano i resti di torri e mura difensive di età longobarda diffusi nei centri irpini un tempo parte dei gastaldati di Montella, Rota, Conza e Quintodecimo, e i castelli di età normanna spesso sorti proprio sulle rovine di epoche precedenti.
3. *Antiche forme di spiritualità e vita religiosa*. Una parte importante della storia culturale dell'Irpinia è costituita dalle diverse forme di spiritualità espresse sin da tempi remoti dai suoi abitanti. Sia i culti pagani sia le antiche o più recenti devozioni cristiane hanno lasciato il segno di una profonda vita religiosa, evidente nei diversi luoghi sacri presenti sul territorio, nei santuari e monasteri meta di pellegrinaggi per secoli, nei riti e nelle feste che ancora oggi si celebrano nei vari centri irpini e che costituiscono una parte importante del patrimonio culturale immateriale locale.
4. *Mestieri e attività produttive tradizionali*. Come si è visto nei paragrafi precedenti, buona parte dei materiali raccolti nell'archivio, in particolare nella forma di etnotesti, ma anche di fotografie e filmati, riguarda il tema dei mestieri e delle pratiche produttive tipici dell'Irpinia. Il patrimonio di saperi che ruota intorno alle

[19] Si vedano in particolare i lavori di Sornicola (2017b); Abete (2016; 2017; 2018); Abete, Vecchia (2018); Retaro, Abete (2018); Vecchia (2017; 2018a; 2018b).

[20] Sempre nell'ambito del progetto *Voci, parole e testi della Campania*, alla comunità albanofona di Greci e all'arbëresh sono state dedicate una specifica "Serie" nell'archivio on-line, una narrazione multimediale e una *mnemoteca*, con sede proprio nel comune di Greci, di cui è responsabile scientifico Emma Milano.

diverse attività locali, spesso di antica tradizione, rappresenta anch'esso un bene immateriale di enorme interesse culturale nonché linguistico. I paesi dell'Irpinia presentano tutti un'antica vocazione contadina e pastorale e un'economia da sempre legata all'ambiente montano e rurale caratteristico del territorio. I mestieri tradizionali riguardano pertanto principalmente i settori dell'agricoltura, della pastorizia e delle altre attività legate alla montagna. Proprio attraverso le parole e il racconto degli informatori che hanno svolto in prima persona queste attività è stato possibile ricostruire il ricco sistema di competenze tecniche e prassi tradizionali, di oggetti, attrezzi, strumenti e prodotti, per lo più codificato in un lessico dialettale altamente specializzato, che contraddistingue le attività maggiormente rappresentative del luogo (v. Abete, 2017). Alcuni dei mestieri documentati nell'archivio sono oggi scomparsi, in altri casi invece si tratta di attività su cui è ancora incentrato il lavoro di tante aziende presenti sul territorio impegnate nel settore agroalimentare. In particolare, nella postazione multimediale dedicata a questo tema sono stati descritti nel dettaglio attraverso testi, fotografie e filmati di supporto: 1) i cicli produttivi connessi alla cerealicoltura, alla castanicoltura e alla viticoltura, con un'ampia sezione dedicata ai processi di trasformazione agroalimentare dei prodotti provenienti da tali attività; 2) la pastorizia transumante e la produzione casearia; 3) il lavoro dei carbonai nelle diverse fasi di taglio e lavorazione del legname per la produzione del carbone vegetale.

Il terzo e ultimo percorso è di tipo audio-visivo e comprende una raccolta di filmati e brevi documentari di tipo divulgativo sui musei, i siti archeologici e i luoghi di culto dell'Irpinia. Per la visione di questo materiale è stata allestita un'apposita sala proiezione all'interno della *mnemoteca* di Ariano. Tutti i documentari sono stati ideati da noi e girati da videomaker professionisti. Ogni filmato vede inoltre la collaborazione delle persone del posto: guide, direttori e addetti delle diverse strutture, ma anche semplici appassionati locali che compaiono nei vari documentari per descrivere e raccontare nella loro particolare prospettiva interna i diversi luoghi presentati, che hanno sì un interesse storico e culturale generale, ma anche un importante valore identitario per le comunità locali. Il percorso audio-visivo contiene anche un *reportage* sull'antica pratica della transumanza còlta e ripresa nel suo reale svolgersi da Giovanni Abete che ha seguito un gruppo di pastori della valle del Calore e le loro mandrie di bovini podolici nel viaggio verso i pascoli invernali della Puglia.

Parte del materiale fotografico e i testi di carattere divulgativo esposti nella *mnemoteca* di Ariano Irpino e nelle altre sedi sono stati raccolti in un catalogo e in alcune brochure di presentazione del progetto e dei musei.

6. *Conclusioni*

Con l'esperienza del progetto *Voci, parole e testi della Campania* abbiamo cercato di raccordare in maniera innovativa la ricerca dialettologica e sociolinguistica con attività che, anche sfruttando le possibilità offerte dalle nuove tecnologie, promuovano la conoscenza dei beni culturali immateriali dell'entroterra campano. Le fonti orali raccolte nell'archivio, oltre ad avere un importante valore sul piano prettamente linguistico[21], costituiscono infatti il tramite di espressione più autentico della cultura locale, essenziale per il recupero dei suoi aspetti profondi e peculiari. L'archivio si propone dunque come una risorsa a cui è possibile attingere per ricostruire la dimensione culturale delle comunità irpine nei suoi molteplici risvolti, in una chiave non scontata e improntata sulle testimonianze dirette delle persone che sono parte integrante di questa realtà e che vivono concretamente il territorio. Gli etnotesti alla base del nostro racconto devono tuttavia essere accompagnati da ulteriori documenti di approfondimento sul quadro storico, sociale e ambientale al cui interno si è formato, nelle sue diverse fasi di sviluppo, il patrimonio culturale locale. Il lavoro svolto per la creazione delle narrazioni multimediali e delle *mnemoteche* è stato proprio quello di mettere in evidenza le linee di connessione tra i contenuti culturali espressi dalle fonti orali e il loro specifico contesto di riferimento.

Attraverso le diverse strategie di divulgazione sperimentate con il progetto *Voci, parole e testi della Campania*, abbiamo cercato di rispondere alla riconosciuta esigenza di "restituzione" del dato alle comunità locali (Sornicola, 2017a; Matranga, 2018: 595-596), e al contempo di proporre un modello per la comunicazione e per la valorizzazione delle specificità culturali di quei luoghi, come l'Irpinia, meno noti e per lo più estranei alle rotte turistiche tradizionali. Le narrazioni e le *mnemoteche* rappresentano infatti solo una delle possibili modalità di impiego dei materiali contenuti nell'archivio: da questa prospettiva, l'archivio si rivolge non solo agli ambienti della ricerca, ma anche alle stesse comunità locali, alle loro istituzioni, associazioni, enti e piccole aziende, e in generale a chi voglia beneficiare del suo potenziale per iniziative volte alla promozione e al recupero del patrimonio territoriale, soprattutto per ciò che riguarda i beni culturali immateriali.

Il nostro contributo di studiosi è stato dunque un impegno di mediazione per lo sviluppo di una rete di relazioni e di un confronto costruttivo tra gruppi di ricerca e comunità locali: l'archivio e le *mnemoteche* sono stati concepiti infatti come un cantiere aperto, pronto ad accogliere sempre nuove testimonianze, in cui le persone del posto non rappresentano soltanto l'oggetto della ricerca, ma partecipano attivamente all'ideazione di nuove forme di comunicazione del proprio territorio.

Riferimenti bibliografici

Abete, G. (2011). Methodological issues for the study of phonetic variation in the Italo-Romance dialects. In *Travaux neuchâtelois de linguistique*, 53, 111-125.

Abete, G. (2016). La pratica della transumanza nella formazione dello spazio linguistico centromeridionale: problemi e ipotesi di ricerca. In Marcato, G. (Ed.), *Il dialetto nel*

[21] Per esempi di analisi linguistiche basate sui materiali degli archivi messi in rete per il progetto *Voci, parole e testi della Campania* si vedano Como, Milano (1999); Sornicola, Calamai (2014); Abete (2017); De Blasi (2019).

tempo e nella storia. Atti del Convegno Internazionale di Studi, Sappada/Plodn, 2-5 luglio 2015. Bologna: Cleup, 379-386.

Abete, G. (2017). *Parole e cose della pastorizia in Alta Irpinia.* Napoli: Giannini.

Abete, G. (2018). La catarina: storia di una parola simbolo della cultura pastorale in Alta Irpinia. In Stromboli, C. (Ed.), *Parole e cose. Il lessico della cultura materiale in Campania*. Firenze: Franco Cesati Editore, 51-76.

Abete, G., Vecchia, C. (2018). Variabilità degli esiti di -ll- in Irpinia: dettagli fonetici e implicazioni diacroniche. In Antonelli, R., Glessgen, M. & Videsott, P. (Eds.), *Atti del XVIII Congresso Internazionale di Linguistica e Filologia Romanza*, Roma, 18-23 luglio 2016. Strasbourg: ELiPhi, 448-459.

Benaiteau, M. (1986). Il principato Ultra dal 1266 al 1861. In Galasso, G., Romeo, R. (Eds.), *Storia del Mezzogiorno, Vol. V: Napoli capitale e le Province*. Roma: Edizioni del Sole, 331-386.

Berruto, G. (1996). Ancora di cose svizzere. Dialettologia ed etnografia in Ticino: la Valle di Blenio. In *Rivista Italiana di Dialettologia*, 20, 167-173.

Bertinetto, P.M., Calamai, S. (2016). L'unità documentale negli archivi sonori. In Marazzini, C., Maconi, L. (Eds.), *Italiano elettronico. Corpora, archivi testuali, archivi sonori. La Piazza delle Lingue 2014.* Firenze: Accademia della Crusca, 95-112.

Bonfiglio-Dosio, G. (2010). Il sistema archivistico italiano. Archival Organization in Italy. In *Documenta & Instrumenta*, 8, 29-45.

Brancaccio, G. (1991). Trasporti e strade. In Galasso, G., Romeo, R. (Eds.), *Storia del Mezzogiorno, Vol. VIII (1): Aspetti e problemi del medioevo e dell'età moderna*. Roma: Edizioni del Sole, 349-385.

Calamai, S. (2011a). Ordinare archivi sonori: il progetto Gra.fo. In *Rivista Italiana di Dialettologia*, 35, 135-164.

Calamai, S. (2011b). Gli archivi sonori e la conservazione della memoria. In *Rivista Italiana di Dialettologia*, 35, 343-344.

Calamai, S., Bertinetto, P.M. (2012). Per il recupero della Carta dei Dialetti Italiani. In Telmon, T., Raimondi, G. & Revelli, L. (Eds.), *Coesistenze linguistiche nell'Italia pre- e postunitaria.* Atti del XLV Congresso internazionale di studi della Società di Linguistica Italiana, Aosta/Bard/Torino, 26-28 settembre 2011. Roma: Bulzoni, 335-356.

Calamai, S., Bertinetto, P.M. (2014). *Le soffitte della voce. Il progetto Grammo-foni*. Manziana (Roma): Vecchiarelli.

Calamai, S., Biliotti, F. & Bertinetto, P.M. (2014). Fuzzy archives. What Kind of an Object Is the Documental Unit of Oral Archives?. In Ioannides, M., Fink, E., Brumana, R., Patias, P., Doulamis, A., Martins, J. & Wallace, M. (Eds.), *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection.* Proceedings of the 5th International Conference, EuroMed 2014, Limassol, Cyprus, November 3-8, 2014. New York: Springer International Publishing, 777-785.

Canazza, S., Calamai, S., Bertinetto, P.M. & De Dominicis, A. (2012). A protocol for the preservation of speech documents archives: towards the digital curation of the Carta dei Dialetti Italiani. In Ferrari, A. (Ed.), *Proceedings of the 5th International Congress "Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin"*, Istanbul, Turkey, 22nd-25th November 2011. Roma: Valmar, 84-93.

Canobbio, S. (2004). Etnotesto. In Beccaria, G.L. (Ed.), *Dizionario di linguistica e di filologia, metrica, retorica*. Torino: Einaudi, 306.

Como, P. (2006). Elicitation techniques for spoken discourse. In Brown, K. (Ed.), *Encyclopedia of Language and Linguistics*, Second edition, vol. 4. Amsterdam: Elsevier, 105-109.

Como, P., Milano, E. (1999). L'archivio di parlato dei dialetti campani: un esperimento di rappresentazione della variazione linguistica. In Marcato, G. (Ed.), *Dialetti oggi.* Atti del convegno "Tra lingua, cultura, società. Dialettologia sociologica", Sappada/Plodn (Belluno), 1-4 luglio 1998. Padova: Unipress, 123-132.

Corti, L. (2003). *I beni culturali e la loro catalogazione.* Milano: Mondadori.

D'Agostino, M., Ruffino, G. (2005). *I rilevamenti sociovariazionali. Le linee progettuali, Materiali e ricerche dell'Atlante Linguistico della Sicilia* (Vol. 16). Palermo: Centro di Studi filologici e linguistici siciliani.

De Blasi, N. (2019). La parola, il gesto, il significato: «mica» come 'piccola quantità' nella documentazione della «Mnemoteca» di San Mango sul Calore. In Del Puente, P. (Ed.), *Dialetti: per parlare e parlarne.* Atti del quinto Convegno Internazionale di Dialettologia, Potenza, Lauria, Vaglio, 1-3 dicembre 2016. Venosa (PZ): Osanna Edizioni, 31-42.

De Dominicis, A. (Ed.) (2002a). *La voce come bene culturale*. Roma: Carocci.

De Dominicis, A. (2002b). La voce come bene culturale, la linguistica come strumento di catalogazione e restauro. In De Dominicis, A. (Ed.) (2002a), 7-26.

Dentoni-Litta, A. (1999). Il censimento curato dall'Ufficio centrale per i beni archivistici. In *Archivi sonori.* Atti dei seminari di Vercelli (22 gennaio 1993), Bologna (22-23 settembre 1994), Milano (7 marzo 1995). Roma: Ministero per i Beni e le Attività culturali. Ufficio centrale per i beni archivistici, 135-142 (Pubblicazioni degli Archivi di Stato, Saggi 53).

Di Blas, N. (2016). *Storytelling digitale a scuola*. Milano: Apogeo Education.

Di Guglielmo, N. (2002). Pastorizia e transumanza nel quadro dell'economia dell'Alta Irpinia in età moderna. In Ivone, D. (Ed.), *La transumanza nell'economia dell'Irpinia in età moderna*. Atti del convegno di studi, Andretta (AV), 21-22 giugno 2001. Napoli: Editoriale Scientifica, 115-169.

Grimaldi, P., Porporato, D. (2011). I 'Granai della memoria'. Un Percorso didattico e di ricerca. In *Atti del Convegno DIDAMATICA 2011*, Politecnico di Torino, 4-6 Maggio 2011. Torino: Politecnico di Torino, 1-10.

Marino, J.A. (1991). *L'economia pastorale nel regno di Napoli.* Napoli: Guida.

Matranga, V. (2018). Lessico dialettale e dispositivi mobili. Il progetto *BookAlive*. In Retali-Medori, S. (Ed.), *Actes du colloque de lexicographie dialectale et étymologique en l'honneur de Francesco Domenico Falcucci*, Corte-Rogliano, 28-30 Octobre 2015. Alessandria: Edizioni dell'Orso, 595-605.

Paolini, P., Di Blas, N. (2014). Storytelling for cultural heritage. In Contin, A., Paolini, P. & Salerno, R. (Eds.), *Innovative technologies in urban mapping*. Berlin: Springer, 33-48.

Paternostro, G., Sottile, R. (2010). Parlante, identità e (etno)testo fra storia della lingua e dialettologia. L'esperienza dell'Atlante Linguistico della Sicilia. In Ruffino, G., D'Agostino, M. (Ed.), *Storia della lingua italiana e dialettologia*. Palermo: Centro di studi filologici e linguistici siciliani, 597-613.

Retaro, V. (2015). Digitalizzazione e messa in rete dell'archivio *ADiCa*: gli standard di descrizione archivistica per la dialettologia. Manoscritto, Università degli Studi Napoli "Federico II", materiali di lavoro del progetto *Voci, parole e testi della Campania*.

Retaro, V., Abete, G. (2018). Sull'importanza delle aree intermedie: i dialetti del Vallo di Lauro. In Antonelli, R., Glessgen, M. & Videsott, P. (Eds.), *Atti del XVIII Congresso Internazionale di Linguistica e Filologia Romanza*, Roma, 18-23 luglio 2016. Strasbourg: ELiPhi, 957-968.

Romito, L., Lio, R. (2007). I corpora vocali come patrimonio culturale. Archiviazione, conservazione e restauro di materiale sonoro calabrese. In *Atti del III Convegno Nazionale AISV "Scienze Vocali e del Linguaggio Metodologie di Valutazione e Risorse Linguistiche"*, Trento, 29/11-1/12/2006. Torriana (RN): EDK Editore, 399-414.

Sornicola, R. (1999). La variazione dialettale dell'area costiera napoletana. Il progetto di un archivio di testi dialettali parlati. In Marcato, G. (Ed.), *Dialetti oggi.* Atti del convegno "Tra lingua, cultura, società. Dialettologia sociologica", Sappada/Plodn (Belluno), 1-4 luglio 1998. Padova: Unipress, 103-122.

Sornicola, R. (2017a). Small islands, global worlds: aspetti linguistici e storico-culturali delle isolee flegree. In Aveta, A., Marino, G.B. & Amore, R. (Eds.), *La Baia di Napoli. Strategie integrate per la conservazione e la fruizione del paesaggio culturale*. Napoli: artstudiopaparo, 437-441.

Sornicola, R. (2017b). The outcomes of the geminate lateral in the South of Italy and their conservation in certain types of linguistic community. In *Sociolinguistic Studies*, 11/2-3-4, 413-433.

Sornicola, R., Calamai, S. (2014). Sound archives and linguistic variation. The case of the Phlegraean diphthongs. In Calamai, S., Celata, C. (Eds.), *Advances in Sociolinguistics*. Amsterdam: Benjamins, 169-185.

Valenti, F. (2000). *Scritti e lezioni di archivistica, diplomatica e storia istituzionale* (a cura di Grana, D.). Roma: Ministero per i Beni e le Attività culturali, Ufficio Centrale per i Beni Archivistici (Pubblicazioni degli Archivi di Stato, Saggi 57).

Vecchia, C. (2017). La variazione fonetica degli esiti di -ll- in Irpinia. Processi di rotacizzazione e di retroflessione nelle varietà dell'alta valle del Calore. Tesi di dottorato, Università degli Studi di Napoli "Federico II".

Vecchia, C. (2018a). Elementi innovativi nel dialetto di Bagnoli Irpino (Av). In Marcato, G. (Ed.), *Dialetto e Società*. Padova: Cleup, 111-117.

Vecchia, C. (2018b). Dialetti e comunità dell'alta valle del Calore. Profilo linguistico e socio-culturale dei centri dell'Irpinia centro-meridionale. In *Rendiconti dell'Accademia di Archeologia Lettere e Belle Arti in Napoli*, N.S., 80, 213-224.

DOI: 10.17469/O2106AISV000006

MARIA FRANCESCA STAMULI

Fonti orali, documenti e archivi: riflessioni e proposte per la nascita di un 'archivio vivo'

In Italy, the relation between the terms 'archive', 'record', and 'oral source' has still not been fully explored. This paper discusses about these terms starting with the understanding of their meaning and use in the Italian archival studies and in research areas based on fieldwork producing, using and preserving records related to oral sources.

Key words: oral source, record, archives, field studies.

1. *Introduzione*

In queste pagine vorrei proporre una riflessione teorica sulla nozione di 'fonte orale' e la costellazione terminologica ad essa correlata nell'assetto disciplinare degli studi che a tale nozione fanno ricorso (par. 1. *Assetto disciplinare*). Si proporrà quindi un approfondimento degli assunti fondanti la nozione di 'fonte orale' in ambito archivistico, dei limiti di tali assunti e delle possibili virtuosità (par. 2. *L'archivistica e le 'nuove fonti'*). Si individueranno, quindi, le metodologie di trattamento archivistico della documentazione legata a tali fonti e le ricadute che tale trattamento potrebbe avere in termini di 'usabilità' e accessibilità da parte delle discipline ad esse interessate (par. 3. *Archivi e fonti orali*).

1.1 Assetto disciplinare

Le 'fonti orali' sono considerate fonti primarie e imprescindibili per le discipline umanistiche che si fondano sul metodo della ricerca sul campo: la linguistica, l'antropologia e la storiografia, soprattutto nella configurazione assunta dal più recente ambito della *oral history*.

Tali discipline rilevano sin dalla fine del XIX secolo, nell'ambito di una riflessione metodologica necessaria, costante e non di rado conflittuale, gli elementi critici della individuazione delle 'fonti orali': il loro reperimento, la loro costituzione, il loro uso. La 'non neutralità' dei rilievi, i limiti riscontrati nelle diverse tecniche adottate (questionario strutturato e traduttivo, osservatore partecipante, *tasks* in cuffia, etc.) e i diversi scopi con i quali le fonti sono rilevate, osservate e conosciute, sono stati elementi centrali nella riflessione su alcune delle loro caratteristiche, sia in seno alla specifica disciplina di riferimento sia in seno a riflessioni di ambito interdisciplinare[1]. Tra questi nodi, la 'neutralità' rispetto al posizionamento di intervistatore/intervistato, l'autenticità rispetto alla fedeltà del dato che si rileva e l'estensione d'uso possibile delle 'fonti orali' sono probabilmente quelli di maggior rilievo rispetto alla valutazione del loro livello di scientificità e, conseguentemente, delle discipline che ne fanno uso.

L'assetto disciplinare assunto dagli studi linguistici, antropologici e storiografici nel quadro della ricerca scientifica italiana nel secolo scorso non ha aiutato a sciogliere tali questioni. Mi riferisco, in particolare, alla scissione tra ricerca linguistica e antropologica che, iniziata già nel primo decennio del '900 con la nascita dell'etnologia, si consumava definitivamente negli anni '60[2]. Probabile ricaduta della critica all'estetica crociana e, al contempo, effetto di una tardiva acquisizione dei modelli strutturalisti, da una parte, e della moderna sociolinguistica dall'altra, le due discipline per un importante pezzo del XX secolo non sono state più dialoganti. Ciò ha determinato una progressiva divaricazione anche rispetto alla possibile trasversalità delle 'fonti orali' e ha portato a trattare come 'distinti' gli 'oggetti' raccolti sul campo nel corso di ricerche afferenti alle diverse discipline. Anche quando, saltuariamente e raramente, tali discipline si sono incontrate per confrontarsi sulla possibilità di un loro uso comune, le riflessioni metodologiche sembrano ambire a rivendicare i diversi ambiti d'azione e le diverse tipologie di analisi cui sottoporre il testimone e/o il documento prodotto dal testimone (cfr. Leydi, Telmon, 1999). Questo ha probabilmente contribuito a determinare una 'diaframmatizzazione' delle 'fonti orali', un processo, cioè, in base al quale oggetti prodotti in ricerche afferenti a un ambito disciplinare restano sconosciuti o poco utilizzati per un altro: si pensi, per esempio, al *Catalogo delle tradizioni orali non cantate* curato da Alberto Mario Cirese e Liliana Serafini e pubblicato nel 1975 dall'allora Discoteca di Stato (attualmente Istituto Centrale per i Beni Sonori e Audiovisivi), repertorio della campagna di raccolta di fiabe, proverbi e miti promossa dallo stesso istituto tra il 1968 e il 1972 e poco noto ai linguisti. Di converso, le grandi raccolte afferenti alle esperienze degli atlanti sono pochissimo note agli antropologi culturali.

Tale diaframmatizzazione, la cui origine è probabilmente di natura scientifico-disciplinare, è ulteriormente rafforzata da due aspetti:

– l'assetto legislativo italiano, per il quale la tutela delle lingue è tra le materie a legislazione concorrente, diversamente dai cosiddetti 'beni demoetnoantropologici' che, a partire dal testo unico del 1999, figurano tra i beni culturali sottoposti a tutela dello Stato;

[1] La bibliografia sulla problematicità delle 'fonti', del loro uso e riuso è notevole. Si riportano, qui, solo pochi riferimenti: alcuni di ordine generale per le singole discipline (Magnani, 2018; Bacchiddu, 2012; Contini, 2006; Matranga, 2002; Carucci, 1993) altri di ordine transdisciplinare (Leydi, Telmon, 1999; Bernardi, Poni & Triulzi, 1978).

[2] Si veda, al riguardo, Cirese (1996). L'intera tematica va sicuramente approfondita e inquadrata nella più ampia dialettica tra la ricezione delle scienze positive in Italia e l'impianto disciplinare ed ermeneutico dell'estetica crociana, impianto determinante per la conformazione assunta sia dagli studi dialettologici che da quelli demologico- folklorici in cui linguistica e antropologia si radicano.

– il ritardo di una terza disciplina, l'archivistica, nel trattamento e nella definizione delle fonti orali.

Tralasciando per ora il primo, approfondirò il secondo elemento nel paragrafo successivo.

2. *L'archivistica e le 'nuove fonti'*

Potenzialmente, l'approccio archivistico alle fonti orali permetterebbe di affrontarne la specificità con un posizionamento, in rapporto alle discipline che di queste fonti fanno uso, di tipo sovra-disciplinare. L'archivistica, infatti, si occupa proprio di conservare e, quindi, di rendere accessibili le fonti senza distinzione rispetto agli scopi finali che possono essere di volta in volta amministrativi di riferimento 'interno' rispetto all'attività nella quale esse sono prodotte o di ricerca storica in senso più ampio. Rispetto alle fonti orali però, la disciplina presenta, soprattutto in Italia, un'incertezza metodologica notevole, determinata dal ritardo con cui esse sono rientrate nell'ambito disciplinare di sua pertinenza.

L'archivistica nasce nel XIX secolo. Erede degli studi eruditi e della diplomatica del secolo dei lumi, si forma sulla documentazione conservata negli archivi istituzionali, e in particolare sulla necessità di garantirne l'autenticità per esigenze sia di fede pubblica che di ricostruzione della memoria storica. È, quindi, una disciplina relativamente giovane che ha allargato lentamente, e con modalità non di rado conflittuali, i suoi ambiti applicativi. In Italia, per esempio, solo negli anni '60 del secolo scorso l'archivistica si è aperta alla trattazione dei cosiddetti 'archivi culturali' (come quelli delle aziende e delle persone), archivi, cioè, nei quali i documenti sono conservati da chi li produce per proprio riferimento e memoria e che, in tal senso, sono destinati a usi non strettamente amministrativi, ma di natura scientifico-culturale.

Le fonti orali, quindi, sono tra le ultime pervenute alle attenzioni degli archivisti e sono state, non a caso, ricondotte al gruppo delle cosiddette 'nuove fonti' (René-Bazin, 1988; Carucci, 1993).

La documentazione a 'fonte orale', in effetti, presenta caratteristiche di segno opposto rispetto alla documentazione istituzionale di cui tradizionalmente l'archivistica si è occupata, e sulla quale la disciplina si è formata metodologicamente. La fonte orale, infatti, spesso non prende forma in un documento scritto e strutturato in senso diplomatico, con caratteristiche formali sia testuali che materiali altamente codificate; non è solitamente veicolata da un documento formato e 'tenuto' in un archivio dotato di strumenti di autenticazione come il registro di protocollo, o di tenuta come inventari e segnature. Essa prende forma, piuttosto, in un documento che vuole fissare atti performativi, solitamente di carattere prevalentemente linguistico (e solitamente non dotati di valore legale) e fortemente legati al contesto comunicativo e ad altre forme di espressività corporea; si configura come un documento non di fede pubblica ma di riferimento interno di chi lo 'raccoglie' fissandolo in una determinata forma, spesso poco o nulla codificata.

Tali aspetti, così fortemente estravaganti rispetto a quelli delle 'classiche' fonti documentarie, accompagnati dal dato che le fonti orali, al momento del loro ingresso tra le 'nuove fonti' dell'archivistica si presentavano mediati da supporti tecnologici nuovi al mondo degli archivi, hanno determinato una notevole fluidità terminologica. Forme come 'documento sonoro', 'registrazioni', 'fonti sonore e audiovisive' sono spesso utilizzate in archivistica come sinonimo di 'fonte orale' o, al massimo, come sue sottocategorizzazioni.

Ancora nel 1993, quando Paola Carucci fu chiamata a scrivere la prefazione alla pubblicazione dei risultati del primo censimento degli istituti di conservazione delle fonti orali[3], la fluidità della nomenclatura e delle classificazioni adottate erano tali da rendere i risultati stessi del censimento non completamente affidabili:

> [...] i dati sul materiale archivistico, presente in 75 casi, sono spesso tra loro non comparabili e talora risultavano incomprensibili nelle schede di rilevazione (Carucci, 1993: 22).

In questa sede, quindi, è forse opportuno partire proprio dalle definizioni archivistiche di fonte orale e dalle criticità eventualmente innescate.

2.1 Le fonti orali: medium e contenuto

Le definizioni riscontrate in archivistica fanno perno sull'opposizione 'fonti orali' *versus* 'fonti scritte':

> L'espressione «fonti orali» si contrappone all'espressione «fonti scritte» (Carucci, 1993: 14);

> le fonti orali sono tutte le possibili tracce su banda magnetica lasciate da tutte le possibili voci, registrate per le ragioni e nei contesti più vari (Contini, 2006: 799).

Sempre in Contini (2006), si nota anche una distinzione tra fonti orali e fonti audiovisive che

> differiscono da quelle orali solo perché fermano, oltre al suono della voce, anche l'immagine del parlante e dell'ambiente che lo circonda (Contini, 2006: 795).

Molto più articolata la definizione fornita da Paola Carucci, sempre nel 1993, che, come vedremo, faceva faticosamente emergere le criticità che tale opposizione poneva. La Carucci, infatti, articola la definizione di 'fonte orale' sotto vari aspetti (fisico, concettuale, descrittivo), proprio per la difficoltà individuativa di una definizione che facesse perno sul solo supporto o "aspetto fisico":

> Sotto *l'aspetto fisico* [corsivo mio, n.d.r.] – si è detto – le fonti orali oggi si presentano in prevalenza sotto forma di registrazioni sonore e audiovisive, mezzi cioè utilizzati per le realizzazioni di una pluralità di fonti concettualmente diverse [...]. Sotto l'aspetto fisico dunque la fonte orale rientra nel più vasto ambito delle registrazioni sonore e degli audiovisivi. [...] Sotto *l'aspetto concettuale* invece la fonte orale – in considerazione del suo processo di formazione – mantiene delle caratteristiche proprie, ancorché non univocamente codificate, che ne fanno una tipologia distinta, e

[3] Si veda: Barrera, Martini & Mulè (1993).

> a sua volta passibile di ulteriori articolazioni, nel più generale quadro delle fonti sonore e audiovisive di cui è necessario tener conto ai fini di una adeguata descrizione. [...] Sotto *l'aspetto della descrizione*, infine, le fonti orali si pongono in termini diversi rispetto alle fonti musicali, ai prodotti di editoria multimediale, ai documentari o ai film a soggetto o ai programmi televisivi, i quali inoltre si distinguono in una pluralità di prodotti di natura differenziata (telegiornali, dibattiti, varietà, film, inchieste, pubblicità, eccetera). Nell'ambito di questa varia e articolata tipologia l'attenzione degli archivisti si focalizza sulla conservazione di quei prodotti che presentano caratteri assimilabili a quelli delle fonti documentarie (Carucci, 1993: 16-17).

La definizione archivistica di fonte orale pone quindi un forte accento sulla tipologia del supporto a scapito di altri elementi essenziali, come il contesto di formazione e le unità informative. Tale assunto va probabilmente indagato meglio nella storia degli studi della disciplina archivistica, radicata nella diplomatica e di conseguenza nel riconoscimento della necessaria consonanza tra forma e contenuto per definire il documento nelle sue caratteristiche di originalità e autenticità. Ci interessa, però, qui sottolineare due elementi di crisi legati a tale posizionamento teorico.

Il primo elemento è costituito dall'inefficacia individuativa della definizione, con pesanti ricadute di ordine tassonomico. Da una parte le fonti orali sono accomunate a una congerie di oggetti che, nella produzione culturale, sono affidati al *medium* audio-visivo: come si evince dalla definizione qui in analisi, per esempio, alcuni degli oggetti elencati (film, concerti *live*, rappresentazioni teatrali, documentari, *et similia*) non possono essere considerati come documenti d'archivio (se non in specifiche condizioni), poiché i documenti d'archivio nascono, come vedremo, con precise caratteristiche di contesto.

Dall'altra, dalle fonti orali sono esclusi una serie di documenti che sono individuabili come 'fonti orali' ma che non sono mediati dalla tecnologia audiovisiva ma dalla scrittura: le trascrizioni. Per i linguisti e gli antropologi la trascrizione è un documento che media una fonte orale, per fonte orale intendendo le testimonianze fornite da vive voci all'intervistatore in un determinato momento e contesto comunicativo[4] e fissate (o nuovamente mediate) nelle forme della scrittura sotto forma di trascrizione. Su queste trascrizioni sono nate e si sono fondate metodologicamente tali discipline e a esse hanno fatto riferimento come 'fonti orali' almeno fino all'invenzione e alla diffusione dell'uso di apparecchiature di registrazione della voce e dell'immagine nelle ricerche 'etnologiche' e 'linguistiche' di rilievo sul campo. Ed è, d'altra parte, proprio da queste discipline che è arrivata la chiara consapevolezza di dover riconoscere (e trattare) tali fonti come documenti autonomi. Già nel 1986, infatti, l'antropologo Pietro Clemente rivendicava alle fonti orali la loro condizione 'documentaria':

> Le fonti orali sono documenti, e ogni documento comporta un problema di conservazione e ordinamento [...] il documento orale diventa finalmente un documento come tanti altri (Clemente, 1986: 190).

[4] Sulla tensione tra intervistato e intervistatore come compartecipazione conflittuale nella costituzione o rappresentazione della 'fonte orale' si vedano Contini (2006), Portelli (2014).

Il secondo, che è anche conseguenza del primo, è costituito dalla difficoltà di trattare archivisticamente la fonte orale e quindi di definirla in riferimento al contesto di formazione (che si caratterizza solitamente come un contesto di ricerca legato a determinate discipline che fanno uso di metodologie di rilievo sul campo).

Occorre, quindi, sottrarre la definizione di 'fonte orale' alla descrizione del suo *medium* e ricollocarla, come la stessa Paola Carucci suggeriva già nel 1993, nella dimensione storica del suo "processo di formazione". Occorre, cioè, adottare in archivistica una definizione di fonte orale che individui nelle modalità di raccolta e/o recupero di quello specifico contenuto informativo la sua costituzione e concrezione in un determinato (ma variabile) *medium,* ovvero di uno o più documenti formati nell'ambito di una determinata attività di rilievo.

In questo modo, si rende possibile anche la piena (e necessaria) individuazione della fonte orale in diversi documenti d'archivio che, in quanto tali, sono collegati tra loro e pienamente contestualizzati dall'attività in cui prendono forma.

2.2 Le fonti orali e il vincolo archivistico

Occorre quindi attribuire in prima istanza alle 'fonti orali' la loro dimensione di 'documento' in quanto 'rappresentazione di un atto o fatto memorizzata su un supporto di qualunque natura, comunque formata o acquisita da un soggetto produttore nel corso della sua attività e conservata da una persona fisica o da una persona giuridica per proprio o altrui riferimento'[5].

La trattazione archivistica dei documenti legati alle fonti orali ha diverse ricadute: la prima è che ne determina necessariamente la loro collocazione in un contesto archivistico poiché è nel contesto archivistico che il documento esiste e sussiste. L'archivio è, infatti,

> un insieme di documenti formati (creati) e ricevuti da una persona fisica o giuridica come strumento e residuo della propria attività istituzionale, professionale o creativa e conservati per 'riferimento' proprio o 'pubblico' (ovvero come fonte informativo-autoriale, n.d.r.) (Duranti, 2015: 21)[6]

Nel contesto archivistico, quindi, i documenti si formano, si stratificano organicamente e vivono come fonti informative autonome e dalle caratteristiche specifiche. In particolare, nell'archivio il documento si ordina naturalmente rispetto all'attività (nel caso delle fonti orali, tale attività si qualificherà, con ogni probabilità come

---

[5] Cfr. Jenkinson (1922); Carucci, Guercio (2008). Una definizione molto simile di documento, che sottolinea la possibilità e l'opportunità che il *medium* sia diverso dalla scrittura, è peraltro utilizzata, anche nelle recentemente novellate "Regole deontologiche per la ricerca storica" art. 2 c.1, lett. *c* dove si individua come documento "qualunque testimonianza scritta, orale o conservata su qualsiasi supporto che contenga dati personali.", cfr. par. 3.2.

[6] Si veda anche la definizione del fr. *archive* nel Multilingual Archival Terminology dell'International Council of Archives: "Documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité. Le mot archives est couramment employé dans le sens restrictif de documents ayant fait l'objet d'un archivage, par opposition aux archives courantes".

attività di ricerca di un gruppo o di un singolo ricercatore afferente alle discipline sopradette) e si determina solo in relazione agli altri documenti.

Uno dei cardini fondanti l'epistemologia archivistica, e la sua metodologia, è il 'vincolo archivistico', "la relazione che ogni documento archivistico ha con i documenti archivistici che partecipano alla stessa attività e si accumulano nel corso delle funzioni del produttore" (Duranti, 2015: 24); il legame, cioè, che mette in relazione i documenti tra loro e con la prassi che organicamente li produce. Il vincolo archivistico[7] è quindi caratteristica originaria e necessaria del documento d'archivio, poiché è solo in relazione agli altri che il documento archivistico assume il suo specifico significato, significato di cui la relazione è parte. Tale relazione è, quindi, anche quel fattore indispensabile a determinare la possibilità che singoli oggetti e segmenti informativi siano individuabili, accessibili e 'autentici' rispetto all'informazione che veicolano.

In questo senso, le fonti orali non potrebbero essere sottratte alla propria dimensione di 'fonte documentaria', dimensione che è pienamente significativa solo in rapporto agli altri documenti con cui si costituiscono in archivio.

Il riconoscimento della dimensione archivistica delle fonti orali eviterebbe, quindi, vere e proprie operazioni di 'smembramento' di archivio, operazioni che hanno portato le registrazioni ad essere custodite in collezioni tipologiche avulse dal contesto di ricerca[8]. La dimensione archivistica, inoltre, rafforzerebbe la percezione che "le testimonianze raccolta su banda magnetica in forma di voce", per citare Pietro Clemente, sono il primo momento di formazione della 'fonte orale', prodotto centrale di una determinata attività di ricerca e non uno strumento di lavoro 'interno', ad uso momentaneo e strumentale e, per questo, non destinate a conservazione di lungo periodo. Non raramente in passato, visti anche i costi dei supporti, i ricercatori, i gruppi di ricerca (si veda, *infra,* l'esperienza dell'Atlante Lessicale Toscano) o persino organizzazioni, come la Oral History Research della Columbia University hanno cancellato le registrazioni conservandone solo la trascrizione, invertendo, cioè, il rapporto valoriale tra formato documentario 'originale' e 'copia'.

La dimensione archivistica, infine, renderebbe possibile un'efficace e necessaria descrizione delle 'fonti orali' nelle caratteristiche formali in cui si concretano che, come abbiamo visto, sono estravaganti, andando a sottolineare, in che modo (*medium*) una determinata informazione si concretizza e si media (se in un documento scritto, documento scritto, illustrato, documento sonoro, documento cinetico filmico, *et alia*).

3. *Archivi e fonti orali*

Come ogni archivio, quindi, anche quello che si forma attraverso la costituzione di documentazione che presenti fonti orali va strutturato in base alle funzioni svolte dal soggetto produttore.

---

[7] Per la piena articolazione della nozione di vincolo archivistico si veda Cencetti (1939).

[8] Si pensi a collezioni come quelle di Gilberto Giuntini, costituite dalle sole bobine e conservate presso la Fondazione Lavoratori Officine Galileo senza nessuna nota contestuale, o alle registrazioni di Diego Carpitella, conservate presso la Biblioteca di Arezzo.

L'ordinamento della produzione documentaria in base alle funzioni del soggetto produttore è effettuato predisponendo a monte un 'quadro' o 'piano di classificazione' definibile come una struttura di voci, articolata logicamente e gerarchicamente, che individua in modo astratto, ai fini dell'ordinata e organica sedimentazione dei documenti, attività e funzioni svolte dall'ente o materie a esso attribuite. In base a tale struttura si 'classifica', si va, cioè, a raggruppare logicamente per materie e funzioni la documentazione che si produce o riceve svolgendo un determinato tipo di attività.

Nel caso di riordino di archivi a valle della loro formazione, il quadro è ipotizzato a partire dai documenti stessi e dagli strumenti in cui si stabiliscono le funzioni svolte dall'ente o comunque dal soggetto produttore.

L'archivio sarà quindi ordinato e inventariato in fondi, serie, fascicoli e documenti che rispecchino le attività e le funzioni che hanno determinato la formazione di quel complesso documentario.

A titolo esemplificativo, possiamo ipotizzare di essere un gruppo di ricerca o un ricercatore che, per ambito disciplinare e metodo, abbia necessità di reperire e formare 'fonti orali'. Dovremmo aspettarci, un quadro come quello ipotizzato in Tabella 1:

Tabella 1 - *Esempio di quadro di classificazione*

I. Attività amministrativa	A. Statuto del gruppo di ricerca	
	B. Attività di finanziamento	
	C. Organigramma	
II. Attività scientifica	A. Metodologia	1. Quadro teorico 2. Linee guida 3. Apparato tecnologico
	B. Campagne di rilievo: luogo	1. Registrazioni audio 2. Registrazioni video 3. Trascrizioni 4. Annotazioni 5. Fotografie
	C. Produzione scientifica	1. Pubblicazioni 2. Comunicazioni 3. Siti web 4. Cataloghi

I documenti prodotti o ricevuti saranno opportunamente classificati in base al quadro. Ogni documento sarà a sua volta corredato di una descrizione minima costituita da datazione, autore, oggetto, tipo di supporto, eventuali allegati o rimandi ad altri documenti e, soprattutto, la segnatura, ovvero l'identificativo univoco costituito possibilmente dagli elementi logici del piano di classificazione: per esempio, seguendo la tabella 1, una potenziale segnatura corrispondente a un documento audio potrebbe essere 'II.B.1.1.1'. Nel caso delle interviste, in particolare, sarebbe opportuno annotare il nome dell'intervistatore, nel caso in cui è diverso dall'autore, e il/gli intervistati, nonché ricondurre l'oggetto a eventuali questionari e tematismi individuati in base alla metodologia adottata dal progetto.

L'archivio che si viene così a creare potrà poi essere reso accessibile da strumenti di accesso (guida e/o inventario) e i cosiddetti 'apparati di corredo'. Anche qui, immaginando un gruppo di ricerca o un ricercatore, potremmo/dovremmo aspettarci biografie dei testimoni, biografie degli intervistatori, indice dei luoghi, indice degli argomenti o tesauro, bibliografia[9].

3.1 Gli standard descrittivi

Gli elementi descrittivi dell'archivio e dei singoli documenti che sono stati appena enumerati andranno individuati e forniti in base agli standard archivistici individuati nazionalmente e internazionalmente.

In particolare, gli standard di contenuto da seguire sono: ISAD (G) per la descrizione degli archivi[10]; ISAAR (CPF)[11] per le intestazioni di enti, persone e famiglie, che è interoperabile con le NIERA (EPF)[12] ovvero le "Norme italiane per l'elaborazione dei record di autorità archivistici di enti, persone, famiglie". Gli standard di contenuto vanno ovviamente accompagnati agli standard di struttura, volti alla descrizione e alla possibilità di condivisione dei formati, quali EAC[13] e EAD[14] nelle versioni più aggiornate. A tali standard andranno, con ogni probabilità, affiancati anche quelli di gestione degli oggetti digitali e di conservazione, tanto del sistema di gestione quando degli oggetti in esso contenuti: framework METS[15], MAG[16] e DC[17] per la descrizione delle risorse digitali e strutture del tipo OAIS[18] per la conservazione degli archivi digitali.

Oggetto di standardizzazione dovrebbe essere, vista la tipologia di fonti, anche quello relativo ai tematismi. L'adozione di linguaggi con tesauri che presentano una base terminologica modellata in XML/SKOS possono produrre descrizioni semantiche interoperabili nel web semantico. Si pensi, per esempio al thesaurus multi-

---

[9] Sebbene datato, un utile strumento alla catalogazione delle fonti orali può essere l'*Oral History Cataloguing Manual* di Marion Matters (cfr. Matters, 1995).

[10] Per gli standard, si veda in generale la pagina dedicata sul sito dell'Istituto Centrale degli Archivi, articolazione del Ministero dei Beni e delle Attività Culturali deputata, in accordo con gli organisimi internazionali, all'individuazione ed emanazione degli standard archivistici in Italia: http://www.icar.beniculturali.it/index.php?id=107. Per ISAD in particolare, si veda http://www.icar.beniculturali.it/index.php?id=54.

[11] *Ibidem*: http://www.icar.beniculturali.it/index.php?id=55.

[12] *Ibidem*: http://www.icar.beniculturali.it/index.php?id=102.

[13] *Ibidem*: http://www.icar.beniculturali.it/index.php?id=58.

[14] *Ibidem*: http://www.icar.beniculturali.it/index.php?id=57.

[15] *Ibidem*: http://www.icar.beniculturali.it/index.php?id=59.

[16] Si veda il documento pubblicato dall'ICCU nel 2006, in particolare i parr. 8. *Sezione audio* e 9. *Sezione video*.

[17] Si veda la pagina in *Dublin Core Metadata Initiative:* https://www.dublincore.org/specifications/dublin-core/.

[18] Si veda http://www.oais.info/.

disciplinare sviluppato dalla Biblioteca Nazionale Centrale di Firenze nell'ambito del Nuovo Soggettario[19], utile alla creazione di sistemi pre-coordinati (i cataloghi per soggetto) e di strumenti di ricerca post – coordinata che fanno uso dei soli descrittori per marcare i documenti con parole chiave, raccordandosi tanto ai record descrittivi quanto agli oggetti digitali descritti[20].

3.2 Pubblicazione e accessibilità delle fonti orali: le nuove "Regole deontologiche per il trattamento a fini di archiviazione nel pubblico interesse o per scopi di ricerca storica"

Affinché le fonti orali siano accessibili alla ricerca storica occorre che già in fase di formazione e raccolta siano rispettate le regole per l'accesso alle fonti. È del 15 gennaio la pubblicazione in Gazzetta della delibera n. 513/2018 del Garante della Privacy con la quale si aggiornano, recependo il Decreto Legislativo n. 101 del 10 agosto 2018 "Disposizioni per l'adeguamento della normativa nazionale al regolamento relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali", le "Regole deontologiche per il trattamento a fini di archiviazione nel pubblico interesse o per scopi di ricerca storica"[21]. Rispetto alla materia qui d'interesse, il Decreto Legislativo n. 101 del 2018 disponeva che il Garante formulasse una serie di Regole che andassero ad emendare il Codice di deontologia e di buona condotta per la ricerca storica, laddove incompatibile con il regolamento generale sulla protezione dei dati (in inglese *General Data Protection Regulation* e meglio noto negli acronimi RGDP o GDPR).

Tra le disposizioni ritenute incompatibili dal Garante, e che sono quindi attualmente annullate, troviamo proprio quella all'art. 8 del "Codice di deontologia e buona condotta per il trattamento dei dati personali per scopi storici" (allegato A2 al D. Lgs. 196 del 2003 "Codice della Privacy") e riferita alle fonti orali. Tale disposizione consentiva al titolare del trattamento dei dati personali di fornire un'informativa semplificata in caso di fonti orali.

Le nuove regole, invece impongono che gli intervistati abbiano espresso il proprio consenso in modo esplicito, eventualmente anche in forma verbale. Ugualmente, gli archivi che acquisiscono fonti orali, devono richiedere all'autore dell'intervista una dichiarazione scritta dell'avvenuta comunicazione degli scopi perseguiti con la registrazione e del relativo consenso manifestato dagli intervistati.

Parimenti è stato reputato non conforme l'esonero all'obbligo di fornire l'informativa agli interessati nei casi di raccolta di dati personali presso soggetti terzi, quando ciò risulti impossibile o comporti uno sforzo sproporzionato (art. 1 c. 5 del "Codice di deontologia e buona condotta per il trattamento dei dati personali per scopi storici"). Occorre, quindi, che le 'fonti orali' siano sempre accompagnate dal consenso al trattamento dei dati personali da parte dell'intervistato e, al momento

[19] Si veda la relativa pagina dell'ICCU https://thes.bncf.firenze.sbn.it/.

[20] Esistono programmi che permettono di ordinare in modo standardizzato ed esportabile archivi con documenti di tipo diverso: l'open source X-Dams, tra questi, ha un'apposita estensione per la descrizione di documenti audio e audiovideo, cfr. https://www.xdams.org/xdams/xdams-cose/.

[21] Si veda http://www.gazzettaufficiale.it/eli/id/2019/01/15/19A00178/sg/Accessed 26/04/ 2019.

dell'acquisizione in archivio, dalla dichiarazione scritta dell'autore dell'intervista di aver richiesto e ottenuto il consenso al trattamento. I dati personali trattati per scopi storici, infatti, possono essere ulteriormente utilizzati per tali scopi e restano soggetti alla medesima disciplina (art. 3, c. 4).

Restando nell'ambito dell'accessibilità e del trattamento delle fonti documentarie per scopi di ricerca storica si rammenta che, in generale, l'accesso agli archivi pubblici e agli archivi privati di straordinario interesse culturale è libero. Fanno eccezione i documenti riservati, per i quali l'utente può richiedere al Ministero degli interni l'autorizzazione alla consultazione anticipata presentando all'ente che conserva i documenti un progetto di ricerca che, in riferimento ai documenti riservati, ne illustri le finalità di ricerca e le modalità di diffusione, che dovranno evidentemente essere rispettose delle prescrizioni contenute nell'art. 11 "Diffusione dei dati personali per ricerca storica del Codice di deontologia e buona condotta per il trattamento dei dati personali per scopi storici". L'autorizzazione è concessa previo parere del direttore dell'Archivio di Stato o della Soprintendenza archivistica competente (art. 10, ibidem) e la documentazione può essere resa consultabile dall'archivista con le 'dovute cautele': (anonimizzazione, pseudonimizzazione, estrazione di fascicoli, oscurazione di parti utili a rilevare l'identità; art. 10, c.7).

È chiaro che se si volesse pubblicare l'archivio nella sua interezza occorrerà riflettere, rispetto ai contenuti, su cosa in base alla legge va sottratto alla consultazione e su quale licenza d'uso assumere rispetto ai diritti d'autore. Nel caso di archivi volti alla ricerca storica, si opta solitamente per una licenza d'uso, rispetto ai diritti d'autore, del tipo Creative Commons 4.0 Non Commercial Share Alike (BY-NA-SA). È fondamentale che alle risorse digitali pubblicate siano attribuiti identificativi persistenti (standard di definizione sono ARK, DOI, PURL, Handle).

4. *Conclusioni*

Dalle note sin qui proposte si configura la necessità di trattare le fonti orali come tutte le altre fonti della ricerca scientifica, la cui documentazione è resa accessibile e affidabile dal trattamento archivistico, che ne dovrebbe garantire, peraltro, l'autenticità e la conservazione. In questo senso, i documenti che veicolano e rappresentano la fonte orale vanno considerati come tutti gli altri prodotti documentari formati durante una determinata attività di ricerca, creativa o amministrativa.

Nel contesto determinato e presupposto dalla buona tenuta archivistica dei documenti, le fonti orali si presenteranno dotate di un apparato informativo necessario alla loro efficace individuazione e utilizzazione. La buona tenuta archivistica, come abbiamo visto, prevede alcuni elementi fondamentali, riassumibili in: 1. ordinamento logico dei documenti, ovvero classificazione della documentazione in base alle funzioni del gruppo/ricercatore finalizzata a ottenere e mantenere l'organica contestualizzazione delle fonti; 2. descrizione normalizzata delle intestazioni e dell'archivio in 'serie archivistiche' nel rispetto di un eventuale quadro di classificazione o comunque restituendo la documentazione, ove possibile, alle attività che ne hanno determinato la formazione, collocandole quindi nel tempo e nei legami rispetto agli altri documenti; 3. descrizione standardizzata dei documenti audio e audio-video come documenti d'archivio, ovvero come singole unità informative; 4. conoscenza e uso degli standard di metadati maggiormente utili alla interoperabilità delle informazioni; 5. rispetto delle regole deontologiche per la ricerca storica; 6. collaborazione con le istituzioni deputate alla tenuta e alla valorizzazione degli archivi.

Al riguardo vale forse la pena rammentare i caposaldi dell'amministrazione archivistica nel nostro Paese. Gli archivi degli enti pubblici sono considerati beni culturali *ab origine,* e per questo sono sottoposti alle misure previste dal Codice dei Beni Culturali in termini di tutela e valorizzazione. Sono sottoposti agli stessi vincoli gli archivi privati cui sia riconosciuto particolare interesse storico con provvedimento emesso dalla Soprintendenza Archivistica e Bibliografica competente per regione ai sensi degli artt. 10 e 13 del D. Lgs. 42/2004 "Codice dei Beni Culturali".

I soggetti preposti alla tutela degli archivi statali sono gli Archivi di Stato di riferimento in cui opera l'organo periferico cui l'archivio appartiene. I soggetti preposti alla tutela degli archivi degli enti pubblici e degli archivi privati dichiarati sono le Soprintendenze archivistiche e bibliografiche (lo stesso regime, infatti, è stato esteso dal 2014 alle biblioteche, la cui tutela è passata dalle Regioni allo Stato nell'agosto di quell'anno). La valorizzazione è materia regionale con coordinamento del Ministero per i Beni e le Attività Culturali e per il Turismo.

La casistica di archivi nei quali è possibile rintracciare fonti orali è molto variegata: fonti del genere, come si evince da tutti i censimenti condotti, possono essere in archivi pubblici (università, centri di ricerca), archivi privati di associazioni o istituti di ricerca (primi fra tutti gli istituti storici della resistenza, ma anche associazioni legate a movimenti di revival folkloristico), archivi di persona (ricercatori afferenti, tra le altre, proprio alle discipline già richiamate che possono essere affiliati o meno a istituzioni universitarie ma la cui produzione documentaria resta comunque indipendente).

In questa sede voglio soffermarmi, vista la naturale destinazione di questo intervento, sui primi, e in particolare sugli archivi delle università, anche qui per ricordare quali siano le previsioni di legge e quali i limiti di tali previsioni nella buona conservazione della documentazione legata alle fonti orali (limiti, probabilmente, dovuti anche al ritardo cui si è accennato sopra dell'archivistica e delle istituzioni archivistiche nel riconoscimento di tali documenti come documentazione archivistica, utile e affidabile nel tramandare memoria e fare ricerca storica)[22].

Gli archivi universitari sono soggetti all'attività di vigilanza della Soprintendenza. Le università si dotano di massimari di scarto. In base a tali massimari, nessuna università italiana è tenuta per legge a conservare in archivi dedicati l'operato dei gruppi di ricerca: si conservano solo i risultati e l'informazione amministrativa circa l'esistenza di un dato progetto. In nessun caso sembra vi sia interesse nel formare e tenere archivi della ricerca: si formano siti dedicati, ma non si mantiene traccia dei quadri epistemologici e della produzione documentaria 'endoprocedimentale' di

[22] Oltre a Portelli (2014), si veda anche Cartosio (2013), incentrato sulla difficoltà di riconoscimento delle fonti orali tra le fonti storiche da parte della stessa tradizione storiografica.

questi progetti che potrebbero essere invece particolarmente utili e interessanti dal punto di vista culturale. I massimari di scarto attuali prevedono il mantenimento dei soli elaborati finali quali tesi di laurea e dissertazioni dottorali.

È questo, quindi, il motivo per il quale anche nuclei archivistici formatisi nel corso di attività istituzionali restano, poi, nei cassetti o comunque nella disponibilità (e nella responsabilità storica) di quanti a questa attività si sono dedicati.

Si prenda, per esempio, il caso della ricerca che ha portato alla pubblicazione dell'Atlante Lessicale Toscano, attualmente disponibile in CD-ROM e in rete su ALT-web (cfr. http://serverdbt.ilc.cnr.it/altweb/). Dell'impresa, che già nel 1983 portò alla pubblicazione dell'Atlante, non è stato conservato 'l'archivio' (cfr. Calamai, Biliotti, 2012), vale a dire la documentazione prodotta nell'ambito di quella specifica ricerca dall'Università di Firenze, che di quella ricerca fu finanziatrice e promotrice attraverso la cattedra di dialettologia di Gabriella Giacomelli. Della documentazione prodotta resta quindi quanto oggi è stato conservato ad iniziativa di altri soggetti, diversi dal 'produttore': il fondo Giacomelli, che si trova presso l'Accademia della Crusca e che raccoglie tesi di laurea e relazioni di seminario, e la documentazione sonora digitalizzata nell'ambito del progetto Gra.fo, che contiene solo il 36% delle campagne di inchiesta, sia perché molte registrazioni sono andate disperse sia perché i raccoglitori non erano tenuti a registrare o a conferire i nastri se acquistati autonomamente, ma soltanto a fornire la trascrizione fonetica delle risposte rese dagli informatori.

Della ricerca non resta, quindi, la documentazione eventualmente stilata a corredo delle interviste, né quella afferente alla scelta di non mantenere come riferimento la registrazione ma la trascrizione. Diversamente, se l'Università di Firenze avesse dovuto (o avuto l'opportunità di) conservare e trattare archivisticamente quanto prodotto dal gruppo della Giacomelli o dalla cattedra della Giacomelli nell'ambito della ricerca dedicata all'Atlante Lessicale Toscano, si potrebbe avere oggi un accesso 'contestualizzato' rispetto al farsi della ricerca stessa con il mantenimento della documentazione 'endoprocedimentale', comprese le, eventuali, registrazioni o la documentazione che ne motiva storicamente l'assenza.

Nella direzione fin qui adombrata, di trattamento archivistico della documentazione legata alle fonti orali, vuole andare la progettazione di 'Archivio Vi.Vo', archivio delle fonti orali della Toscana, in cui i documenti di testo si integrino a quelli orali e cinematici nell'intero contesto documentario prodotto dai soggetti e dai sistemi di ricerca che li hanno prodotti.

Tale progettazione è nata dalla collaborazione tra Soprintendenza Archivistica della Toscana, Università di Siena, il nodo italiano CLARIN dell'Istituto di Linguistica Computazionale di Pisa, il Centro Ricerca Documentazione Didattica del Casentino. La fase pilota del progetto sarà finanziata da Regione Toscana nell'ambito dell'accordo di valorizzazione per il triennio 2019-2021 tra la Soprintendenza e la Regione stessa e avrà come primo obiettivo il trattamento archivistico delle fonti orali raccolte e prodotte dalla *folk singer* e ricercatrice toscana Caterina Bueno.

La creazione di *repositories* standardizzati e di sistemi informativi di interesse interdisciplinare si pone non solo come una necessità per la valorizzazione e la conservazione del patrimonio cosiddetto immateriale, ma soprattutto come una sfida metodologicamente significativa nel quadro delle discipline umanistiche accomunate dall'uso delle fonti orali.

Bibliografia

Andreini, A., Clemente, P. (Eds.) (2007). *I custodi delle voci. Archivi orali in Toscana: primo censimento.* Firenze: Centro Stampa Regione Toscana.

Bachiddu, E. (2012). Fonti orali. Approcci e dialoghi tra antropologia e storia orale. Introduzione. In *Lares,* 78, 1-2, 5-20.

Barrera, G., Martini, A. & Mulé, A. (Eds.) (1993). Fonti orali. Censimento degli istituti di conservazione. In *Quaderni della Rassegna degli Archivi di Stato*, *71*. Roma: Pubblicazioni degli Archivi di Stato.

Bergamin, G., Lucarelli, A. (2013). The Nuovo Soggettario as a service for the linked data world. In *JLIS*, 4, 1, 213-226.

Bernardi, B., Poni, C. & Triulzi, A. (1978). (Eds.) *Fonti orali: antropologia e storia.* Milano: Franco Angeli.

Bertinetto, P.M., Calamai, S. (2016). L'unità documentale negli archivi sonori. In Marazzini, C., Maconi, L. (Eds.), *Atti del convegno Piazza delle Lingue 2014: Italiano elettronico. Vocabolari, corpora, archivi testuali e sonori.* Firenze, 6-8 novembre 2014, 95-112.

Cartosio, B. (2013). Storia orale e storia. Manoscritto, Società di Mutuo Soccorso Ernesto de Martino. https://smsdemartino.files.wordpress.com/2013/11/storia-orale-e-storia.pdf/ Accessed 11/09/2019.

Carucci, P. (1993). Prefazione. In Barrera, G., Martini, A. & Mulé, A. (Eds.), *Fonti orali. Censimento degli istituti di conservazione*. *Quaderni della Rassegna degli Archivi di Stato, 71*. Roma: Pubblicazioni degli Archivi di Stato, 13-23.

Carucci, P., Contini, G. (Eds.) (1988). Le fonti orali. In *Rassegna degli archivi di stato XLVIII, 1-2.* Roma: Ministero per i beni culturali e ambientali, ufficio centrale per i beni archivistici, divisione studi e pubblicazioni.

Cencetti, G. (1939). Il fondamento teorico della disciplina archivistica. In *Archivi*, 2, 6, 7-13.

Cirese, A.M. (1996). Lo studio delle tradizioni popolari. In Brioschi, F., Di Girolamo, G. (Eds.), *Manuale di letteratura italiana, vol. IV*. Torino: Bollati e Boringhieri, 931-941.

Clemente, P. (1986). Voci su banda magnetica: problemi dell'analisi e della conservazione dei documenti orali. Note italiane. In *Gli archivi per la storia contemporanea. Organizzazione e fruizione. Atti del seminario di studi,* Mondovì, 23-29 febbraio 1984, 185-191.

Contini, G. (2006). Le fonti orali e audiovisive. In Pavone, C. (Ed.), *Storia d'Italia nel secolo ventesimo: strumenti e fonti, vol. III: le fonti documentarie*. Roma: Ministero per i beni e le attività culturali. Dipartimento per i beni archivistici e librari. Direzione Generale per gli Archivi, 795-820.

Duranti, L. (2015). Il documento archivistico. In Giuva, L., Guercio, M. (Eds.), *Archivistica. Teorie, metodi e pratiche.* Roma: Carocci, 19-34.

Fischetti, A. (2008). *Creazione e gestione delle fonti orali.* In Bermani, C., De Palma, A. (Eds.), *Fonti orali. Istruzioni per l'uso.* Venezia: Società di Mutuo Soccorso Ernesto De Martino. http://aisoitalia.org/creazione-e-gestione-della-fonte-orale/Accessed 11/09/2019.

IFLA Study Group on the Functional Requirements for Bibliographic Records (2009). *Functional requirements for bibliographic records: final report*. https://www.ifla.org/files/assets/cataloguing//frbr_2008.pdf/ Accessed 24.09.2018.

International Council of Archives. *Multilingual Archival Terminology* http://www.ciscra.org/mat/mat/term/1418/1801/Accessed 26.04.2019.

International Council on Archives. Experts group on Archival Description (2016). *Record in Contexts. A conceptual model for Archival Description. Consultation Draft v.01*. https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf/Accessed 24.09.2018.

Leydi, R., Telmon, T. (1999). Documenti sonori e ragioni della ricerca. In *Archivi sonori: atti dei seminari di Vercelli (22 gennaio 1993), Bologna (22-23 settembre 1994), Milano (7 marzo 1995).* Roma: Ministero per i beni e le attività culturali, 17-34.

Magnani, F. (2018). Le documentazioni audiovisive nelle schede di catalogo per i beni culturali demoetnoantropologici. In Tucci, R. (Ed.), *Le voci, le opere e le cose. La catalogazione dei beni culturali demoetnoantropologici*. Roma: Ministero dei beni e delle attività culturali e del turismo, Istituto centrale per il catalogo e la documentazione, 262-274.

Matranga, V. (2002). Come si fa un'indagine dialettale sul campo. In Cortelazzo, M., Marcato, C. & De Blasi, N. (Eds.), *I Dialetti Italiani,* 64-80. Torino: UTET.

Matter, M. (1995). *Oral History Cataloging Manual*. Chicago: Society of American Archivists.

Mc Luhan, M. (1967). *Gli strumenti del comunicare.* Milano: Il Saggiatore.

Mulè, A. (2003). Le fonti orali in archivio. Un approccio archivistico alle fonti orali. In *Archivi per la storia*, 16, 1, 111-125.

Ong, W.J. (1986). *Oralità e scrittura. Le tecnologie della parola*. Bologna: Il Mulino.

Portelli, A. (1999). Sulla diversità della storia orale. In Bermani, C. (Ed.), *Introduzione alla storia orale. Storia, conservazione delle fonti e problemi di metodo*. Volume 1. Roma: Odradek, 149-166.

Portelli, A. (2014). Sulla diversità della storia orale. Manoscritto, Aiso. http://aisoitalia.org/wp-content/uploads/2014/04/Alessandro-Portelli-Storia-orale-un-lavoro-di-relazione.pdf/Accessed 11.09.2019.

René-Bazin, P. (1988). La création et la collecte des nouvelles archives. In *Rassegna degli archivi di Stato*, 48, 14-50.

Acknowledgments

Thanks to Silvia Calamai, who gave me the chance to reconnect my multiple lives in one and to Giovanni Contini, who introduced me to the wonderful world of Tuscan Voices and Music.

DOI: 10.17469/O2106AISV000007

DANIELA-CARMEN STOICA, ANYLA SARAÇI

Oral Histories and the Dialectal Research in Albania

The present article discusses the use of oral histories as a key methodological tool in linguistic and dialectal research in Albania, and as a crucial part of historical documentation. The article draws mainly on the empirical experience of both authors, who have performed research mainly in the Korça area, an area of complex linguistic contacts (Albanian-Slavic Macedonian-Greek). The two researchers recorded a number of oral histories in two distinct speeches; one belonging to the Tosk dialect of the Albanian language and the other belonging to the Aromanian dialect of the Romanian language. These two speeches are in danger of becoming extinct for two main reasons: the first being due to the linguistic homogenization of the Albanian dialects into the standard variant (Shkurtaj, 1999), the other being due to the cultural and linguistic assimilation of the Aromanians living in Albania (Capidan, 1943).

Key words: oral history, Tosk dialect, Devoll speech, Aromanian dialect, Farsherot speech.

1. *Introduction*

The present article aims to discuss the advantages and limitations of oral histories (in the form of both video recordings and interviews), as sociolinguists' main tools for linguistic and dialectal research, as well as for historians to document various historical moments in Albania. The oral histories analyzed in the article were collected between 2007 and 2010 (for the Devoll speech) and during the period 2010-2015 (for the Farsherot speech[1]) mainly in the Korça area, situated in South-Eastern Albania, an area of complex linguistic contacts (Albanian-Slavic, Macedonian-Greek), where the Tosk dialect of the Albanian language and its various speeches also come in contact with the Aromanian dialect, particularly with its Farsherot speech.

The geographical position of the Devoll area to the South of the Shkumbin River (which divides Albania into the two main dialectal areas; that of the Geg dialect to the North and that of the Tosk dialect to the South) includes the Devoll speech into the Tosk dialect area. This speech is spoken in the villages located on the two sides of the Devoll River, including those villages found where the river springs up on the mountain and down, up to the village of Cangonj, where it becomes a natural boundary, separating the Devoll area from the Korça area. On the one hand, this speech presents common features with the Geg dialect, particularly

[1] In this study the term "speech" has been used observing the corresponding terms in Albanian and Romanian dialectology: *e folme* (Albanian) i.e. *e folmja e Devollit-* the Devoll speech and *grai* (Romanian) i.e. *graiul fărșerot-* the Farsherot speech.

with some of its speeches, but it also presents special features which include it in the Eastern group of speeches belonging to the Northern Tosk sub-dialect.

Within the Northern Tosk sub-dialect there are two main speech groups: the Western and the Eastern group. Due to the geographical position of Devoll, the Devoll speech falls under the Eastern group. The territories where the Eastern speeches are spoken used to belong to the various administrative units, separated from each other. Thus, historically, the Western speeches have been spoken in the territories belonging to the Vlore and Berat Sanjaks[2], whereas the Eastern speeches have been spoken in the Oher and Selasfor Sanjaks. Before the Ottoman conquest, these two groups were divided again on a territorial basis, since Vlore and Berat areas were under the domination of the Muzakas[3] (Hopf, 1873). The Devoll speech is very interesting from a dialectological point of view, not only because of its geographical position, but also due to the specific historical and social circumstances under which it developed, turning it into a speech with distinct features, as compared to other Tosk speeches and standard Albanian. As a borderline South-eastern speech spoken by an Albanian community living next to a Greek speaking community, this speech preserves old features of the Albanian language, especially at the phonetic level.

Two ethnographic works have dealt with the Devoll speech at all linguistic levels, being accompanied by dialectal texts and a glossary of regional words (Xhaçka, 1956 and Gjinari, 1960). In addition, there are also two works on several customs and rites of the Devoll area written by Xhaçka (1959a; 1959b), as well as a volume of folklore from Upper Devoll collected by Zdruli (1987).

The Aromanians of Albania belong to the Farsherot group (Capidan, 1931; Nevaci, 2011), speaking a dialect of the Romanian language, known as the Aromanian dialect, which, together with the Meglenoromanian and Istroromanian, form the historical dialects of the Romanian language to the South of Danube.

The Aromanian dialect is widely spread in Albania, Greece, Bulgaria, the Republic of North Macedonia and Romania, as well as other European countries. The dialect has also spread across Australia and the United States due to the immigration of the Aromanian community over the last few decades.

Based on the linguistic differences of their speeches, Aromanians have been divided into several groups or branches: the Pindens (most of them compactly concentrated in Greece, in the Pindus Mountains), Farsherots (most of whom live in Albania, having their origins in Frashër, Albania), the Gramostens (having their origins from Gramoste) and the Grabovens (from Grabova, Llënga and Niçea in Albania, near Pogradec, about 30 km from the city of Korça).

Most of the Farsherots live in Albania, though we may come across Farsherot communities today in the Republic of North Macedonia and Greece. This is due to the extraordinary mobility of the Farsherots and the nature of their nomadic lifestyle, historically. From all the Aromanians, the Farsherots are the only ones who keep their ethnic name without its prosthesis *a-*, but with an initial apical trill, *rrămăn*,

[2] Former administrative unit during the Ottoman Empire.
[3] Albanian aristocratic family during the Middle Ages.

similar to the sound reproduced by the Albanian *rr* grapheme (multi-vibration apical trill). The other populations call them *Vlachs*, a term by which the Romanized populations living to the north and south of the Danube have generally been designated. Albanians call them *çobeni* ("shepherds", so by their traditional profession) or *rëmeri* (probably from Romanus), while the Greeks call them αρβανιτόβλαχοι (i.e. "Vlachs from Albania") or καραγούνοι ("bearers of black furs"), this latter name pointing to the traditional black sheep or goat fur these shepherds used to wear in the past. Being a population of shepherds, the Farsherots have known a wide spread in Greece too, where many of them finally settled. The case of the Aromanians is a particular one because, even though they are among the oldest ethnic groups in the Balkans, speaking a Romance idiom, they have never managed to form a state. Most of the Farsherots are settled in Albania, and live in communities spread like small islands on the map of Albania, in the southern part of the country (the area of contact with the Tosk dialect of the Albanian language), both in the mountainous or higher parts of the southeast, as well as in the western plain, Muzachia Plain (Alb. Myzeqe). Therefore, the Aromanians living in Muzachia Plain, in the districts of Vlorë, Fier, Lushnja, some of the Berat district call themselves in arom. *mizuk'ări*. Depending on their places of origin, there are other names assigned to the different groups, such as *culuńáț* (from Kolonja), *jărcăńóț* (from Jarcani), *uiańíț* (from Ujanik), *căstrăńóț* (from Kustrec), *plisóț* (from Plasa).

At present there are no exact statistics on the number of Aromanians in Albania or in the Korça area. According to the information obtained from the Aromanian associations in Albania, their number would be roughly 200.000, with between 4.000 and 5.000 being in the Korça area. Our investigations show that a large number of Aromanians living in the city of Korça today have their origins in the old Aromanian centers Moscopole (Alb. Voskopojë) and Șipsca (Alb. Shipskë). They are considered to be old town dwellers, and their homes are generally located in districts 1 and 2 at the very heart of the city, the historical area of Korça. In the last 20 years, a new district appeared, inhabited mostly by Aromanians, whose places of origin are Plasa and Mborie (the latter being located just a few kilometers from Korça city and practically joining it). In addition, our interviews show that some of the neighboring villages (Mborja, Drenova, Boboshtica, Bellovoda, and Voskopojë) were populated by Aromanians after 1945, when they were forced to settle down under the Communist regime.

During the Communist regime and its subsequent fall in 1991, the Aromanians of Albania did not enjoy the status of "national minority", because, according to the Albanian state, they lacked "a motherland", or, to use another term, "a kin-state". Officially, the Aromanians and Romas had the status of "ethnic-linguistic minorities", which neither de facto, nor de iure granted any language rights to these communities. Officially, this situation has come to an end with the entering into force of Law 96/2017, where 9 national minorities are recognized today in Albania: Greek, Macedonian, Aromanian, Roma, Egyptian, Montenegrin, Bosnian, Serb and Bulgarian.

1.1 The Use of Movies for Dialectal Studies

Albanian art films are a very good source for dialectal studies too. The film history in Albania begins relatively late, but their production from 1953, with the Soviet-Albanian joint productions, until the 1990s had been abundant. It is worth mentioning that the Albanian films of the period 1958-1980 focused largely on topics of war, historical themes, topics dealing with rebuilding the country and changing the mentality under the new Communist ideology. As a result, films were made with a large cast of actors where the community was on focus, not the individual, the action, and not the psychological analysis. In the films of this period, a large group of amateur actors were involved, actors who maintained the essential characteristics of their dialect, whereas the professional actors adjusted their speech to the dialect, depending on the film theme.

The Korça region played an important role in both broadcasting and providing actors. Many of the films of this period were made in Korça and the surrounding areas because they presented a distinct mentality from the Albanian mentality as a whole. Moreover, Korça has had a rich theater tradition, enabling more qualified actors and directors.

Devoll's speech is decisive in the characters of some of the films, such as "The First Years" (1965), "The Eighth in Bronze" (1970), "The Stars of the Long Nights" (1972), "The Confrontation" (1976). Meanwhile, Korça and the surrounding areas are represented in a large number of films (apart from those we mentioned), such as "The Emperor's Bed" (1973), "Beni Walks by Himself" (1975), "The Narrow Streets Looking for the Sun" (1976), "The Last Winter" (1976), "The Lady of the City" (1976), "The Man with a Canon" (1977), "A Friend from the Village" (1980) (Qezari et al.: 2004)[4].

2. *Methodology*

2.1 Introduction

During their field research, often accompanied by their students, the two experienced researchers managed to record an important number of oral histories (approximately 5 hours each) in various speeches of the Tosk and Aromanian dialects, speeches which are in danger of becoming extinct because of the latest linguistic homogenization of the Albanian dialects into the standard variant (Shkurtaj, 1999), on the one hand and due to the cultural and linguistic assimilation of the Aromanians living in Albania, on the other hand (Capidan, 1943). More than that, the oral histories collected represent two important databases not only for linguistic and dialectal research in this particular geographic area of Albania, but they may become valuable data for historians and social scientists since they focus on particu-

[4] "Vitet e para" (1965), "I teti në bronz" (1970), "Yjet e netëve të gjata" (1972), "Përballimi" (1976). "Krevati i perandorit" (1973), "Beni ecën vetë" (1975), "Rrugicat që kërkonin diell" (1976), "Dimri i fundit" (1976), "Zonja nga qyteti" (1976), "Njeriu me top" (1977), "Një shoqe nga fshati" (1980).

lar events from the history of Albania, such as the creation of the Albanian modern state, the Second World War and the liberation from the Germans, the instauration of the Communist regime in Albania and its consequences, the transition from the traditional nomadic lifestyle to the sedentary one within the borders of the Albanian state of the Aromanians during the Communist regime etc. By focusing on the speakers' unique life experiences, these oral histories become an opportunity for all these recorded people who have been "hidden from history" to have their voices heard.

Finally, oral histories are looked upon as narratives, which according to many authors (Schiffrin, 1996; De Fina, 2003, 2006; Georgakopoulou, 2007) represent an important locus for the articulation of identities because they give the tellers the opportunity to present themselves as actors in social worlds, while at the same time negotiating their present self with other interactants.

The dialectal texts for the Devoll speech have been collected in 22 dialectal points. Some texts were recorded by means of a dictaphone, others by video camera. The questionnaire used for the interviews was the one compiled by S. Floqi in 1971. This questionnaire continued and extended the one compiled by Çabej and Domi in 1953 (containing 63 questions in all, 20 about phonetics, 30 about morphology and syntax and 13 general ones about lexicology) since it contained 304 questions in all, 74 about phonetics, 54 about morphology, 20 about syntax and 60 about lexicology. Relying on the guidance offered by the cited questionnaire, the informants were preselected according to their ability to speak clearly and to tell stories at a good capacity (Floqi, 1971: 377) In addition, extra parameters taken into account were whether the informant had his/her origins in the village where the interview was conducted; if the informant had different origins, it was important to know how long he/she had been living there. The transcription of the texts was done by means of the IPA system using diacritical signs for various characteristics of the sounds.

The texts used for the Aromanian dialect have been collected through interviews conducted in Korça area, which includes Korça city and several neighboring villages, such as Voskopojë, Shipskë, Mborje, Drenova, Boboshtica, Bellovoda, Plasa. The number of the informants was 60, males and females, in a balanced number, of various ages and having various educational levels. The informants belonging to the age group 60 to 80 were retired; most of them used to work on farms especially during the Communist regime, playing various professional roles within the agricultural cooperatives of production. Before the instauration of this regime in Albania, the Aromanian informants belonging to this age group used to have a largely nomadic lifestyle, characterized by transhumance (i.e. in summertime they used to live high up in the mountains, where they were leading their flocks of sheep, and in winter they used to lead them near the sea, where they built shelters for them), many of them having no formal education or only the basic level (the primary school). The other groups identified were adults aged 30 to 50, professionally active, having graduated high school or university, working in the public or private sector, and children aged 7 to 15, all attending public schools where standard Albanian, quite similar to the Tosk dialect, is the main language of instruction.

The questionnaire contained several open questions, whose topics depended on the informant's age, asking him/her to describe old professions, traditions and customs (e.g. how they used to organize baptisms, engagements, weddings, funerals or how they used to make bread, cheese, wool etc.) and several open questions requiring them to describe more contemporary professions and preoccupations (e.g. a visit to the doctor/dentist, at the supermarket, travelling abroad, a day at school/university etc.).

2.2 Goals

The recordings for the Devoll speech contain speech samples from informants of various ages too. Due to the special features of the speech, most of the informants belong to the ages of 60 to 80, as they have generally preserved the old features of the dialect, particularly the informants living in the villages from the Upper Devoll area. These villages are mainly inhabited by Christian orthodox communities, whose members have massively immigrated to the USA. The other informants, belonging to the age groups 40 to 60 or 5 to 15 and 16 to 30 were recorded in order to measure the influence of the following sociolinguistic factors over the dialect and consequently over the speech:

1. the massive influence of the standard Albanian due to the obligatory education in standard Albanian (especially from 1971 till 1990);
2. the Code-switching phenomena due to the immigration before 1945, mainly in the USA, and after the 1990s in Greece and the USA.

Our method of research combines the sociolinguistic interview with the ethnographic qualitative research and discourse analysis, focusing mainly on *indexicality* (especially linguistic – e.g. sounds, accents, words, expressions that may be associated with aspects of the larger context when they evoke specific traits, ideas, activities and properties that may be seen as typical of certain social identities), *local occasioning* (i.e. identity seen as a communicative process taking place within concrete social contexts and practices, paying attention to details of local talk in order to understand how identities are brought about and negotiated), *positioning* as relational processes (i.e. identity seen as a socially grounded process that involves a continuous confrontation of the self with others; cfr. Wilson, 1990; De Fina, 1995 and Van Dijk, 2010).

2.3 Phonetic Transcription

An important part of each research has been the transcription of the interviews, partly recorded by video camera and partly by dictaphones. For the Tosk dialect, the system of transcription used was IPA, following the tradition of the Albanian dialectological school which started in the 1950's, also present in *The Dialectological Atlas of the Albanian Language* (Volume I, II, 2007, 2008). For the Aromanian dialect, the system was the one used by the Romanian school of linguistics and dialectology, particularly by the *Romanian Linguistic Atlas* taken over, with a few additions, from the collaborators to the *New Romanian Linguistic Atlas by Regions* and to the Romanian Phonogrammic Archive. In the latter case, many symbols are accompanied by diacritical signs in order to reflect specific characteristics of certain vowels (degree of openness, centrality, intensity, length etc.) and consonants (palatalization, nasality etc.), which unfortunately could not be transferred into ELAN, the linguistic annotator, which the researcher attempted to use in order to further analyze the linguistic corpora collected, especially quantitatively. The transcription of diphthongs, whose presence was massive and specific for both speeches, remained a challenge in both cases.

2.3.1 Phonetic Transcription of Devoll Speech

The oral histories in the Devoll speech were mainly recorded by dictaphone using the questionnaire compiled by Floqi (1971), whereas the others were video recorded. The transcription of the oral histories were based on the IPA system (See *Appendix 1*), but the researcher experienced difficulties transcribing them since this was done by ear and not using a special computer program. Nevertheless, the phonetic transcription system used for the consonants has been adapted to the graphic system of Albanian, similarly to the one used by S. Floqi: e.g (<sh> for [ʃ], <ll> for [ɫ] (Floqi, 1971: 381-382), and not according to the IPA system).

One of the challenges was the transcription in the case of the affrication of the post dental [t] after the post-alveolar [ʃ]: in IPA system there is no special sign for the transcription of the affricated post-dental [t]. The solution found at that moment was the use of the symbol for the affricate <ç>, regardless of the degree of the affrication (**shç**atë "seven", **shç**izë "spear", **shç**ytkë "diaper", **shç**jere "pour", ja**shç**ë "out").

Regarding the vowels' system of the Devoll speech, it does not contain diacritical signs because the vowels do not have the characteristics of nasality or length.

The most problematic transcription was that of the vowel [ə] which is more closed towards [e] (so [ɘ]). The corresponding vowel in the other speeches of the Tosk dialect is <â> placed before one of the nasal consonants [m], [n], [nj]: E.g. *shkëmp* "rock", *dhëmp* "tooth", *dhëntret* "groom", *i mëngjër* "left".

Xhaçka emphasizes that: "... here too, like all over the Tosk dialect, we have the stressed [ə] in those positions where in the Geg dialect there is [â]" (Xhaçka, 1956: 108). This [â] before nasal consonants has been replaced by [ə] only in some Turkish words such as: *musëndër* "a kind of wardrobe", *nallëne* "sabot", but in other Turkish words it has remained unchanged: e.g. *ashllama* (a kind of food), *Osman, Hasan* (proper names), *aksham* "evening", *amamçe* "Turkish bath", *ajvan* "donkey" etc. According to Gjinari, in the speech of the old Devoll inhabitants, this [ə] is pronounced more backwards and more rounded than in the speeches of Berat, Skrapar, Myzeqe and Mallakastra (Gjinari, 1960: 105). However, we may add that today, even though it is pronounced as a mid vowel in words such as *këngë* "song", *këmbë* "leg", *dhëmb* "tooth" etc., this vowel is pronounced slightly more frontal that in other speeches, which makes it closer to [e]. In the speeches of Labëria and Çamëria (two other sub-dialects of Tosk) vowel [ə] is clear in every position.

Another challenge was the transcription of the diphthongs in various positions in the words:

– diphthong ***u̯a*** – its transcription by means of the diacritical sign of the non-syllabic u̯; ***Jorgu̯a, u martu̯a*** "was married", ***u çpu̯a*** "was pierced", ***kurisku̯a*** "humpback", ***u ndru̯a*** "got dressed", ***u lëçu̯a*** "rushed on".
– diphthong ***ɥe-*** may be found in any position in a word, ***lërɥer*** "got dirty", ***(me) të pɥeturë*** "asked", ***rëfɥerë*** "told", ***rëmbɥerë*** "grabbed". It is worth emphasizing that in this diphthong, in certain positions, the closed vowel [y], loses its initial element and reduces to [e]: ***janë ktherë-janë kthyer*** "have returned", ***ndë kre-në krye*** "in front", ***më ka rëmberë-më ka rrëmbyer*** "has grabbed me". In some cases, this diphthong appears as a result of a more frontal labialization of the vowel [u], which changes it into[y] because of the presence of the alveolar-palatal consonants in front of it: *mbles>lues>lyes* (in standard Albanian *mbles* "matchmaker") (Gjinari, 1960: 105), *Shuec>Shyec* (Shuec- official name of a Devoll village).

Furthermore, the diphthong may decompose, not because the constituent vowels may be divided into different syllables, but because in their middle the consonant [v] may insert itself, a phonetic phenomenon which is against the hiatus, like in the case of the toponym *Shyec-Shyvec*. Researching the dialectal material collected in the villages Vishocicë, Hoçisht, Ziçisht, Vërlen, we have come across the form *byéj* (the plural form of the noun *buall* "buffalo"), pronounced with a diphthong. The change of the diphthong [u̯a] into [yé] has taken a grammatical value by means of the metaphony and this is an old plural form (Çabej, 1958).

2.3.2 Phonetic Transcription of Aromanian Speech

For the Farsherot speech of the Aromanian dialect, the system was the one used by the Romanian school of linguistics and dialectology, particularly by the *Romanian Linguistic Atlas* taken over, with a few additions, from the collaborators to the *New Romanian Linguistic Atlas by Regions* and to the Romanian Phonogrammic Archive.

Many symbols are accompanied by diacritical signs in order to reflect specific characteristics of certain vowels (degree of openness, centrality, intensity, length etc.) and consonants (palatalization, nasality etc.), which unfortunately could not be transferred into ELAN, the linguistic annotator designed by the Max Planck Institute for linguistic research, which the researcher attempted to use in order to further analyze the linguistic corpora collected, especially quantitatively (See *Appendix 3*).

3. *Indexicality*

The following sub-chapter will demonstrate that certain phonetic traits and styles of speaking in the case of the Albanians living in Devoll area and Aromanians from Korça become symbolically associated with complex systems of meaning such as ideologies, social representations about group membership, social roles and attributes, presuppositions about aspects of social reality, individual and collective stances, practices and organization structures.

From the collected material, it appears that in the case of the Aromanians living in the Korça area, the frequency of pronouncing the apical *r* with several vibrations is more frequent with men, who, according to Capidan (Capidan, 1943: 78-79), have lost their uvular *r*, due to their social mobility, to their bilingualism or even tri-lingualism. Uvular *r* appears only in the speech of the Aromanian women, especially from rural areas, a sound which changes the usual sound of the nearby vowels. Uvular *r* is considered to have been uttered by all speakers in the past (Capidan, 1931: 124), regardless of their gender, but it has been preserved only by women because they were not socially mobile like men, usually being confined to their households and villages where they were living and raising their children or grandchildren to whom they were teaching the "authentic" mother tongue. This was also confirmed by our interviews, since all the ten women, with ages between 60 and 80, interviewed in the villages of Voskopojë, Shipskë, Plasa, Drenova, Mborja, as well as in Korça city (who had lived most of their lives in villages and later had moved to the city) presented the uvular *r*. In the case of women informants whose ages were between 30 and 50, only the ones living in villages did present this phonetic feature, whereas the others, living in Korça city, had lost it.

In the case of the old Albanian informants of Devoll (60-80 years old), an important phonetic characteristic is the preservation of the consonant groups [kl], [gl], as in *klishë* (>kishë "church"), *glu* (>gju "knee"), present in the old Albanian and today in the Arbëresh speeches of Greece and Italy (Çabej, 1958; Çabej, 1969; Topalli, 2007). These groups have undergone palatalization, evolving today to the sounds represented by the graphemes <q> and <gj> in standard Albanian. In connection to these consonant groups, J. Gjinari (1960), based on the dialectal material collected by V. Xhaçka (1959), states that they appear only in the speech of the old speakers (60-80 years old) of Christian religion from the villages of Upper Devoll. This conclusion is also endorsed by the old toponimy of the area, existing even in Devoll villages inhabited by Muslim population today, such as *Varet e Klishës* ("The Graves of the Church") in Bitinckë, *Klishka* ("The Church", a place where there used to be a church) in Kurilë, Eçmenik, *Lisi i madh de Klisha* ("The Big Oak Tree of the Church", a place where there used to be an oak tree, 150 years old) in Stropan. Thus, these speakers, of Christian religion and living mainly in Upper Devoll area, are considered to be the only ones left still preserving this feature of the Devoll speech, while the oral histories of the other age groups present more phonetic characteristics of the standard variant.

The transformation of the consonant groups [kl] and [gl], into <q> and <gj>, respectively, has already come to an end. What is interesting is that this evolution has not taken place step by step like in the case of some Geg speeches: e.g. *kl>kj, gl>gi* (*kian* "cry", *giuri* "knee") (Çabej, 1977: 107). If in three Northern sub-dialects these consonant groups have evolved, this has not happened in the speeches of Çamëri, of several villages in Korce (Dardha, Sinica, Qyteza etc.), as well as in the speeches of the old Arberesh communities settled in Greece, Ukraine and partly in the speeches of the Arberesh of Italy (Topalli, 2007: 267). Gjinari highlights the fact that the old people from Sinica (a village in Upper Devoll, inhabited mainly by Christians) have changed the pronunciation of these consonant groups on purpose so that they may not be discriminated from the territorial community whose members they were part of (Gjinari, 1960: 110). However, since these consonant groups have appeared till recently on the linguistic map of this area, this is a clear proof of linguistic conservation. Furthermore, the preservation of [kl] group was present in the words where it finally changed into the velar [k] and not into the palatal [c]: e.g. *klisha>kisha* "church"; *klyçi>kyçi* "wrist"; *klinda>kinda* "crinkle". More than that, the groups [kl] and [gl] have been preserved in the late borrowings, which entered the dialect and the speech after the transformation of these consonant groups, like in the Latin borrowing *glasë* "bird droppings" (< Lat. *galliatia), or Slavic borrowing *kleçkë* "splinter, piece of wood" (< Sl. *kléčka), or in several toponyms like Kokogllavë (the name of a mountain). This phenomenon has been observed in 7 informants from Upper Devoll villages, over 70 years old. Nowadays it is almost impossible to find informants who use these consonant groups.

4. *Local occasioning*

At the lexical level, our analysis has dealt mainly with the borrowings and code-switching phenomena, using several linguistic criteria (i.e. frequency, adaptation and attestation in dictionaries), so that we may conclude on the degree of influence of the Tosk dialect over the Aromanian one, on the one hand, and of the standard Albanian on the Devoll speech, on the other hand. In addition, we have taken into account the gender, age, education level of the Albanian and Aromanian informants in order to see whether there are variations in this respect too. In the case of Aromanian, the inserted language is Albanian (i.e. the Tosk dialect spoken in South-Eastern Albania) and the analysis is limited to the dialect spoken by the Aromanians from the Korça area, who live together with the Albanian majority. They are predominantly bilingual, using Aromanian mostly at home, in the family environment and within the community, and Albanian, both in the community and for official communication. Younger generations tend to abandon their mother tongue and use Albanian in verbal interactions both within the family and the Aromanian community, motivated by the strong influence of the Albanian language on the Aromanian children and youths during the schooling process.

In the case of the informants aged 60-80, we have analyzed, at lexical and syntactical levels, the borrowings and code-switching phenomena in their communication, which highlighted their dynamic and fluid identity under specific political, social and economic circumstances in the history of Albania. If before Communism, most of them were *păstori* "shepherds" (<Lat. *pastor, -is*) or *çobenj* (<Turkish *çoban*) leading their sheep to better pastures in the mountains, during the Communist regime most of them gave up on their traditional profession and started working on farms, playing various professional roles within the agricultural cooperatives of production, and enjoying the status of pensioners later on in their lives: *magazinér*

"warehouseman" (<Alb. *magaziner*), ***mecaníc*** "mechanic"(<Alb. *mekanik*), ***krúe-veterinér*** "chief vet"(<Alb. *kryeveteriner*), ***pensioníști*** "pensioners"(<Alb. *pensionistë*).

As previously stated, we have analyzed the borrowings into Aromanian from the Albanian language, based on Carol Meyers-Scotton's classification (Meyers-Scotton, 2005: 215). In the category of cultural borrowings, we have included words, which at the semantic level, designate new notions, either abstract or concrete. Cultural borrowings appear as a necessity, especially in the case of the rural communities of Aromanians, which began to come in contact with the new realities of the Albanian state once they have shifted from the nomadic life to the sedentary one and once they were integrated within the Albanian state borders after WWII.

Over time, the Aromanian speakers have been faced with the need to enrich their mother tongue with new lexical items, and sometimes replace older terms with newer ones because of the latters' superior functionality. This is why they have been labeled as core borrowings. The biggest number of core borrowings belong again to the semantic field of professions and social categories, some of them pointing to the more recent professional categories and social roles Aromanian have taken over in the new social context: e.g. ***măsús*** "teacher" (<Alb. *mësues*) used instead of ***dháscalu***, ***príft*** "priest"(<Alb. *prift*) used instead the Aromanian ***préftu***, ***pasonári*** "wealthy"(<Alb. *të pasur*), instead of ***avúți***, ***besimtári*** "believers" (<Alb. *besimtarë*) instead of ***pistipsíț/pistiméñi***, ***mbrétărlu*** "the king" (<Alb. *mbretri*) used instead of ***amirălu***, ***crúșci*** "inlaws" (<Alb. *krushq*) instead of ***cuscri***. Informants between 7 and 15 years old, as well as those between 30 and 50, males and females, both from rural and urban areas, made use of core borrowings up to 80%, whereas those with ages between 60 and 80, only up to 20%.

As in the case of morphologically and syntactically less integrated borrowings, sentence-level code-switches point to moments in communication when bilingual speakers switch from one code to another. They activate one idiom or another according to the identity they assume or the point of view they adopt when reporting (Albanian citizen or playing various professional roles within the Communist farming systems- such as vets, animals' caretakers, storekeepers, supervisors etc.). Thus, we have distinguished three types of thematic specialized texts in which the change of code acts as a mark of the change of perspective or the role that the individual assumes during the narrative: texts dealing with the story of a day at school, texts that include a visit to a physician, texts in which subjects describe professional activities from the past or present. Often the latter texts describe a modern, specialized profession where, besides the cultural loans needed to designate the new realities, there are also many cases of code switching at a lexical level.

Analyzing these examples, we may conclude that borrowings, either cultural or core, appear with all speakers regardless of age and gender. This demonstrates the strong influence of standard Albanian over the Aromanian dialect. Thus, with the male speakers, with a low level of formal education (usually just primary school), who have been professionally active, the rate of cultural borrowings is about 50% higher than in the case of the female speakers having the same level of education, but who have played the role of housewives all their lives. More than that, phenomena of code switching are quite frequent with these male informants, especially when they are narrating the professional roles they held during the Communist regime (in all oral histories narrated by male informants there are frequent code-switching phenomena both at lexical and syntactical level). Interestingly enough, code switching phenomena have been observed in the case of women aged between 60 and 75, a sign that certain verbal practices act constantly and impose themselves as preferred variants in certain contexts (e.g. when they want to communicate something personal to their husbands and are afraid of being overheard by people around).

Men and women aged between 30 and 50 living in towns, with a higher degree of education (high school or university) and having an active professional life very frequently resort to code switching (all our informants included in this category used code-switching at least twice at a sentence level). The same can be observed with children aged 7 to 15 who study in public schools only in the Albanian language, regardless of the environment, rural or urban. Generally, Aromanian children or youths living in urban areas tend to abandon their mother tongue because they are afraid of being discriminated by their Albanian colleagues and friends. According to our adult informants living in urban areas, their children were still active speakers of Aromanian when the latters' grandparents were still alive.

5. *Positioning*

Within all age groups we have identified the phenomena of positioning themselves in relation to the others: for example, the Aromanians, as a cultural and linguistic minority in opposition to Albanian majority speaking the prestige language. Here we find interesting the distinction made by the inhabitants of Albanian origin from Korça regarding the Aromanian community living here. Thus, they divide the Aromanians into two categories: *çobenj fushe* (in lit. translation "shepherds i.e. Aromanians from the plain") and *çobenj mali* ("shepherds i.e. Aromanians from the mountains"). By *çobenj fushe*, Albanians generally designate the old Aromanian inhabitants of Moscopole or Shipska, who had settled down in Korça centuries ago and had adapted themselves to the urban lifestyle, thus forming a kind of bourgeoisie before the Communist regime. *Çobenj mali* Albanians designate the Aromanians living in the communes and villages around Korça (Mborja, Drenova, Boboştiţa etc.) or those Aromanians who settled down relatively late in the city of Korça or, in a more general sense, those who have become sedentary late. However, about 70% of the Aromanians interviewed, with ages between 30 and 80, males and females, disagreed with this distinction, claiming that *çobenj fushe* designated the Aromanians, having their origins in Myzeqe Plain (in Albanian *fushë* meaning "plain"), and came to settle down in Korça area.

Furthermore, in relation to the Communist regime, in their oral histories, certain Aromanian informants place themselves in total opposition, for various reasons: first of all because they used to be rich (*pasonári* <Alb. *të pasur*) and the

Communist regime confiscated all their wealth (which included not only all their money, usually gold coins, Ar. *flurí*, but also their animals, their flocks of sheep, Ar. *tutípta*, the source of their income and the basis of their lifestyle); secondly because by establishing the separating state borders and banning travel abroad, the Albanian Communist regime separated them from other members of their family (including mothers, fathers, brothers, sisters etc.) who remained or were forced to remain on the other side of the border (e.g. the one dividing the Albanian state from the Greek state). There are some impressive stories in which they confess being reunited with all the members of their family after many years, when the Communist regime in Albania came to an end. This was the case of two old couples (with ages over 70) interviewed in Voskopojë.

Most of the villages in Upper Devoll are inhabited by Christians. Nevertheless, Sinica is the only village with religiously mixed population, whereas Arrza, located near the Christian villages, is entirely inhabited by Muslim population. We need to emphasize that the old inhabitants of Arrza, in their oral histories, use to say that they are cousins of Dardha and Sinica inhabitants (i.e. they used to be Christians too in the past). As far as the old Albanian inhabitants of Devoll, of Christian religion, living in Upper Devoll, are concerned, they usually define themselves in opposition to the Albanian inhabitants of Muslim religion living in Lower Devoll, whose speech does not contain some of the characteristics of the old Albanian language.

6. *Conclusions*

We have highlighted the advantages of using oral history recordings in order to illustrate regional dialects and speeches. Oral histories give us the possibility of mapping changes in dialects through time by comparing older and younger voices from the same area, as well as complex systems of meaning such as ideologies, social representations about group membership, social roles and attributes, presuppositions about all aspects of social reality, individual and collective stances, practices and organization structures. In sociolinguistics the primary focus for analysis is the original speech and thus accurate and consistent methods of transcription are necessary.

In the dialectological research, the creation of the digital archives for speeches is also very important, in order to highlight specific phonetic traits, as well as to preserve samples of speeches in danger of becoming extinct, as part of our cultural heritage.

As far as the transcription system is concerned, the adaptation of the transcription system for the Albanian oral texts is imperative: the IPA system is more useful since it allows us to use diacritical signs for various characteristics of the sounds.

Finally, the use of special computer programs may help only after a well established knowledge of the phonological system has been defined. Consequently, the use of a software may be helpful as part of a "computer-assisted transcription" procedure.

In the case of the Aromanian dialect, there is a stringent need for the unification of the transcription systems used in written media not only from Albania, but also from Bulgaria, Republic of North Macedonia etc.

Bibliography

Capidan, T. (1943). *Limbă și Cultură*. București: Fundația Regală pentru Literatură și Artă.

Capidan, T. (1931). *Fărșeroții. Studiu lingvistic asupra românilor din Albania*. Bucharest: Cartea românească.

Çabej, E. (1977). *Studime gjuhësore III*. Pristina: Rilindja.

Çabej, E. (1977). Ngulimet shqiptare në Itali dhe gjuha e tyre. In *Konferenca Kombëtare e Studimev Albanologjike*. Tirana 28-30 june 1976, 95-110.

De Fina, A. (1995). Pronominal choice, identity, and solidarity in political discourse. In *Text*, 15(3), 379-410.

De Fina, A. (2003). *Identity in Narrative. A Study of Immigrant Discourse.* Amsterdam: John Benjamins.

De Fina, A. (2006). Group identity, narrative and self representations. In De Fina, A., Schiffrin, D. & Bamberg, M. (Eds), *Discourse and Identity*. Cambridge: Cambridge University Press, 351-375.

Georgakoupoulou, A. (2007). *Small Stories, Interaction, and Identities*. Amsterdam: John Benjamins.

Gjinari, J. (1960). Mbi të folmen e Devollit. In *Bul. I Univ.Shtet. të Tiranës, Seri e shkencave shoqërore*, 4.

Gjinari, J. (1972). Mbi disa elementë sllavë në të folmet veriore të Shqipërisë jugore. In Kostallari, A. (Ed), *Studime mbi leksikun dhe formimin e fjalëve II*. Tirana: Universiteti Shtetëror i Tiranës, 267-279.

Floqi, S. (1971). Pyetësor për mbledhjen e materialit dialektor të gjuhës shqipe. In Ginari, J. (Ed), *Dialektologjia shqiptare I*. Tirana: IGJSH, ASHSH, 371-447.

Gjinari, J., Beci, B., Shkurtaj, Gj. & Gosturani, Xh. (2007). *Atlasi dialektologjik i gjuhës shqipe I*. Napoli: Università degli studi di Napoli "L'Orientale". Dipartimento di studi dell'Europa orientale.

Gjinari, J., Beci, B., Shkurtaj, Gj. & Gosturani, Xh. (2008). *Atlasi dialektologjik i gjuhës shqipe II*. Napoli: Università degli studi di Napoli "L'Orientale". Dipartimento di studi dell'Europa orientale.

Hopf, C. (1873). *Chronique gréco-romane inédites ou peu connues publiées avec les notes et tables généalogiques.* Berlin: Librairie de Weidmann.

Meyers-Scotton, C. (2005). *Multiple Voices. An Introduction to Bilingualism*. Oxford: Blackwell Publishing.

Nevaci, M. (2011). *Graiul aromânilor fărșeroți din Dobrogea*. Bucharest: Editura Universitară.

Qezari, N., et al. (2004). *Filmografi e filmit shqiptar 1953-2003, vëllimi I*. AQSHF Tirana: Toena.

Schiffrin, D. (1996). Narrative as self-portrait: sociolinguistic construction of identity. In *Language in Society*, 25(2), 167-203.

Shkurtaj, G. (1999). *Sociolinguistika*. Tirana: SHBLU.

Topalli, K. (2007). *Fonetika historike e gjuhës shqipe*. Tirana: Dituria.

Van Dijk, T. (2010). Political identities in parliamentary debates. In Ilie, C. (Ed.), *European Parliaments under Scrutiny: Discourse Strategies and Interaction Practices*. Amsterdam: John Benjamins, 29-56.

Wilson, J. (1990). *Politically Speaking: The Pragmatic Analysis of Political Language.* Oxford: Blackwell.

Xhaçka, V. (1958). Një shikim mbi të folmen e krahinës së Devollit. In *BUSHT, SSHSH*, 2.

Xhaçka, V. (1959a). Disa zakone në festa kalendarike popullore në Devoll. In *Bul. I Univ. Shtet. të Tiranës, Seri e shkencave shoqërore*, 2.

Xhaçka, V. (1959b). Lindja dhe martesa në Devoll. In *Bul. I Univ.Shtet. të Tiranës, Seri e shkencave shoqërore*, 1.

Zdruli, K. (1987). *Folklor nga Devolli i Sipërm*. Tirana: Botim i ASHSH.

Appendix 1

Dialectal text from Upper Devoll

Transcription with Albanian Alphabet

e me çupë ce pe të Rënjave je tinë// dhe unë ngaj Bradvica jam// jam martu̯arë ktu Vërlen// u arrova// më e voglë se kto çupat isha/// atere martoshim çpejt// kam xbritur nuse këtunë që para luftës// aterë/ kur martoshim/ vishmin xibunë/ unë kisha xibunë e mamasë// bë tërë anët me klinda/// kur jam martu̯arë këtu dë fshat kishte shumë të krishterë// tani kemi mbetur pak/// tërë kanë ikur dë Greqit// ja katrë shpi kemi mbetur// dhe bë shpit time bënim poçe/// tani ktu dë Vërlen ëshçë vetëm një djalë që bën// po zˈdi shumë si ata të qëmoçmit// e kush i mer tashinë// blenë ngaj këto plastiket për kollajxhe ce zˈdinë që bulmeti ëshçë më mirë dë poçet e baltit///

(Androniqi Pine, *91 vjeç, e krishterë, pa arsim, lindur në Bradvicë, martuar në Vërlen*)

Transcription with IPA

e me ʧupə ʧe pe tə ɾənjave je tinə// ðe unə ngaj bɾadvica jam// jam maɾtu̯aɾə ktu vəɾlen// u aɾova// mə e voglə se kto ʧupat iʃa/// ateɾe maɾtoʃim ʧpejt// kam ʥbɾituɾ nuse kətunə cə paɾa luftəs// ateɾə/ kur maɾtoʃim/ viʃmin ʥibunə/ unə kiʃa ʥibunə e mamasə// bə təɾə anət me klinda/// kuɾ jam maɾtu̯aɾə kətu də fʃat kiʃte ʃumə tə kɾiʃteɾə// tani kemi mbetuɾ pak/// təɾə kanə ikur də gɾecit// ja katɾə ʃpi kemi mbetur// ðe bə ʃpit time bənim potʃe/// tani ktu də vəɾlen əʃʧə vetəm ɲə djalə cə bən// po zdi ʃumə si ata tə cəmoʧmit// e kuʃ i mer taʃinə// blenə ngaj kəto plastiket pəɾ kołajʤe ce zdinə cə bulmeti əʃʧə mə miɾə də potʃet e baltit///

[Hey girl, from Rënja's family are you// I'm from Bradvica too// I'm married here, in Vërlen// I was younger than these girls///back then we used to marry very young// I came down here as bride before the war// when we got married we used to wear xibun (a kind of wool manteaux with embroidery)// I had my mom's xibun// it was all over the sides with crinkles/// when I got married, there were many Christians here in this village//now we are few/// they have all fled to Greece// we are only four families left// we have made pots in my house too/// now it's only a guy that makes them (pots)// but he doesn't know how to make them as well as our ancestors// eh, who is taking (buying) them now// it's easier for all to buy these plastics but they don't know that dairy is better in clay pots///]

(Androniqi Pine, *91 old, Christian, without formal education, born in Bradvicë, married in Vërlen*)

Appendix 2

Dialectal text in Farsherot Speech from Korça

noi̯ ĉilimáńii̯ ţe̯-avémụ cρăscútă cu papụ ş-cu dádă u̯-avémụ grítă límba / límba nǫ́stρă di căndụ aρámụ ĉilimáń/ di căndụ adukímụ ţi zbuρámụ / ş-míni u̯-amụ zbuρắtă că míni avi̯ámụ pápu acásă / avi̯ámụ şi păρínţyi / ma avi̯ámụ ş-pápu / ş-cu pápu detürimişti̦ u zbuρám vláhica / ose rămänę́şte ţi cumụ ḑăţémụ noi̯ rămắńi / şă ş-nafǫ́ρă căndụ zbuρámụ nafǫ́ρă noi̯ avi̯ámụ únă θécsi [**nuc e di si iθónă**] avi̯ámụ únă θécsi áltă cu límba / cu límba ţi zbuρám ụ / aρvanicá/ şń n-adukámụ /şń míni di fumę́ya nǫ́stρă / míni mi̯-adukéstụ ma múłtă di θécsi / θécsi áli ... áli [k] alụ zbóρoρụ / şń u̯-avę́mụ zbuρắtă múltă límba / ş-míni nu u̯-amụ agńrşę́tă ş-nu pǫ́ti su̯-agńrşéstụ tútă bána / ş-fiĉóρu a ńeu̯ căndụ aρá ma ńicụ avi̯á dádă ş-papụ ş-cu păρínţyi a ńei̯ zbuρá multụ ĝíni / ma dicáρa eyi muρíρă fiĉóρu nu u zbuρá / u̯-avi̯á ca ... ca cu sicléti s-u zbuρá / vétăm zburá nafǫ̯áρă cu sóţăl' / d-apói̯a u-nviţắ italíşti / nviţắ anglíşti / nviţắ frănĝíşti / ş-nu că nu u̯-adukę́şti / u̯-adukę́şti multụ búnă / ş-ári zbǫ́ρă ţi li ḑắţi multụ búni / ma căndụ aρá ńicụ zbuρá multụ ... multụ búnă/ u zbuρá ca míni / tóρa nu u zbuρắşte / ş-noi̯ tóρa căndụ avi̯ámụ păρínţăyi /aρámụ ca detürit ta s-u zbuρắmụ mult búnă límba / tóρa u̯-avémụ ałăsátă nih´ámă / cu tútă aţá límba ţi-i nviţátă di cu ńícụ nu s-agńrşę́şti puté // tóρa să spunụ ti lucρụ / míni estụ măρtátă / amụ un fiĉóρụ / tóρa s-fáţi di γ´íṇγ´iţ ań / bằρbátu ę́sti mecanícụ / ári lucρátă la únă fabrícă ţi făţi̯á păρρóḑ / di cáρa s-fę́ţiρă d-apói̯a privatizími / noi̯ avémụ lu̯átă dáu-tρe makíni / ş-avę́mụ zdăfáptă un lucρụ de̯-anóstụ / lucρámụ dól'i: băρbátu ş-míni / elụ lucρáḑă ş-dipụ la aţę́li mecaníkili că míni nu li şti ahắntụ ma łu̯-aĝútụ ş-míni / avémụ zdăfáptụ aístụ lucρụ / avémụ vắρă ḑáţi ań ţi lucρắmụ / ş-h´imụ ĝíni / avémụ sculátă nih´ámụ di iconomiíi̯ / ş-ţi să-ţ ḑăcụ altţivá //

[we who grew up with our grandparents/ we have spoken our mother tongue since we were kids/ I have spoken it too because I grew up with my grandpa at home/ I had my parents too/but I also had my grandpa/ to whom I obligatorily spoke in our idiom or Rumanian, as we, the Rumanians, call it/and outside, when we were speaking outside, we had a certain accent [Alb. I don't know how we call it]/ we had a different accent from the Albanians/ so we were different/and in my family I have a different accent/ when I speak/ and we all have spoken our mother tongue/ and I haven't forgotten it/ and I am not going to ever forget it as long as I live/ my son, when he was a child, grew up with his grandparents and he used to speak with them a lot/but when they died, he stopped speaking/he started having … difficulties, speaking his mother tongue/ he used to speak it only outside with his friends/ and later he learnt Italian/English/French/ now he understands it very well/and even speaks some words/ but when he was a child / he used to speak it very well/ just like me/now he doesn't speak it anymore/when we had our grandparents/ we were somewhat forced to speak it well/ now we have abandoned it a little bit/ Nevertheless the language you have learned since you were children can never be forgotten/ now let me tell you about my work/ I am married/ I have a son/ now he turns twenty and/ my husband is a mechanic/ he used to work in a factory which made socks/ after that it was privatized/ we have bought 2-3 machines/ and we have started our own business/ we work together now/ my husband and I / he takes care of the mechanical things I don't know anything about/ but I help him/we have been running this business for ten years now/ and we are doing well/ we have made some savings/ what else can I say…]

Appendix 3

Transcription System for the Farsherot Speech of the Aromanian Dialect of the Romanian language

a = the most open vowel from the central series (a – ă – î)
ă = open-mid vowel from the central series (a – ă – î)
ă̧ = open *ă* (between *ă* and *a̧*)
â = vowel from the central series, between *ă* and *î*, closer to *î*
b = voiced bilabial plosive
c = voicelss velar plosive
ĉ = soft voiceless post-alveolar affricate
d = voiced dental plosive
ḑ = voiced alveolar affricate
δ = voiced interdental fricative
e = open-mid vowel from the front series
e̯ = semivocalic e
ę = open *e* (between *e* and *e̯*)
f = voiceless labio-dental fricative
g = voiced velar plosive
ĝ = soft voiced post-palatal affricate
ǵ = voiced palatal plosive
γ = voiced velar fricative

γ́ = voiced palatal fricative
h = voiced velar fricative
ɧ = voiced pharyngeal fricative
h́ = voiceless palatal fricative
i = the closest vowel from the front series
į = open *i*
i̩ = short *i*
i̯ = semivocalic *i*
î = the closest vowel from the central series
j = voiced prepalatal fricative
jˈ = soft *j*
k = voiceless velar plosive identical with *c*; the symbol is used before the vowels from the front series (e.g.: *kepţu̩,* "chest" in the Grabova speech)
ḱ = voiceless palatal velar plosive
l = lateral alveolar *l*
lˈ = lateral palatalized *l*
ł = lateral velarized *l*
m = nasal bilabial plosive
n = nasal dental plosive
ň = nasal plosive slightly palatalized
ń = nasal palatal plosive
ŋ = velar nasal plosive, followed by *c, g*
o = mid-open vowel from the back series
o̯ = semivocalic *o*
ǫ = open *o* (between *o* and *o̯*)
p = voiceless bilabial plosive
r = apical-alveolar trill
r̄ = *r* with several trills
ρ = uvular *r*
ρρ = ρ with several trills
s = voiceless alveolar fricative
ş́ = voiceless prepalatal fricative
b = soft *ş*
t = voiceless dental plosive
θ = voiceless interdental fricative
ţ = voiceless alveolar affricate
u = the closest vowel from the back series
u̩ = short *u*
u̯ = semivocalic *u*
v = voiced labio-dental fricative
w = voiced bilabial fricative
y = voiced palatal fricative (between γ́ and *i̯*)
z = voiced alveolar fricative

Diacritical signs

´	= (above a vowel or a diphthong) main stress: *cásă* "house", *lúcru* "work", *pátru* "four"
~	= (above a vowel) nasality: ĩşí "ieşi", Eng."he(she) went out"
̥	= (under voiced vowels or consonants) voiceless: ị
̨	= (under vowels) openness: į
̑	= (under vowels) it signals that the vowel is the consonant element of a diphthong; e.g: *i̯áspi* in standard Romanian "viespe", Eng. "wasp", *l'au̯* in standard Romanian ortography "iau", Eng. "I take"
ị, ậ, ụ etc.	= slightly audible sounds

DOI: 10.17469/O2106AISV000008

GRAZIANO TISATO

Acquisizione dell'intero AIS (*Sprach- und Sachatlas Italiens und der Südschweiz*)

The paper presents the results of the NavigAIS project, born to acquire the text of the whole AIS, the Linguistic and Ethnographic Atlas of Italy and Southern Switzerland (*Sprach- und Sachatlas Italiens und der Südschweiz*) (Karl Jaberg and Jakob Jud, 1928-1940). The project was developed between 2009 and 2019 in three complementary phases: 1) the realization in 2009 of NavigAIS, a high resolution digital version of the AIS atlas, provided with a search and navigation engine, to allow the exploration of the 1705 maps contained in the atlas (www3.pd.istc.cnr.it/navigais). NavigAIS assured the real time access to the AIS maps to check on the field the speaker's answers, during the recording sessions of AMDV (the Multimedia Atlas of Veneto Dialects) (www.pd.istc.cnr.it/amdv) in 2009-2010, and to implement the database of the AMDV lemma and the related AIS lemma collected in 1921 in the Veneto region; 2) the NavigAIS online version in 2014 (www3.pd.istc.cnr.it/navigais-web), which does not require a software installation; 3) In 2016, within the *AIS Reloaded* project (www.rose.uzh.ch/de/forschung/forschungamrose/projekte/AIS-reloaded.html), NavigAIS has been provided with a specific OCR, to acquire the AIS text in acceptable times. The estimate is to complete 50% of the AIS tables by the end of 2019.

Key words: linguistic geography, AIS atlas, ancient document acquisition.

1. *Introduzione*

Si presenta in questo lavoro il progetto NavigAIS di acquisizione digitale del testo dell'intero AIS, l'Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale (*Sprach- und Sachatlas Italiens und der Südschweiz*) (Jaberg, Jud, 1928-1940), e si discutono le problematiche e le soluzioni adottate nel software implementato a questo scopo fra il 2009 e 2019.

Si illustrano le tre fasi in cui si è sviluppato il progetto NavigAIS:

1. La versione digitale e navigabile dell'AIS, completata nel 2009 (Tisato, 2010) www3.pd.istc.cnr.it/navigais, per la realizzazione dell'Atlante Multimediale dei Dialetti Veneti (AMDV) www.pd.istc.cnr.it/amdv.
2. La versione online, consultabile in rete senza necessità di installazione) realizzata nel 2014, www3.pd.istc.cnr.it/navigais-web.
3. La versione con OCR (*Optical Character Recognition*), implementata nel 2015-2016 per il progetto *AIS Reloaded* (Michele Loporcaro e Stephan Schmid, Università di Zurigo, finanziato dal SNSF *Swiss National Science Foundation*, https://www.rose.uzh.ch/de/forschung/forschungamrose/projekte/AIS-reloaded.html; vedi negli atti di questa conferenza: Negrinelli, Donzelli, *Il pro-*

getto AIS reloaded: un archivio sonoro per 36 varietà dialettali della Svizzera meridionale).

Si discutono le caratteristiche del simbolismo AIS, l'architettura e le funzionalità del sistema, le modalità di acquisizione del testo con l'OCR, la validazione dei risultati, la valutazione obiettiva del riconoscitore, ed i tempi di acquisizione dei lemmi.

Si propone, infine, NavigAIS come un modello facilmente adattabile ad altri atlanti linguistici storici. L'implementazione può avvenire in due fasi complementari: una iniziale più rapida, che opera la scansione e il trattamento digitale delle pagine di un atlante, lo indicizza e lo rende navigabile. La seconda forzatamente più lenta, ma più interessante dal punto di vista linguistico, che acquisisce tutto il contenuto testuale dell'atlante con un OCR predisposto al compito. Questo approccio evita che una miniera di informazioni linguistiche ed etnografiche finisca per rimanere sepolta ed inutilizzata negli atlanti cartacei, per i problemi legati alla conservazione dei volumi, alle limitazioni di accesso delle biblioteche, e alle difficoltà di ritrovare l'informazione cercata per mancanza di indicizzazione.

Fig. 1 - *NavigAIS: versione con OCR (NavigAIS-web K. 548.31)*

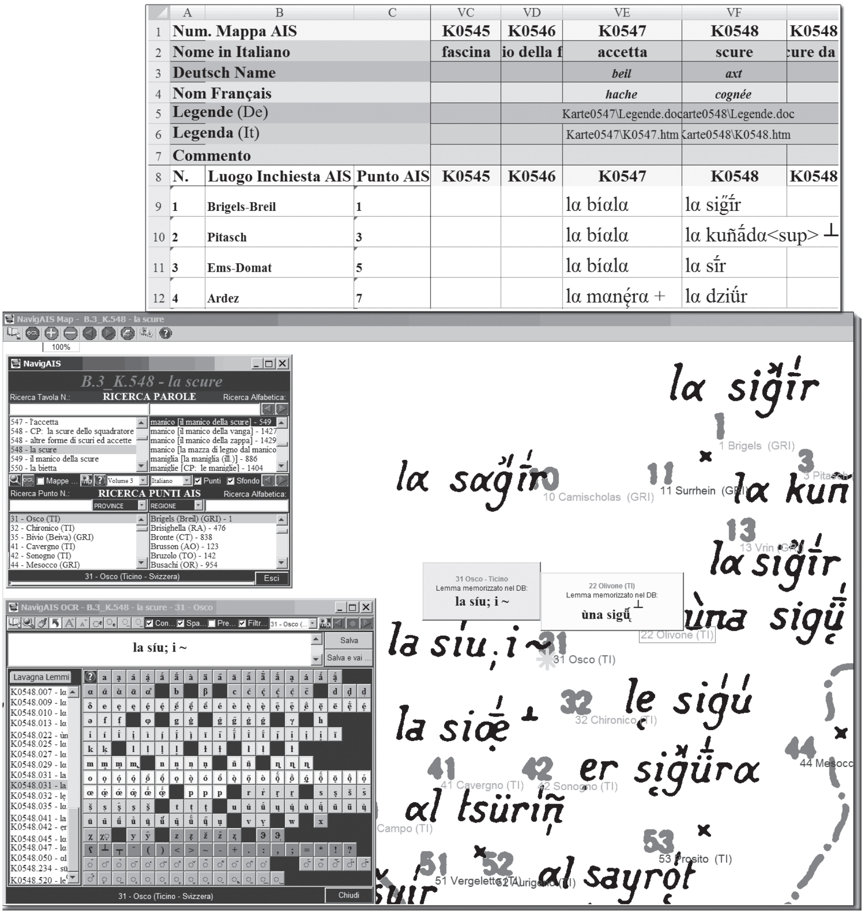
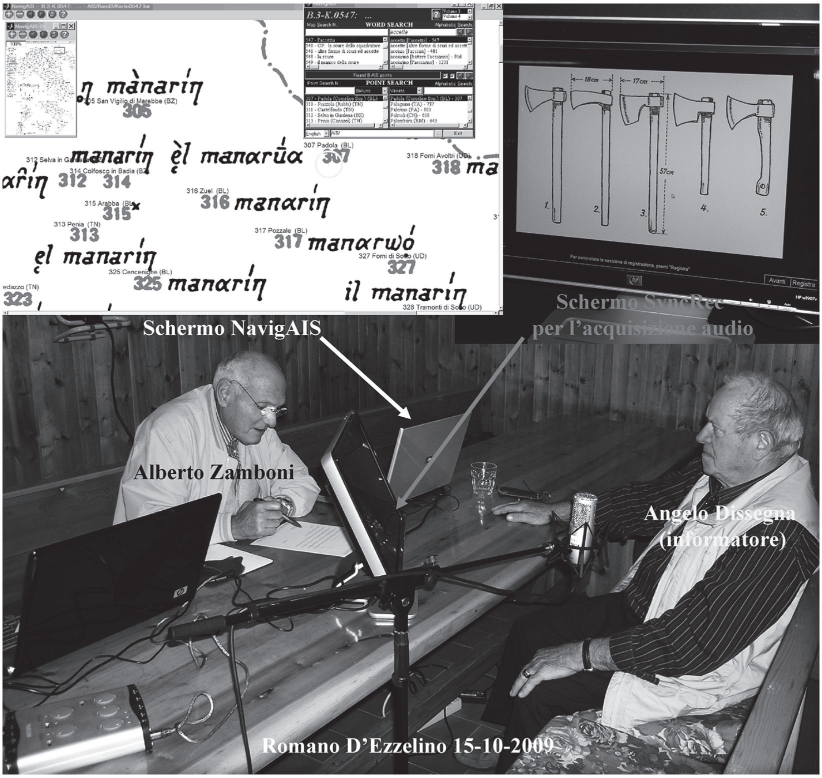


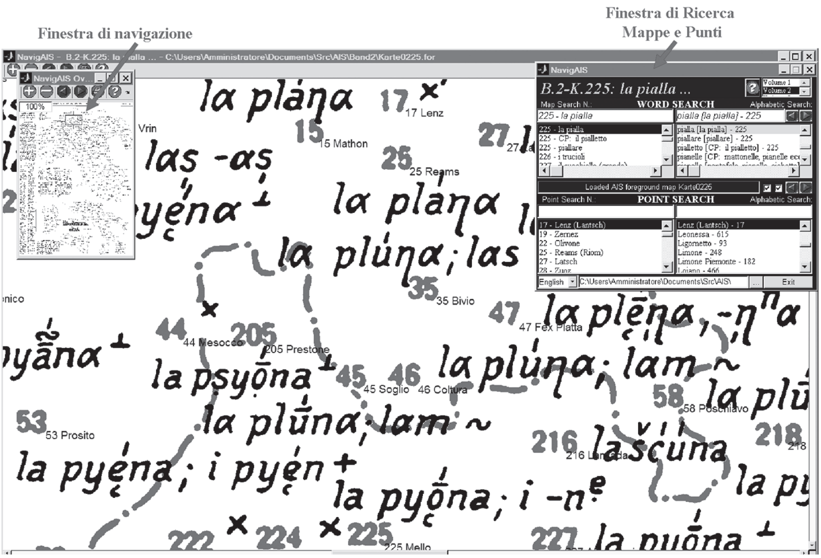
Fig. 2 - *Inchieste AMDV: a sinistra, la postazione del dialettologo, che può confrontare le risposte dell'informatore con quelle AIS sulla mappa di NavigAIS (NavigAIS-web K. 547.307). Il software SyncRec registra l'audio e mostra sequenzialmente l'oggetto corrente all'informatore sullo schermo di fronte a lui, e al responsabile audio sullo schermo nell'angolo sinistro in basso*



2. *NavigAIS Prima Fase: Versione Desktop*

NavigAIS è nato nel 2009 per soddisfare le specifiche esigenze del progetto dell'Atlante Multimediale dei Dialetti Veneti (AMDV, www.pd.istc.cnr.it/amdv) (Tisato, Barbierato, Ferrieri, Gentili & Vigolo, 2013). L'AMDV è stato realizzato fra il 2010 e il 2015 presso il Dipartimento di Discipline Linguistiche, Comunicative e dello Spettacolo dell'Università di Padova, e l'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR di Padova, con un finanziamento della Cassa di Risparmio di Padova e Rovigo per i progetti d'eccellenza del 2008.

Fig. 3 - *NavigAIS: versione desktop (2009) (NavigAIS-web K. 225.35)*



L'idea portante dell'AMDV era il confronto diacronico fra il lessico registrato per l'AMDV nel 2009-2010 e quello raccolto nel Veneto nel 1921 dall'AIS. L'AMDV trae ispirazione da recenti atlanti parlanti: ALEPO (Telmon, Canobbio, 1985), ALD (Goebl, 1994), VIVALDI (Kattenbusch, 1995), ecc., ed in particolare da un precedente lavoro realizzato dall'autore, su scala più ridotta, per i dialetti trentini *Il Trentino dei contadini* (Mott, Kezich & Tisato, 2003). Rispetto all'approccio puramente lessicale dei tradizionali atlanti cartacei, l'AMDV si proponeva di recuperare la dimensione sonora del parlato dialettale e di integrare tutti gli aspetti fonetici, etimologici ed etnografici dei dialetti veneti, in un insieme che permettesse di capire una realtà complessa e articolata come quella dialettale.

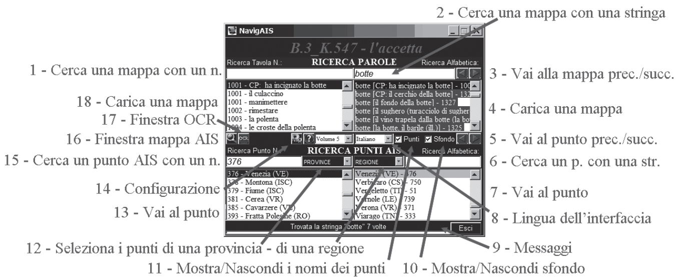
La versione navigabile dell'AIS si rese necessaria non appena divenne evidente che la consultazione dell'atlante non poteva avvenire sui volumi cartacei senza una perdita di tempo enorme, tenuto conto degli orari e delle limitazioni di accesso (feste, ferie, orari, ecc.) della biblioteca del dipartimento universitario, che custodiva i preziosi volumi.

Anche la semplice fotocopia delle pagine AIS era proibitiva, poiché la dimensione fuori del comune (44x58 cm, circa un foglio A2) obbligava a più passaggi per coprire, con un certo margine utile, l'intera superficie, e richiedeva la collaborazione di due persone per girare sottosopra il volume, dargli l'orientamento più appropriato, mantenere la pagina complanare sulla superficie della fotocopiatrice e spostarla poi in un nuovo settore.

Era impossibile pensare di ripetere questa operazione per le 1705 pagine dell'AIS su una normale fotocopiatrice A2 o A3, sia per il problema non trascurabile della conservazione dei volumi originali, sia per la difficoltà operativa di ripescare poi, qualora servisse, una pagina (o una sua parte) in una pila di migliaia di fogli.

La decisione di affrontare alla radice il problema della consultazione dell'Atlante AIS portò nel 2009 a realizzare NavigAIS (Figg. 3, 4), come un progetto completamente autonomo rispetto all'AMDV, quando questo ancora non aveva mosso i primi passi.

Fig. 4 - *Finestra di NavigAIS per la ricerca delle mappe e dei punti AIS. L'esempio mostra i risultati della ricerca della stringa "botte" nell'indice delle mappe AIS*

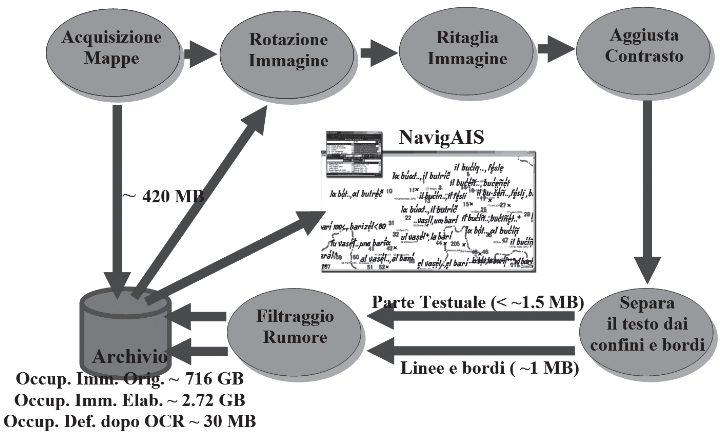


L'operazione preliminare fu la scansione a 600 dpi delle pagine dell'AIS su una macchina professionale messa a disposizione dal Comune di Padova.

L'elaborazione successiva è schematizzabile in 5 fasi (Fig. 5), che possono essere eseguite in un'unica sequenza automatizzata oppure isolatamente, una fase alla volta, sotto il controllo di un supervisore. Al software era richiesto prima di tutto di correggere la rotazione della pagina, dovuta al posizionamento del volume sullo scanner, che non risulta mai quello ottimale. L'aggiustamento della pagina è richiesto per una corretta riproduzione dell'immagine, ma è indispensabile soprattutto per il processo di riconoscimento automatico dei caratteri (OCR), che si prevedeva di attuare in un secondo tempo per acquisire l'intero testo AIS. Dopo la rotazione della pagina, l'immagine era automaticamente ritagliata in corrispondenza dei bordi esterni, in modo da ridurre le dimensioni finali al minimo possibile e l'inclusione di parti della pagina confinante.

Il terzo passo prevedeva un aggiustamento automatico del contrasto dell'immagine con la conseguenza di virare i colori verso le tonalità più scure.

Fig. 5 - *NavigAIS: acquisizione ed elaborazione delle mappe AIS (NavigAIS-web K. 1325.1)*



Il quarto passo serviva a separare il testo dagli altri elementi (bordi, confini regionali, linee orografiche, simboli e numeri indicanti le località dell'Atlante) con un'operazione di sottrazione della parte colorata in arancio, ed a restituire le due mappe complementari della pagina originale. Questo passaggio elimina tutte le componenti di colore inferiori ad una soglia prefissata (tanto per il testo in primo piano quanto per il background), spazzando via completamente lo sfondo rosa pallido ed il rumore di fondo, dovuto, ad esempio, alla comparsa in trasparenza del testo del verso della pagina.

Subito dopo si procedeva ad una ripulitura del "rumore" casuale, e cioè dei punti isolati, con dimensioni inferiori ad una quantità prefissata (il cosiddetto filtraggio *Salt & Pepper*).

L'ultimo stadio salvava le due componenti dell'immagine come maschere binarie fatte unicamente di 0 e 1 ("binarizzazione" dell'immagine originale), per indicare la presenza del colore nero oppure del colore arancio. Questo stratagemma ha permesso di ridurre l'occupazione totale delle 1705 pagine (più le prefazioni) dell'intero Atlante, che in partenza era di 716 GB, a circa 2.72 GB (circa 270 volte inferiore all'occupazione originale).

In base ai dati ottenuti con il progetto AMDV e con il progetto *AIS Reloaded*, la dimensione finale di tutto l'AIS testuale risulterà inferiore a 50 MB.

Una volta ottenute le immagini grafiche delle pagine AIS, è stato messo a punto il software di navigazione in una prima versione *stand-alone* (o *desktop*, che richiede quindi un'installazione sul proprio computer) con l'indicizzazione degli argomenti delle tavole AIS e con un sistema di ricerca delle mappe e dei punti di inchiesta.

Fig. 6 - *AMDV: Ricerca di un tratto fonetico (nasalità) nel database 2009-2010, riscontrato in 1972 lemmi dei 14521 acquisiti in 26 punti di inchiesta del Veneto*

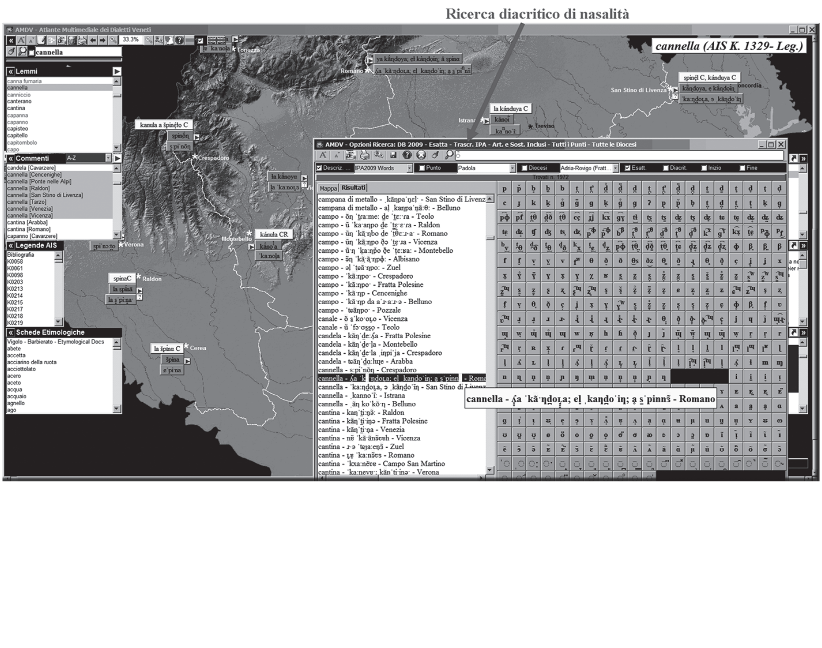
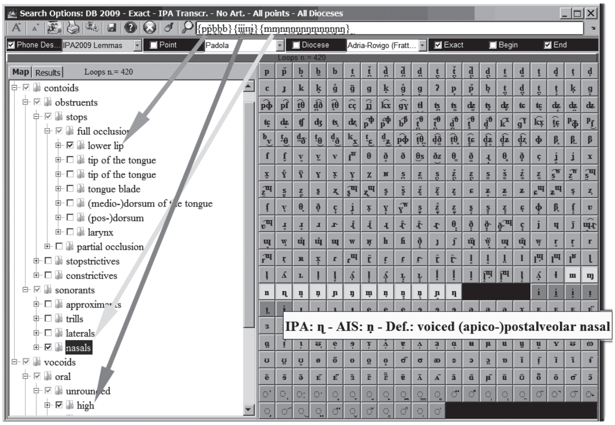


Fig. 7 - *AMDV: Ricerca combinatoria di tutte le occlusive bilabiali, seguite dalle vocali alte e dai contoidi nasali, per un totale di 420 combinazioni distinte*

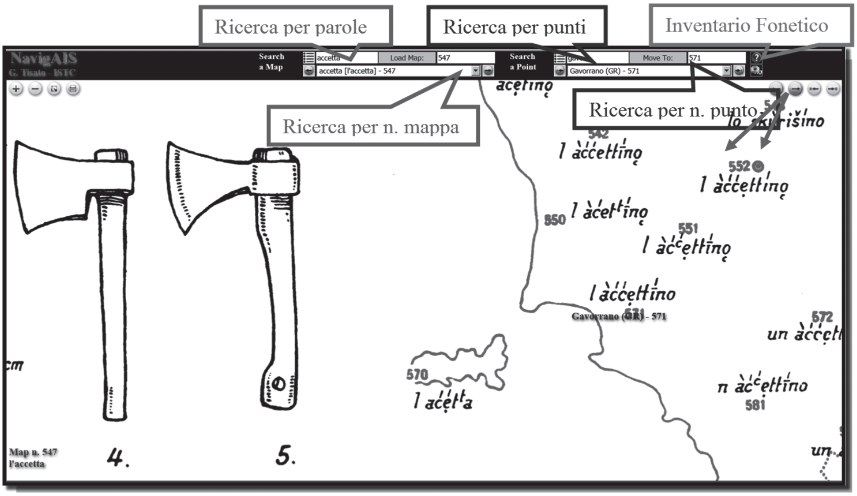


Benché progettato con l'obiettivo finale a lungo termine dell'acquisizione del testo con l'OCR, gli scopi più immediati erano due:

- NavigAIS doveva permettere il confronto in tempo reale sul campo delle risposte degli informatori AMDV con quelle raccolte dall'AIS nel 1921. La Fig. 2 mostra una tipica sessione di lavoro AMDV sul campo: durante l'inchiesta il dialettologo di turno aveva sotto gli occhi le risposte raccolte dall'AIS novant'anni prima, in quello stesso punto e nei punti vicini, e poteva quindi interagire efficacemente con l'informatore dei nostri giorni.
- NavigAIS era anche essenziale per costruire il database delle risposte AIS del 1921 in tempi rapidi e con un minimo tasso di errore. Una volta salvata in un database, la trascrizione corrente poteva essere immediatamente visualizzata fianco a fianco con quella dell'immagine grafica originale, in modo da permettere di correggere gli eventuali errori. Questa strategia ha consentito di realizzare la trascrizione in due mesi di lavoro con un errore inferiore al 2%.

Visto l'interesse che NavigAIS poteva avere per studenti, ricercatori ed esperti della materia, nel gennaio del 2010, l'atlante fu reso disponibile in un sito web: www3.pd.istc.cnr.it/navigais.

Fig. 8 - *NavigAIS-web versione online dell'AIS (2014): le mappe AIS e i punti d'inchiesta sono accessibili con parametri aggiunti al normale indirizzo web (NavigAIS-web K. 547.571)*



3. *NavigAIS Seconda Fase: Versione Online*

La seconda fase, implementata nel 2013-2014, ha permesso di realizzare una versione online dell'atlante AIS (www3.pd.istc.cnr.it/navigais-web) (Fig. 8), che ha il vantaggio di non richiedere una installazione del software ed è consultabile in rete da qualsiasi parte del mondo.

La novità rilevante, rispetto alla versione desktop del 2009, è la possibilità di aprire in rete una certa mappa AIS in un certo punto d'inchiesta, per evidenziare il lemma voluto direttamente dal proprio computer o dalla propria applicazione mediante tre parametri, inseriti nell'indirizzo web:

– map = xxxx (xxxx è il numero della tavola AIS 1-1705)
– point = yyy (yyy è il n. identificativo del luogo di inchiesta 1-990)
– loc = luogo (luogo è il nome del luogo di inchiesta)

Si forniscono alcuni esempi concreti di utilizzo dei parametri:

– www3.pd.istc.cnr.it/navigais-web (apre la mappa n. 1 nel punto n. 1, Brigels);
– www3.pd.istc.cnr.it/navigais-web?map=1401 (apre la mappa n. 1401, nel punto n. 1);
– www3.pd.istc.cnr.it/navigais-web?map=1401&loc=Teolo (mappa n. 1401, punto Teolo);
– www3.pd.istc.cnr.it/navigais-web?map=1434&point=229 (apre la mappa n. 1434 nel punto n. 229, corrispondente a Sonico).

La seconda novità è l'inserimento nel software dell'inventario fonetico AIS, tratto dai Cap. 3-4 del volume introduttivo dell'AIS (Jaberg, Jud, 1928), con link diretti ad esempi concreti sulle tavole AIS, sfruttando il meccanismo di indirizzamento appena spiegato, *www3.pd.istc.cnr.it/navigais-web/AIS_symbols.htm*.

Fig. 9 - *NavigAIS-web: Inventario fonetico con esempi* (*www3.pd.istc.cnr.it/navigais-web/AIS_symbols.htm*)

*pag. 42-55	Simboli Fonetici	Esempi/Examples	Phonetic Symbols
ʕ	Occlusiva glottidale – "coup de glotte", tipica di 947 Fonni (*pag. 46).	K. 1325 – botte 947 Fonni **ṣa ˤarrā́δa …**	Laryngeal stop typical of 947 Fonni (*pag. 46).
b	Come in italiano.	K. 1427 Leg. – vanga 346 Tarzo **la badī́la; ẹl badī́l**	As in Italian.
ḅ	/b/ lene (*pag. 53).		Lenition of /b/ (*pag. 53).
β	/v/ bilabiale = /b/ come in sp. "haba".	K. 1072 – il branco di pecore 955 Laconi **una γẹ́ḍḍa de eβrẹ̄́γeṣ**	Bilabial /v/ = /b/ as the Sp. "haba".
ć	Come it. "cena".	K. 1229 – cerchio della ruota 376 Venezia **I. ũⁿ ṣẹ́rćọ**	As the It. "cena" ("dinner").
ć̣	Simile a /tṣ/ – Vicino a /ć/.	K. 1664 – l'afferrò per il collo 322 Tuenno **l a ć̣apā́ pẹr ɑl kọ̄́l**	Sound between /tṣ/ and /ć/.
ć̩		K. 982 – cucchiaio 372 Raldon **el kuć̩ár**	
ć̯		K. 1642 – aspetta un tantino 42 Sonogno **šp̌ẹ́ć̯ ụm mọ́mẹntíṇ**	
ć̥	Suono derivato da *tr* e rilevato da Scheu. a 222 Germasino (*pag. 48).	K. 1585 – un cappello 222 Germasino **üṇ kapẹ́l inć̥a butẹ̄́ǧ̥ọ**	Sound derived from *tr* observed by Scheu. in 222 Germasino (*pag. 48).
c̋	= *tg* in soprasilv. *latg* = "latte". Suono schiacciato mediopalatale.	K. 966 – secchio di legno 374 Teolo **sẹ́c̋a dẹ ẹ́ñọ**	= *tg* in soprasilv. *latg* = "latte".
d	Come in italiano.	K. 901 – armadio 363 Vicenza	As in It…

4. *NavigAIS Terza Fase: Versione con OCR Integrato*

La terza fase del progetto NavigAIS riguarda l'acquisizione del testo dell'intero Atlante AIS. Questa fase ha lo svantaggio di richiedere tempi di esecuzione lunghi, sia per l'addestramento e l'implementazione dell'OCR, sia per la correzione degli errori del riconoscitore, ma ha d'altra parte il vantaggio enorme di generare un database di dati interrogabili di grande interesse per tutto il campo della dialettologia romanza.

Preliminare a questo passo è la scelta di una rappresentazione dell'informazione che sia adatta alle necessità linguistiche e computazionali successive. Per poter codificare il lessico AIS, che impiega simboli latini e greci ed un numero molto alto di diacritici, non si può ricorrere ad un font specifico (o proprietario), per alcune buone ragioni: *a*) significherebbe ricorrere a diverse centinaia di caratteri diversi l'uno dall'altro (con un aumento delle difficoltà di addestramento dell'OCR); *b*) creerebbe grosse difficoltà computazionali nella ricerca di una stringa di caratteri e difficoltà nel digitare la sequenza corrispondente al carattere voluto; *c*) comporterebbe la necessità di installare il font sulla macchina su cui deve girare il programma applicativo (che richiede a sua volta una fase di installazione), e la conseguente preclusione dell'utilizzo in rete dell'informazione, qualora si volesse una versione online del software.

La soluzione a questi problemi è l'adozione di un font generico, basato sullo standard Unicode, che abbia un insieme di diacritici sufficientemente esteso da coprire tutte le necessità dell'inventario fonetico che si voglia usare. Il Times New Roman, ad esempio, a partire dalla versione 5.0, ha 112 diacritici, sufficienti per esprimere la grafia AIS abbastanza fedelmente. Il vantaggio di questa soluzione è che permette di implementare l'algoritmo di ricerca di una parola e delle sue varianti in maniera molto più semplice ed efficace, consentendo all'utilizzatore di cercare una sequenza o un tratto fonetico anche senza presupporre la conoscenza a priori della combinazione di tasti per richiamare un simbolo specifico (Figg. 6, 7).

Nell'AMDV, ad esempio, si possono individuare le realizzazioni fonetiche contrassegnate con il diacritico di nasalità [õ] digitandolo nel riquadro di ricerca (come indicato dalla freccia in Fig. 6).

In maniera analoga alla nasalizzazione, si possono ricercare altre caratteristiche come: accento (primario e secondario); durata (crono e semicrono); sollevamento e abbassamento della lingua, avanzamento e arretramento della base della lingua, varie tipologie articolatorie (dentale, alveolare, laminale, ecc.), centralizzazione, arrotondamento, rilascio non udibile, lateralizzazione, palatalizzazione; tipologie di fonazione mista e altre ancora.

Si può inoltre cercare la sequenza all'inizio o alla fine di parola (caselle da spuntare in alto a destra in Fig. 7). Si può restringere la ricerca alle parole riscontrate solo in uno dei punti di inchiesta oppure a quelli appartenenti ad una diocesi particolare (2° e 3° *check-box* in alto in Figg. 6, 7). Si può ottenere la ricerca combinatoria fra vari gruppi fonetici C-V, selezionati dai rami dell'alberatura che compare a sinistra della tastiera virtuale di Fig. 7 (alberatura che fra l'altro mostra le relazioni di parentela fonetica), oppure scrivendo manualmente fra parentesi graffe {} i simboli da combinare nella ricerca. Nell'esempio di Fig. 7, si cercano tutte le combinazioni costituite da un'occlusiva bilabiale, seguita da un vocoide alto e da un contoide nasale (/pim/, /pin/, ecc.), per un totale di 420 (5x6x14) sequenze distinte da individuare.

Per quanto riguarda la trascrizione fonetica dell'Atlante AIS, si decise di attenersi a criteri strettamente filologici e di mantenerla così com'era, senza utilizzare glifi diversi (semplificati), come era stato fatto, per ragioni di leggibilità, nel progetto sui dialetti trentini, e anche senza "tradurla" in IPA, come si sta facendo anche recentemente (ad es., *nell'Atlante Digitale dell'Italia e della Svizzera Meridionale* ADIS; Krefeld, Lücke, 2010[1]).

Si evitava in questo modo di introdurre un secondo livello interpretativo, che si sarebbe aggiunto a quello operato a suo tempo dai raccoglitori dell'AIS (nel caso del Veneto, P. Scheuermeier), con il grave handicap di non disporre più della sorgente sonora originale.

Una controindicazione a questa operazione complessa veniva innanzitutto dal fatto che non esiste una corrispondenza biunivoca fra i simboli dell'inventario fonetico AIS e quello IPA.

La seconda controindicazione era che, inevitabilmente, questo passaggio sfociava nella tentazione di "restaurare" l'AIS, a somiglianza di quanto può avvenire anche in altri campi disciplinari (archeologia, decifrazione di testi antichi ecc.), modificando trascrizioni, giudicate errate, che riportavano invece fedelmente varianti locali significative.

[1] https://www.adis.gwi.uni-muenchen.de/AIS.php.

5. *Scelta di un OCR per l'AIS*

Il compito di riconoscere i caratteri dell'Atlante AIS presenta difficoltà notevoli, da un lato perché sono tracciati a mano, dall'altro per il corsivo e per il numero dei diacritici e dei livelli su cui sono distribuiti, caratteristiche che impediscono l'utilizzo dei normali OCR.

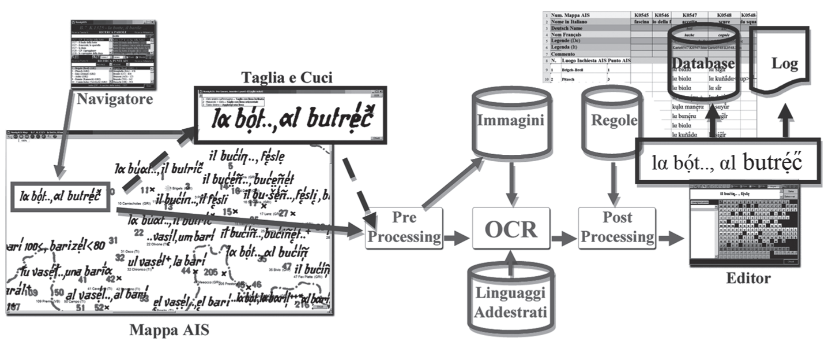
Il riconoscimento dei caratteri a stampa (OCR, *Optical Character Recognition*) e soprattutto dei caratteri scritti a mano (HCR, *Handwritten Character Recognition*) con diacritici ed in corsivo è ancora un campo interdisciplinare aperto ad una grande quantità di ricerche, e allo sviluppo di metodologie e tecnologie innovative (ICR, *Intelligent Character Recognition*, *Machine Learning*, *Pattern Recognition*, *Deep Learning*, ecc.).

Per dare una idea sullo stato dell'arte, si possono citare i risultati di un test, alquanto elementare, come quello di riconoscere una sequenza di 36 caratteri (inglesi, alfa-numerici, maiuscoli, isolati), ripetuta 10 volte, e scritta a mano libera da 7 persone di varia età (Vithlani, Kumbharana, 2015).

La sequenza è stata sottoposta a 6 software OCR sia desktop che online. Il migliore OCR si è dimostrato essere Custom OCR Online, con una percentuale di accuratezza totale del 43.89% di riconoscimento, ovverossia un errore CER (*Character Error Rate*) del 56.11%.

Un risultato così modesto dimostra che siamo ben lontano dalle percentuali vicine al 100% che si ottengono nel caso dei caratteri stampati (quando non siano rovinati o affetti da rumore).

Fig. 10 - *NavigAIS: Schema generale. A sinistra il navigatore. Sulla mappa AIS è indicato con un riquadro la zona dell'immagine passata all'OCR (al centro). A destra, l'Editor per la correzione con la tastiera virtuale e il salvataggio della sequenza riconosciuta*



Da notare che, in genere, le condizioni che si incontrano nel riconoscimento dei caratteri AIS sono peggiori di quelle del test descritto. In generale, si possono considerare le sequenze fonetiche dell'AIS come caratterizzate da un numero molto alto di caratteri a bassa frequenza, con un andamento che può essere descritto dalla legge di Zipf (Zipf, 1949).

La conseguenza inevitabile di questo andamento non può essere altro che il peggioramento dei risultati di un qualsiasi OCR basato su un motore probabilistico. In particolare: *a*) l'atlante AIS non riguarda una lingua sola, ma le lingue di molte regioni con una grande variabilità linguistica. Il numero di simboli da riconoscere aumenta di conseguenza (per es. 230 simboli diversi per il Veneto, ma più di 300 per l'intero AIS) con il conseguente aumento delle probabilità di errore dell'OCR; *b*) il corsivo (circa 70° di inclinazione, vedi Fig. 1-3, 8) dei caratteri AIS abbassa le performance degli odierni OCR a livelli anche inferiori a quelli del testo scritto a mano; *c*) la struttura dei diacritici, articolata su 7 livelli, è al momento irrisolvibile da parte degli attuali OCR (Fig. 9); *d*) manca per l'AIS un modello del linguaggio (o meglio, dei linguaggi), che predica le probabilità della successione di una serie di caratteri e di parole nella(e) lingua(e) da riconoscere, e che permetta all'OCR di scegliere fra varie possibili sequenze quella giusta; *d*) manca per l'AIS un dizionario che elenchi le parole esistenti nei dialetti in questione (forme flesse comprese) e che, come nei correttori ortografici, consenta di individuare la forma corretta fra le varie candidate dall'OCR.

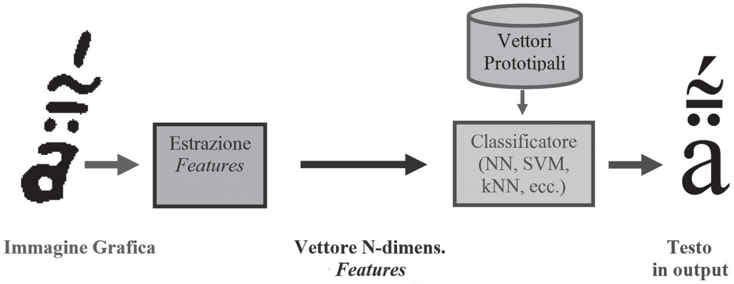
All'inizio della progettazione di NavigAIS furono presi in considerazione tre OCR: i due software commerciali più quotati al momento, Omnipage (Nuance www.nuance.com), e FineReader (ABBYY, www.abbyy.com), ed un software di uso libero Tesseract (sviluppato dalla Hewlett Packard fra il 1985 ed il 1994, diventato open source nel 2005 con licenza Apache 2.0, sponsorizzato da Google per una decina di anni, e ora scaricabile dal sito: https://github.com/tesseract-ocr). Un primo esame rivelò che Omnipage non assicurava una gestione dei diacritici adeguata alle necessità del progetto dell'AIS, e fu scartato.

FineReader permetteva la definizione dei caratteri e dei diacritici AIS secondo la convenzione Unicode. Per quanto riguarda la fase di *training*, il software accettava la ridefinizione delle immagini grafiche dei caratteri problematici con la loro corretta identità, sebbene il meccanismo fosse molto laborioso: in effetti, perché il risultato fosse accettabile, si doveva ripetere l'operazione di identificazione su migliaia di glifi. Da tenere presente anche che questa assegnazione era abbastanza approssimativa, in quanto le immagini catturate da FineReader non potevano essere ripulite dagli spezzoni di linee appartenenti ai caratteri confinanti, con relativi effetti di peggioramento degli errori del riconoscitore.

Dove tuttavia FineReader rivelò la sua inadeguatezza fu nell'impossibilità di integrazione del software nel processo di elaborazione complessivo. In effetti, l'automazione del software consisteva nel depositare da parte dell'utente l'immagine grafica voluta in una cartella Hot Folder, successivamente prelevata e riconosciuta da parte del programma. Purtroppo, l'azione dell'OCR, prevista per funzioni di *Office Automation*, era asincrona e scattava a intervalli temporali discreti con un minimo di 1 minuto. Una simile limitazione comportava tempi di acquisizione inaccettabili per il milione di parole dell'AIS e finiva per rendere la trascrizione manuale più vantaggiosa. A peggiorare le cose si scoprì che il meccanismo di automazione prevedeva solo l'uso di linguaggi standard predefiniti, ed impediva l'uso di un linguaggio addestrato dall'utente, impedendo così l'impiego del software per l'AIS.

Tesseract offriva vantaggi notevoli: 1) metteva a disposizione tutto il codice del programma; 2) poteva essere facilmente integrato nella catena di elaborazione della sequenza fonetica; 3) aveva una percentuale di errore confrontabile con i software commerciali più prestigiosi.

Fig. 11 - *Schema di un generico riconoscitore OCR*



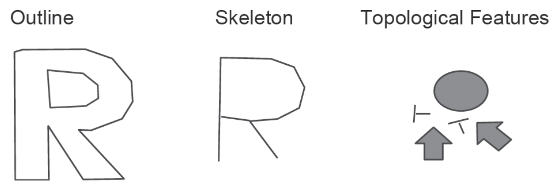
6. *Funzionamento di un OCR*

Il processo di un generico OCR può essere schematizzato in due fasi (Fig. 11):

1. una fase di misurazione, con cui si estraggono dall'immagine grafica in esame un set di misure caratteristiche (*features*) che identificano un glifo;
2. una fase di classificazione con cui si cerca il *matching* fra features ricavati dai caratteri da riconoscere, e prototipi memorizzati nella fase di addestramento (*templates*). Il classificatore deve trovare la minima distanza nello spazio a N dimensioni (con N = numero di features) fra il vettore estratto dal carattere sconosciuto, ed i templates già ricavati dal *training*.

Come è evidente, la scelta del numero di features è la chiave del sistema OCR, poiché porta ad una complessità computazionale ed implementativa, che cresce esponenzialmente con il numero di features. Si tratta di trovare il giusto compromesso fra le esigenze contrastanti della migliore approssimazione possibile e del minor costo computazionale da pagare. I metodi (Fig. 12) più usati sono basati su: 1) il contorno del carattere; 2) la scheletrizzazione del carattere; 3) le proprietà geometriche e topologiche dei caratteri.

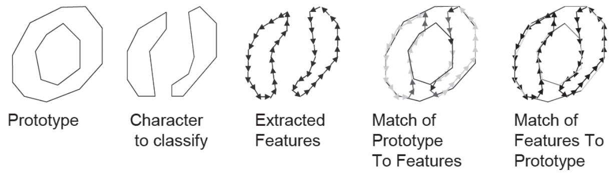
Fig. 12 - *Possibili features da usare nel processo di riconoscimento (Smith, 2014)*



L'approccio che dimostra l'aderenza più stretta ai requisiti ideali di estrazione dei *features* (robustezza, efficienza e peso computazionale) è quello basato su una rappresentazione matematica del contorno dei caratteri.

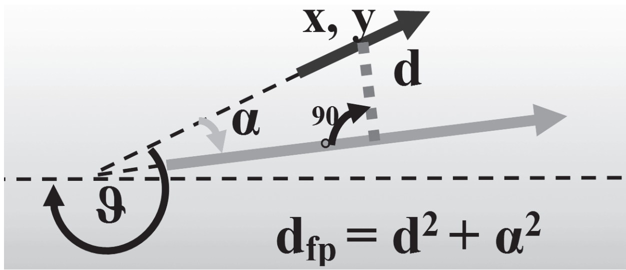
Si evita una descrizione puntuale dell'outline, troppo costosa dal punto di vista computazionale, e si ricorre ad una approssimazione poligonale del contorno del glifo con un numero di lati appropriato, per rendere tutto il processo di riconoscimento sostenibile (Fig. 13).

Fig. 13 - *Approssimazione del contorno di un carattere con una poligonale (linee con le frecce) e matching con features prototipali (linee continue) (Smith, 2014)*



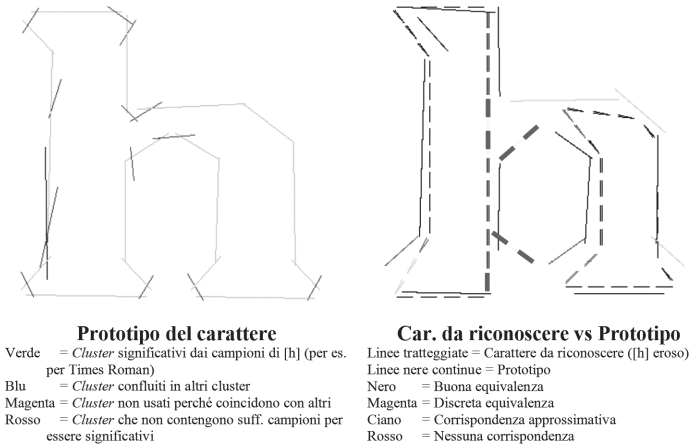
Il contorno è descritto da una serie di vettori 3D (linee in Fig. 14) caratterizzati da tre parametri: la posizione x, y nei vari punti di applicazione del vettore lungo le linee di contorno e l'angolo ϑ che il vettore forma con l'asse orizzontale di base (linea tratteggiata in Fig. 14). Ognuno dei vettori che descrivono il contorno del carattere è confrontato con i vettori prototipali (linea più lunga in Fig. 14) ricavati a priori da un set opportuno di campioni di caratteri della lingua in questione, e memorizzati in un database. La distanza d_{fp} da calcolare per ognuno dei vettori dei modelli di riferimento è data dalla somma della distanza euclidea d (linea tratteggiata in Fig. 14) (perpendicolare al vettore prototipale indicato dalla linea più lunga in Fig. 14) al quadrato e dall'angolo α fra le loro proiezioni al quadrato.

Fig. 14 - *Vettore 3D della poligonale del contorno del carattere, posizionato nel punto x, y e formante l'angolo θ con l'asse orizzontale, e distanza d da un feature (Smith, 2014)*



La minima fra le distanze d_{fp} dai prototipi etichettati fornisce l'identità del candidato più probabile del riconoscimento (Fig. 14). Da notare che l'approssimazione poligonale è un'arma a doppio taglio, che ha il vantaggio di eliminare la dentellatura rumorosa irrilevante, ma che ha il difetto di sopprimere anche informazioni potenzialmente rilevanti.

Fig. 15 - *Estrazione dei prototipi e matching con i features del carattere incognito (Smith et al., 2009)*



7. *La realizzazione del sistema Navigais con OCR integrato*

La soluzione adottata per il progetto AIS è stato un OCR che fosse: *a*) basato sull'OCR Tesseract; *b*) addestrato specificatamente per il riconoscimento dei simboli dell'Atlante AIS; *c*) integrato con il sistema di navigazione e di interrogazione delle mappe e dei punti di inchiesta di NavigAIS (Fig. 10); *d*) seguito da una fase di post-processing per correggere gli errori eventuali (Fig. 17); *e*) provvisto di una tastiera virtuale (completamente riconfigurabile secondo le esigenze dell'utilizzatore) per l'editing della sequenza restituita da Tesseract e dal post-processing (Fig. 18).

Lo schema a blocchi del sistema realizzato è mostrato in Fig. 10: a sinistra in alto compare la finestra di NavigAIS, che chiameremo Navigatore e che permette di individuare le mappe contenenti un certo argomento nell'indice delle tavole AIS, di caricare la mappa voluta (in basso a sinistra) e di scegliere eventualmente il punto (o i punti) su cui operare il riconoscimento.

Al centro della Fig. 10 figurano i blocchi di elaborazione dell'OCR e di post-elaborazione, ed a destra la finestra dell'*Editor* che serve alla correzione della sequenza fonetica e al salvataggio nel database.

Fig. 16 - *L'OCR opera separando i caratteri dai diacritici. La fase di post-processing si avvale di un set di regole ad hoc, per sopperire alla mancanza di un dizionario ed un modello del linguaggio*

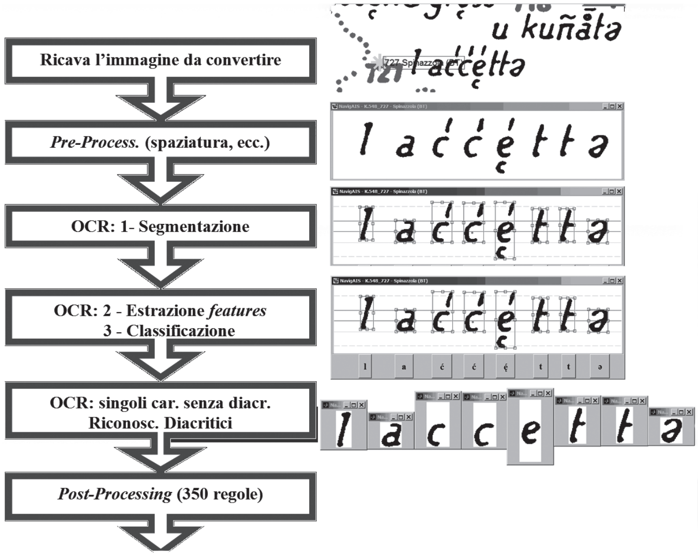
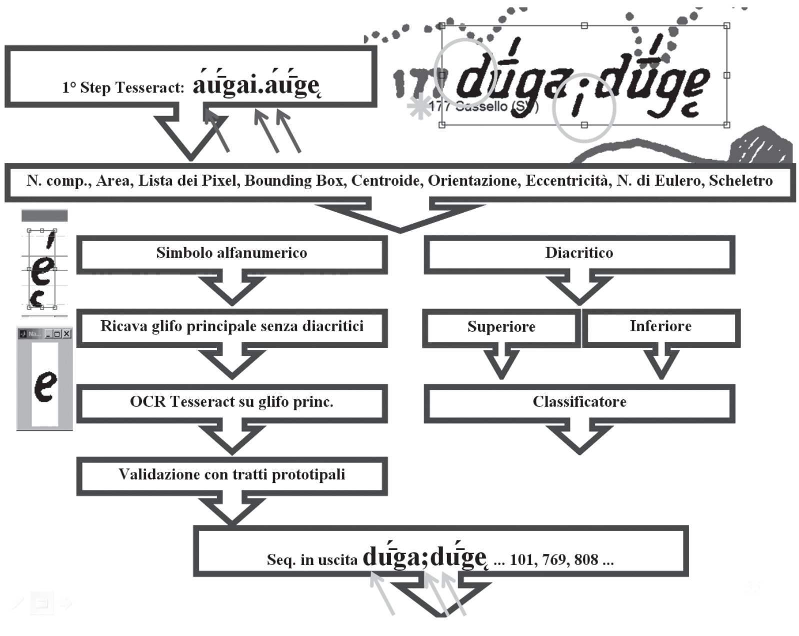


Fig. 17 - *Fase di post-processing. I 12 errori (caratteri + diacritici) commessi dall'OCR Tesseract (in alto a sinistra) sono stati automaticamente corretti dalla fase di post-elaborazione*



Nel processo di acquisizione con l'OCR, il supervisore ha a sua disposizione tre modalità di movimento da un punto all'altro: 1) Il movimento può procedere automaticamente su tutti i punti dell'atlante in sequenza, oppure esclusivamente sui punti appartenenti ad una regione o ad una provincia, oppure infine su una lista di punti selezionati a priori dagli appositi box nell'interfaccia di navigazione, oppure dall'apposito parametro del file di configurazione; 2) il software provvede automaticamente, senza perdita di tempo e senza errori, a spostarsi sui punti di acquisizione voluti secondo un ordine dato; 3) il supervisore svolge due compiti: sceglie quale delle sequenze che circondano il punto sia candidata all'OCR, e controlla nello stesso tempo la correttezza della stringa ritornata dall'OCR.

In alternativa: *a*) l'operatore può selezionare manualmente un punto qualsiasi sulla mappa, per cui un successivo salvataggio avviene nella località che era già selezionata in quel momento; oppure *b*) l'operatore può selezionare un punto cliccando sul nome delle località d'inchiesta, per cui un salvataggio avviene forzatamente all'indirizzo corrispondente.

Dopo la selezione del lemma da acquisire, il programma traccia il riquadro (rettangolo in Fig. 10) con l'area della mappa, in cui i caratteri non siano più distanti di una certa quantità prefissata. Qualsiasi linea tocchi il riquadro esterno è completamente rimossa nel processo di pre-elaborazione, per mantenere solo la sequenza utile al riconoscimento. In questa fase, prima di passare all'elaborazione, l'immagine grafica è salvata su disco, a disposizione per la successiva fase di riconoscimento, ma anche per una eventuale fase di apprendimento dell'OCR.

L'immagine grafica subisce varie fasi opzionali di trattamento, per ripulirla automaticamente dalle porzioni marginali dei glifi, appartenenti ad altre sequenze di caratteri vicini, e per ripararla da piccoli difetti. Se necessario, l'operatore può anche intervenire manualmente, per operare tagli o per "cucire" i rami dei glifi rovinati, nella finestra "Taglia e Cuci" (Fig. 10).

Fig. 18 - *Finestra di lancio dell'OCR e tastiera virtuale per l'editing delle sequenze fonetiche*

Alla fine del riconoscimento, Tesseract restituisce la sequenza fonetica, assieme ai confini dei rettangoli che delimitano ogni singolo carattere (*Bounding Box*), e che possono servire eventualmente a fasi ulteriori di addestramento dell'OCR.

8. *Post-Processing*

Per migliorare la performance, gli OCR comuni ricorrono ad una fase di post-processing che corregge, quando possibile, gli errori dell'OCR, consultando un dizionario ed un modello del linguaggio. Queste componenti possono contribuire anche in maniera essenziale alla fase di segmentazione di una sequenza di caratteri.

Ognuna delle possibili varianti delle parole in uscita dall'OCR è confrontata con il dizionario ed eliminata, se non trova una validazione (ad es. [monte], che esiste, rispetto a [rnonte], che non esiste). Purtroppo, non esiste un dizionario delle parole dialettali dell'Atlante AIS che sia disponibile in un formato accettabile, per poter attribuire una sequenza di caratteri ad una forma certa della lingua in questione. Nel 1960, fu pubblicato postumo (Jaberg era morto nel 1958 e Jud nel 1952) un volume con l'indice delle parole trascritte in maniera tipizzata (ad esempio, "abbághie" invece di [abbáǧǧə]), che non può servire allo scopo che si era prefissato il progetto NavigAIS.

Non esiste neppure un modello del linguaggio (*Statistical Language Model* - SLM), che dia le possibili frequenze di digrammi, trigrammi, ecc., e che risulterebbe particolarmente utile nel riconoscimento della sequenza di caratteri.

Per abbassare la percentuale di errori, si è adottata una specifica strategia per l'Atlante AIS, che ricorre sostanzialmente a due fasi distinte e successive di riconoscimento con Tesseract (Tab. 1):

- Si demanda alla prima fase di riconoscimento svolta da Tesseract il lavoro grossolano di una prima identificazione del carattere, che deve fornire i parametri dei *Bounding Box* indispensabili per le operazioni successive.
- Ottenuta la prima classificazione del carattere, la fase di post-processing provvede a separare il flusso dell'elaborazione in due percorsi diversi (Fig. 17), in cui il primo deve riprocessare e validare la componente di base del glifo da solo con una nuova sottomissione all'OCR Tesseract, mentre il secondo è specializzato nell'elaborare con algoritmi Matlab il gruppo di tutti i diacritici a sé stanti.
- Ogni carattere di base senza diacritici è assoggettato a questo secondo passaggio solitario attraverso Tesseract, con l'uscita ristretta unicamente ai segni alfanumerici senza diacritici, per ridurre contemporaneamente sia il numero dei glifi a soli 91 simboli, sia le possibili ambiguità fra caratteri, e di conseguenza il CER complessivo.
- I diacritici sono separati dal corpo principale del glifo, sono divisi fra diacritici in apice e a pedice, e classificati con regole basate sulle loro caratteristiche topologiche, sulla loro compatibilità reciproca ed anche sulla loro compatibilità con il carattere di base. In questo modo la complessità della casistica infinita delle forme (nel caso delle vocali circa 8480) e dei livelli dei diacritici, si riduce a pochi casi semplici (simboli alfanumerici + diacritici) in numero inferiore a 150.

Ovviamente gli errori sono ancora possibili per i glifi scritti male, rovinati, fusi con i diacritici, e anche per forme male addestrate, ecc.

– Infine, l'ultimo stadio della post-elaborazione applica circa 350 regole di controllo dell'identità dei caratteri, basate sulla topologia, le dimensioni e la conformità fonetica delle componenti dei glifi. In questa fase, per aumentare la velocità di elaborazione, i simboli sono classificati ed elaborati in 5 categorie separate: 1 – Vocali, 2 – Consonanti, 3 – Cifre, 4 – Punteggiatura, 5 – Segni diacritici. Se ad esempio l'OCR restituisce una [i], ma il numero delle componenti reali, ricavato dai pixel dell'immagine, è invece di una sola componente, il carattere, a seconda delle dimensioni e della posizione del baricentro, può essere una [l] oppure una [,]. Se invece le componenti sono due, ma il baricentro è basso (sotto la media risultante dei caratteri AIS) allora si tratta di un [;]. Altrettanto dicasi ad es. di una [é] accentata, che abbia un'unica componente, con la conseguente eliminazione dell'accento.
  Se l'OCR ritorna un [.], ma la componente ha un buco (numero di Eulero = 0), allora (a seconda delle dimensioni e della posizione) può trattarsi di un diacritico [o̊], oppure [o̥], oppure del carattere [o], oppure di [o] in apice [°], ecc. Un altro tipo di controllo interessa anche l'associazione dei diacritici e dei glifi di base, che risulti accettabile per l'inventario fonetico dell'AIS.

Si deve notare che queste regole non sono per nulla definitive e sono state, e possono essere ancora, modificate, raffinate, ed estese in continuazione, sulla base dell'esperienza e di altre conoscenze rilevanti acquisite nel corso del processo di riconoscimento.

Tab. 1 - *Schema delle regole applicate nella fase di post-processing*

Tipologia	**Scopo dell'elaborazione**
N. componenti	Per permettere, ad es., di discriminare: [i] da [l], [;] da [,].
Area	Per distinguere fra oggetti irrilevanti, diacritici e caratteri.
Lista dei Pixel	Per individuare le componenti connesse.
Bounding Box	Per confronti differenziali basati sulle dimensioni.
Estremi	Per distinguere, ad es., le parentesi.
Centroide	Per la dilatazione dei caratteri, linee di riferimento.
Orientazione	Per distinguere, ad es., accento primario [ó] da secondario [ò].
Eccentricità	Per discriminare componenti lineari (linea=1) da ellittici (cerchio=0): ad es. [ō], [õ], rispetto a [ȯ].
N. di Eulero	Per distinzione basata sul numero di "fori" del componente: ad es. [e] da [œ], [g] da [y], [ọ] da [o̥], ecc.
Angoli	Per discriminare alcuni diacritici.
Scheletro	Per distinguere diacritici e caratteri: ad es. [u] da [y], [k] da [h].

9. *Valutazione dei risultati dell'OCR*

Per la valutazione della prestazione del riconoscitore, sono stati realizzati vari test che hanno riguardato sia le tappe del *training* dell'OCR, sia la fase di riconoscimento vero e proprio. Per la fase di addestramento, è stato preparato un test minimale con 50 lemmi (circa 500 caratteri complessivamente) provenienti dal database AMDV.

Il test è stato volutamente limitato, per testare rapidamente l'effetto delle modifiche che si sono succedute nel tempo, e per confrontare fra di loro l'accuratezza dei vari linguaggi addestrati e di tutti i linguaggi fra di loro. La Fig. 19 riassume i risultati dell'OCR con linguaggi addestrati su diversi set di *training*, presi a sé stanti oppure in combinazione fra di loro. Le tre colonne verticali riportano il CER (*Character Error Rate*) per le tre modalità di test principali (Fig. 19-20): 1) OCR Tesseract da solo senza dilatazione dei caratteri (a destra); 2) OCR Tesseract da solo con dilatazione dei caratteri (al centro); 3) OCR con post-processing (a sinistra).

Fig. 19 - *Risultati comparativi di vari test fatti con i linguaggi addestrati su diversi set di training. Le tre colonne riportano il CER per le tre modalità di test principali: OCR Tesseract da solo senza dilatazione dei caratteri (a destra), OCR Tesseract da solo con dilatazione dei caratteri (al centro), OCR con post-processing (a sinistra)*

1	NavigAIS OCR Test											
2	Char Dilatation + Post_Processing				Char Dilatation				OCR only			
3	OCR Language	CER	WER	WER (	OCR Language	CER	WER	WER (	OCR Language	CER	WER	WER (
4												
5	ais01+ais12+ais09+ais11_dilat_post	1,41	6,6	6,6	ais01+ais12+ais09+ais11_dilat	22,05	73,58	70,75	ais01+ais12+ais09+ais11_no_dil	24,51	71,7	66,98
6	ais01+ais12+ais09_dilat_post	1,41	6,6	6,6	ais01+ais12+ais09_dilat	22,05	73,58	69,81	ais01+ais12+ais09_no_dil	25,04	74,53	68,87
7	ais01+ais12+ais10+ais11_dilat_post	1,41	6,6	6,6	ais01+ais12+ais10+ais11_dilat	22,57	84,91	81,13	ais01+ais12+ais10+ais11_no_dil	26,46	79,25	73,58
8	ais01+ais12+ais10_dilat_post	1,41	6,6	6,6	ais01+ais12+ais10_dilat	22,93	83,96	81,13	ais01+ais12+ais10_no_dil	26,63	80,19	74,53
9	ais01+ais09+ais11+ais12_dilat_post	1,41	6,6	6,6	ais01+ais09+ais11+ais12_dilat	23,1	76,42	72,64	ais01+ais09+ais11+ais12_no_dil	25,4	74,53	69,81
10	ais01+ais09+ais12+ais11_dilat_post	1,41	6,6	6,6	ais01+ais09+ais12+ais11_dilat	23,1	76,42	72,64	ais01+ais09+ais12+ais11_no_dil	25,4	74,53	69,81
11	ais12+ais01+ais10+ais11_dilat_post	1,41	6,6	6,6	ais12+ais01+ais10+ais11_dilat	23,1	88,68	83,96	ais12+ais01+ais10+ais11_no_dil	27,16	83,02	77,36
12	ais12+ais01+ais09+ais11_dilat_post	1,41	6,6	6,6	ais12+ais01+ais09+ais11_dilat	23,81	81,13	77,36	ais12+ais01+ais09+ais11_no_dil	26,81	76,42	71,7
13	ais01+ais12+ais09+ais10_dilat_post	1,41	6,6	6,6	ais01+ais12+ais09+ais10_dilat	24,34	76,42	74,53	ais01+ais12+ais09+ais10_no_dil	26,63	75,47	70,75
14	ais01+ais12+ais10+ais09_dilat_post	1,41	6,6	6,6	ais01+ais12+ais10+ais09_dilat	24,69	78,3	76,42	ais01+ais12+ais10+ais09_no_dil	26,98	76,42	72,64
138	ais09+ais10+ais11+ais01_dilat_post	2,12	10,38	9,43	ais09+ais10+ais11+ais01_dilat	31,75	85,85	81,13	ais09+ais10+ais11+ais01_no_dil	32,45	86,79	82,08
139	ais01+ais12_dilat_post	2,47	9,43	7,55	ais01+ais12_dilat	19,58	77,36	74,53	ais01+ais12_no_dil	22,22	75,47	71,7
140	ais11+ais10_dilat_post	2,47	12,26	12,26	ais11+ais10_dilat	29,63	83,02	80,19	ais11+ais10_no_dil	29,1	83,96	78,3
141	ais01+ais12+ais11_dilat_post	2,65	10,38	7,55	ais01+ais12+ais11_dilat	19,22	76,42	75,47	ais01+ais12+ais11_no_dil	21,52	72,64	69,81
142	ais12+ais11_dilat_post	2,65	12,26	9,43	ais12+ais11_dilat	21,52	88,68	85,85	ais12+ais11_no_dil	23,99	84,91	83,02
143	ais11+ais09_dilat_post	2,82	13,21	13,21	ais11+ais09_dilat	28,22	83,02	79,25	ais11+ais09_no_dil	29,1	83,02	78,3
144	ais10+ais09_dilat_post	2,82	14,15	13,21	ais10+ais09_dilat	32,63	86,79	82,08	ais10+ais09_no_dil	32,98	85,85	82,08
145	ais12_dilat_post	3,17	15,09	13,21	ais12_dilat	24,69	99,06	96,23	ais12_no_dil	25,4	92,45	88,68
146	ais10_dilat_post	3,17	16,04	16,04	ais10_dilat	31,22	89,62	83,96	ais10_no_dil	31,57	90,57	84,91
147	ais09_dilat_post	3,88	18,87	17,92	ais09_dilat	34,04	92,45	86,79	ais09_no_dil	34,39	89,62	85,85
148	ais11_dilat_post	6,7	29,25	26,42	ais11_dilat	27,87	85,85	82,08	ais11_no_dil	29,63	85,85	83,02
149	ais01+ais11_dilat_post	7,23	24,53	21,7	ais01+ais11_dilat	20,63	63,21	61,32	ais01+ais11_no_dil	22,57	67,92	64,15
150	ais01_dilat_post	11,82	44,34	39,62	ais01_dilat	21,52	68,87	63,21	ais01_no_dil	24,69	74,53	67,92
151	ais_f_dilat_post	25,93	93,4	89,62	ais_f_dilat	40,56	116,04	109,43	ais_f_no_dil	40,56	108,49	102,83
152												
153	Media errore con post-processing	1,735643			Media con dilat. caratteri	26,47582			Media solo OCR	28,12795		

Il risultato migliore (errore dell'1.41%) è ottenuto dalla combinazione di vari linguaggi (fra cui ais01, ais09, ais10, ais11, ais12) e dall'esecuzione del post-processing, mentre il peggiore, CER = 25.93%, è dato dal linguaggio allenato sul set costruito artificialmente con caratteri di un font che doveva imitare l'AIS, ma che evidentemente è abbastanza lontano dalla realtà grafica delle tavole AIS.

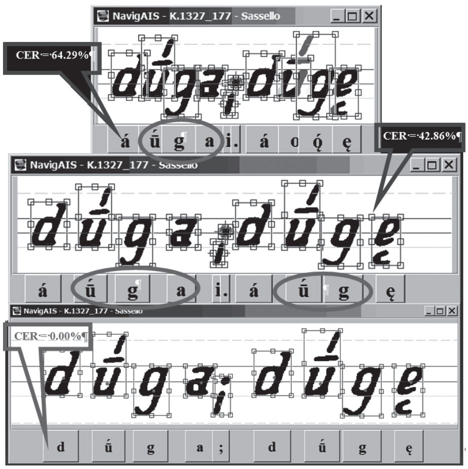
Senza la fase di post-processing, ma con la dilatazione dei caratteri, il CER aumenta e varia dal 22% al 40%, mentre senza la fase di post-processing, e senza la dilatazione dei caratteri, il CER aumenta dal 24% al 40%, con un peggioramento del 2-3% rispetto al caso della dilatazione dei caratteri. La Fig. 19 riporta anche il WER (*Word Error Rate*), l'errore sulle parole intere, che non ha molto interesse per l'AIS.

Il tempo di acquisizione per punto è risultato inferiore ai 14" per sequenze di circa 8.36 caratteri, più che dimezzato rispetto ai 30", ipotizzati inizialmente nel progetto. Si deve ovviamente tenere conto della lunghezza della sequenza, per cui all'aumentare del numero di caratteri il tempo di acquisizione aumenta di qualche secondo nelle frasi più lunghe.

Per quanto riguarda il tasso di errore, invece, questo non dipende dalla lunghezza della sequenza riconosciuta dall'OCR e rimane invariato. L'errore può aumentare di qualche punto percentuale nelle sequenze "anomale", in cui i caratteri siano appiccicati e nel caso il deterioramento dei caratteri sia alto. In questi casi, in effetti, la mancanza di un dizionario dei lemmi dialettali impedisce all'OCR di interpretare correttamente i caratteri lesionati in una certa sequenza.

Un test di acquisizione su dati reali provenienti da 14 tavole complete per 100.000 caratteri complessivi ha dato un CER medio del 3.65%, includendo nel calcolo una qualsiasi inserzione, omissione e sostituzione dei caratteri, diacritici compresi (con la cosiddetta *Distanza di Levenshtein*; Levenshtein, 1965).

Fig. 20 - *Risultati OCR: Senza distanziamento dei caratteri, l'errore CER commesso da Tesseract è del 64.29%. Con il distanziamento, ma senza post-processing sul risultato dell'OCR, l'errore CER si riduce al 42.86%. Infine con distanziamento e post-elaborazione, tutti i caratteri sono riconosciuti correttamente con CER = 0% (NavigAIS-web K. 1327.177)*



10. *Conclusioni*

Basandosi sulla percentuale di errori CER (*Character Error Rate*), ottenuta nei test e sui dati reali di 14 mappe per circa 100.000 caratteri (CER medio inferiore al 3.65%, molto minore di quello ipotizzato inizialmente), e sui tempi di acquisizione per punto (inferiore ai 14" per sequenze con una media di 8.36 simboli, dimezzato rispetto alle stime iniziali), si può ragionevolmente predire che il completamento dell'acquisizione dell'intero AIS avverrà in meno di 2 anni/uomo. Queste stime sono confermate dai risultati già ottenuti, per cui alla fine del 2019 il 50% delle mappe AIS è già stato completamente trascritto.

Bibliografia

Goebl, H. (1994). *L'Atlas linguistique du ladin central et des dialectes limitrophes (première partie, ALD-I).* In Mouton, P.G. (Ed.), *Geolingüística. Trabajos europeos*. Madrid: Consejo Superior de Investigaciones Científicas, 155-168.

Jaberg, K., Jud, J. (1928). *Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweitz*. Halle: Max Niemeyer. [Traduzione italiana a cura di G. Sanga: *Ais - Atlante Linguistico ed Etnografico dell'Italia e della Svizzera Meridionale. Vol. 1: Fondamenti Critici e Introduzione, Vol. 2: Scelta di Carte Commentate*. Milano: Unicopli, 1988.]

Jaberg, K., Jud, J. (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz, Vol. 1-8*. Ringier, Zofingen, Bern: Max Niemeyer. [Ristampa 1971-1981: Nendeln/New York: Kraus Reprint]. www3.pd.istc.cnr.it/navigais-web.

Jaberg, K. (1960). *Index zum Sprach- und Sachatlas Italiens und der Südschweiz: Ein propädeutisches etymologisches Wöterbuch der italienischen Mundarten*. Berna: Stämpfli.

Kattenbusch, D. (1995). *Atlas parlant de l'Italie par régions: VIVALDI*. In Aa. Vv. *Estudis de lingüística i filologia oferts a Antoni M. Badia i Margarit*. Barcellona: Universitat Autònoma de Barcelona, Departamento de Filología Catalana y de Filología Española, 443-455. www.dailybest.it/web/vivaldi-mappa-interattiva-dialetti-italiani.

Kramer, J. (1988-1998). *Etymologisches Wörterbuch des Dolomitenladinischen (EWD), voll. 1-8*. Amburgo: H. Buske Verlag.

Loporcaro, M. (2014). *AIS Reloaded*, https://www.rose.uzh.ch/de/forschung/forschungamrose/projekte/AIS-reloaded.html.

Mott, A., Kezich, G. & Tisato, G. (2003). *Il Trentino dei contadini. Piccolo atlante sonoro della cultura materiale.* San Michele all'Adige (TN): Museo degli Usi e Costumi della Gente Trentina. www.museosanmichele.it/editoria/shop/opere-varie-volumi-esauriti/g-kezich-a-mott-g-tisato-il-trentino-dei-contadini-cdrom.

Scheuermeier, P. (1943). *Bauernwerk in Italien, der italienischen und rätoromanischen Schweiz, Vol. 1-2*. Erlenbach/Zurigo: Rentsch. [Traduzione italiana: Scheuermeier P. (1980). *Il lavoro dei contadini*. Milano: Longanesi.]

Telmon, T., Canobbio, S. (1985). *Atlante linguistico ed Etnolinguistico del Piemonte Occidentale.* Torino: Regione Piemonte.

Tisato, G. (2010). NavigAIS – AIS Digital Atlas and Navigation Software. In *Atti del VI Convegno AISV*, Napoli, 3-5 febbraio 2010, 451-461 (versione scaricabile: www3.pd.istc.cnr.it/navigais, versione online: www3.pd.istc.cnr.it/navigais-web).

Tisato, G., Vigolo, M.T. (2011). *Atlante Multimediale dei Dialetti Veneti (AMDV)*. In Borgato, G., Vanelli, L. (Eds.), *Atti del Convegno "In ricordo di Alberto Zamboni"*. Padova, 25, gennaio, 2011, 99-126.

Tisato, G., Barbierato, P., Ferrieri, G., Gentili, C. & Vigolo, M.T. (2013). Atlante Multimediale dei Dialetti Veneti. In *Atti del IX Convegno AISV 2013, Multimodalità e Multilingualità*, Università Ca' Foscari, Venezia, 21-21 gennaio 2013. Roma: Bulzoni, 445-462.

Tisato, G., Vigolo, M.T. (2016). Dagli Atlanti storici agli Atlanti multimediali: il NavigAIS e l'AMDV (Atlante Multimediale dei Dialetti Veneti). In *Atti del Convegno*

Internazionale di Studi Archivi Etnolinguistici Multimediali, Museo delle Genti d'Abruzzo, Pescara, 6 ottobre 2012, 96-123.

Vithlani, P., Kumharana, C.K. (2015). Structural and Statistical Feature Extraction Methods for Character and Digit Recognition. In *International Journal of Computer Applications*, 120, 43-47.

Zamboni, A. (1974). *I dialetti veneti*. Pisa: Pacini.

Zamboni, A. (1984). I dialetti cadorini. In Pellegrini, G.B., Sacco, S. (Eds.), *Atti del Convegno Internazionale "Il ladino bellunese"*. Belluno, 2-4 giugno 1983, 45-83.

Zamboni, A., Vigolo, M.T. (2011). Tra nomi e cose. Commenti lessicali e onomasiologici allo Scheuermeier veneto. In Perco, D., Sanga, G. & Vigolo, M.T. (Eds.), Scheuermeier, P., *Il Veneto dei contadini 1921-1932*. Vicenza: Angelo Colla Editore, 67-87.

Zipf, G.K. (1949). *Human Behaviour and the Principle of Least-Effort*. Oxford: Addison-Wesley Press.

Ringraziamenti

Un ringraziamento a tutto il team del progetto *AIS Reloaded*, in particolare a Chiara Zanini, Giulia Donzelli e Stefano Negrinelli, che stanno lavorando alla raccolta dei dati nella Svizzera Romanza ed hanno collaborato (assieme a Giacomo Ferrieri) alla preparazione dell'inventario fonetico AIS.

DOI: 10.17469/O2106AISV000009

PETRA VANGELISTI, CATERINA PESCE, MARICA SETARO,
GRETA BIANCHINI, LUCILLA GIGLI, SILVIA CALAMAI

Ritrovare Voci: il lavoro intorno all'archivio di Anna Maria Bruzzone[1]

A Paola Chiama,
che ha reso questo possibile

The Anna Maria Bruzzone archive, donated in 2017 to the University of Siena, contains a vast range of materials: medical records, iconographic materials (photos, drawings and portraits), tape and video recordings, direct testimonies of the last years of activity of the mental hospital of Arezzo. Such an archive paves the way for different lines of research, two of which are outlined in this article. The first one is a linguistic analysis of the relationship between an original audio interview and its rendering in the Bruzzone's book *Ci chiamavano matti*. The second one consists in tracing some of the hospital patients interviewed in 1977 by Bruzzone in order to collect a new testimony from them, starting from the listening of their old interviews.

Key words: audio archives, oral history, psychiatric hospital.

1. *Non solo carta*

Il contributo nasce dalle storie e dalle esperienze che hanno scandito l'internamento manicomiale nell'ex Ospedale Neuropsichiatrico di Arezzo, a partire dagli inizi del '900 fino alla sua dismissione, avvenuta sotto la direzione di Agostino Pirella, uno dei riferimenti fondamentali del movimento nazionale che portò all'approvazione della cosiddetta *legge Basaglia* (n. 180 del 1978).

Una consolidata abitudine – legata ad un'iperproduzione di natura scritta – fa pensare agli archivi come a un luogo fisico pieno di carte e anche gli archivi degli ex ospedali psichiatrici non si sottraggono a questa descrizione. Sarebbe tuttavia limitativo ridurre i precipitati di vita di queste istituzioni alle sole fonti scritte. Una storia lunga un secolo è fatta di altri materiali: alle storie cliniche degli internati aretini registrate nelle cartelle si affiancano il materiale iconografico (foto, disegni, ritratti) e le registrazioni audio, i filmati, le testimonianze dirette rese da chi ha

[1] Questo contributo fa parte di un progetto più ampio, in un gruppo di ricerca composito costituito da linguisti, storici orali, storici della scienza, archivisti. Le voci di una giovane laureata in Linguistica Generale (Petra Vangelisti, autore del secondo paragrafo), di una dottoranda in Storia Contemporanea (Caterina Pesce, autore del terzo paragrafo) e di un'assegnista di ricerca in Storia della scienza (Marica Setaro, autore del primo paragrafo) intendono mostrare le molteplici linee di ricerca aperte dal ritrovamento dell'archivio sonoro di Anna Maria Bruzzone.

ricordato o ricorda gli ultimi anni di funzionamento del manicomio prima della totale dismissione. Con la direzione di Agostino Pirella (1971-1979) il manicomio si apre all'esterno e giungono intellettuali, artisti, giornalisti, fotografi, volontari. Nell'estate del 1977 ad Arezzo arriva, con un mangianastri, anche Anna Maria Bruzzone. Insegnante di lettere in un istituto magistrale torinese, sin da giovanissima indossa gli abiti della storica fuori dal circuito accademico. L'incontro con l'universo manicomiale in cui decide di entrare è precoce. Ancora specializzanda in psicologia, nel 1968 decide di conoscere l'esperienza di comunità terapeutica che Franco Basaglia e la sua équipe stanno sperimentando nel Manicomio di Gorizia. Sono mesi intensi in cui Bruzzone misura su sé stessa le strategie di relazione con i degenti di quel manicomio e cerca di elaborare una metodologia della ricerca. Osserva, appunta tutto nei suoi taccuini, redige questionari a schema fisso tramite i quali ottiene le prime interviste con i pazienti. Gorizia rappresenta una sorta di "battesimo del fuoco" per Bruzzone, consentendole solo nove anni più tardi di ripetere l'esperienza nel manicomio aretino. *Ci chiamavano matti. Voci da un ospedale psichiatrico*, pubblicato da Einaudi nel 1979, è l'espressione compiuta del metodo di ricerca di Anna Maria Bruzzone e di una stagione culturale che aveva avuto come esito l'approvazione della legge 180 del 1978.

Nell'arco di un decennio, la storica torinese dà corpo e voce ai soggetti che spesso la storia ha silenziato e si pone il problema di affinare un metodo. Se a Gorizia per Bruzzone è ancora la testimonianza scritta a ritradurre il parlato, ad Arezzo – già forte del lavoro di testimonianza raccolto dalle ex partigiane piemontesi[2] – si produce uno scarto ulteriore. Il registratore consente di "catturare" le voci, di riprodurle e trascriverle fedelmente, di conservarle su supporti che inaugurano una nuova compagine delle fonti storiche: le fonti sonore, appunto.

Bruzzone non si limita a un lavoro di racconto e di denuncia delle condizioni manicomiali, ma dà voce ai ricoverati: la studiosa, infatti, registra e poi trascrive, relegando il proprio intervento alle note a piè di pagina e a un'introduzione iniziale; niente interrompe la storia che ciascun paziente vuole raccontare di sé. I nastri, dati per dispersi, sono stati ritrovati nel 2016 a Torino, digitalizzati grazie al sostegno economico della Soprintendenza Archivistica e bibliografica della Toscana e donati all'Ateneo di Siena dalla nipote della studiosa, Paola Chiama (Calamai, Biliotti, 2017), insieme ad altra documentazione scritta riferibile al volume.

2. *Il punto di vista della linguistica: affinità e divergenze tra sonoro e testo a stampa*

Nell'archivio di Anna Maria Bruzzone, oltre alle audiocassette su cui sono incise le interviste alla base della stesura del libro *Ci chiamavano matti*, compaiono i quaderni in cui è contenuta una prima trascrizione manoscritta delle interviste e due versioni dattiloscritte di quest'ultime. Avendo non solo la possibilità di accedere

[2] Bruzzone, A.M., Farina, R. (1976). *La resistenza taciuta: dodici vite di partigiane piemontesi*. Bollati Boringhieri.

alle testimonianze integrali dei pazienti così come le avevano pronunciate, ma anche quella di osservare l'intero iter del lavoro di ricerca della studiosa e cercando di capire quali procedimenti abbiano avuto luogo durante la stesura di questa raccolta di interviste, è stato possibile svolgere uno studio di caso sull'intervista al paziente rinominato nel testo come Flavio (Bruzzone, 1979: 80-90).

Obiettivo dell'indagine è verificare come l'oralità viene rappresentata su carta, attraverso l'analisi del documento sonoro originale, delle versioni intermedie (il manoscritto e due versioni dattiloscritte), fino alla versione edita del testo[3].

Per la trascrizione si è fatto riferimento alla proposta metodologica per l'edizione dei documenti orali di Luciano Giannelli e Valeria di Piazza (Giannelli, Di Piazza, 1995). Come suggerito nell'articolo appena citato, si è proceduto per fasi distinte, svolgendo operazioni diverse. Nello specifico:
1. Ascolto generico di tutto il materiale audio dell'intervista.
2. Prima trascrizione attraverso blocchi di audio più brevi, tramite l'uso di oTranscribe (http://otranscribe.com)[4], nella quale si è cercato di carpire il più possibile dal punto di vista contenutistico.
3. Secondo ascolto dei blocchi selezionati di volta in volta a velocità normale per verificare la corretta trascrizione di segmenti che avrebbero potuto essere parzialmente distorti dal rallentamento dell'audio; allo stesso tempo è stata introdotta la punteggiatura.
4. Ascolto integrale dell'intervista seguita tramite la trascrizione compiuta, per controllare che tutti gli elementi orali fossero stati rappresentati in modo sufficientemente soddisfacente nello scritto.
5. Controllo finale.

Poiché fra gli scopi c'era quello di rendere il parlato in un *medium* scritto in maniera quanto più fedele possibile alla fonte, l'utilizzo della punteggiatura è stato cruciale. Si è cercato infatti di trovare un compromesso tra la separazione dei gruppi intonativi compiuta dal parlante e l'utilizzo sintattico dell'interpunzione (Giannelli, Di Piazza, 1995: 54). In particolare, si sono riprodotte le pause compiute dall'intervistato nel corso dell'interlocuzione tramite la punteggiatura, seguendo la concezione invalsa secondo cui quest'ultima è accostata alla positura di pause più o meno lunghe, a seconda che si tratti di virgole, punti o tre puntini. Questa comune interpretazione dell'utilizzo della punteggiatura, per quanto non sia del tutto corretta, viene in ausilio al tentativo di resa del parlato nella pagina scritta. Tuttavia, si è ritenuto che un utilizzo sovrabbondante di segni grafici o grafie diverse per tentare di rappresentare pause, interruzioni o curve intonative, avrebbe potuto rendere la

[3] L'intervista di Flavio si colloca nel capitolo del libro intitolato "L'istituzionalizzazione precoce" (Bruzzone, 1979: 49-90), che raccoglie le testimonianze di figli di ignoti, cresciuti in brefotrofi e tra un istituto assistenziale e l'altro per poi essere richiusi in manicomio intorno ai quindici anni. In particolare, Flavio venne ricoverato nell'ospedale psichiatrico di Arezzo in quanto "frenastenico e incapace di provvedere a sé stesso e pertanto pericoloso a sé e agli altri" (Bruzzone, 1979: 49).

[4] Software di trascrizione che trasforma il computer in un dittafono diminuendo la velocità dell'audio.

lettura eccessivamente faticosa, perciò ci si è limitati ai soli segni diacritici di uso comune e all'utilizzo del corsivo per le sole annotazioni tecniche. Inoltre, sempre per il motivo sopracitato, la punteggiatura è stata inserita per indicare pause, interruzioni e quant'altro, solo laddove il suo impiego non contrastava con le regole grammaticali e sintattiche (fatta eccezione per le virgole, utilizzate anche per indicare sequenze di autocorrezione). Conseguentemente, il risultato è stato quello di un utilizzo della punteggiatura marcato, non sempre coerente con i criteri interpuntivi italiani e le partizioni di ordine testuale (Giannelli, Di Piazza, 1995: 55).

Sempre nel tentativo di rimanere quanto più possibile vicino al parlato, sono riportati tutti quei fenomeni tipici della produzione orale come esitazioni, false partenze, sequenze di autocorrezione e, soprattutto, forme dialettali. Proprio per quanto riguarda queste ultime, nonostante l'intervistato abbia viaggiato molto nel corso della sua vita, risulta marcata l'influenza del dialetto toscano e in particolare di quello aretino, riscontrabile tanto nella fonologia quanto nella morfologia del parlato analizzato. Il confronto fra trascrizione e testo a stampa ha messo in luce sia gli interventi 'macroscopici' di Bruzzone (tagli di contenuto), sia quelli più minuziosi (trattamento dei dati sensibili, alla difficile rappresentazione degli elementi tipici del parlato sopracitati). Assumono un certo rilievo anche le scelte relative alla eventuale resa in scrittura delle caratteristiche dialettali, sia più genericamente toscane che più specifiche del dialetto aretino.

Gli interventi di Bruzzone appaiono – come vedremo – orientati verso la leggibilità e la fruibilità: del resto, il fine ultimo della sua opera non era quello di presentarsi come un saggio di natura accademica, bensì come un atto di denuncia sociale sulla condizione delle istituzioni psichiatriche. Obiettivo di Bruzzone era dar voce a coloro che per anni erano stati vessati nel silenzio e dar loro la possibilità di raccontare le proprie esperienze così come le avevano vissute[5], ragione per cui era necessario mantenere i colloqui con i pazienti in un clima disteso e – se possibile – informale, in una posizione di 'vigile ascolto', per far sì che le voci dei ricoverati non subissero interruzioni e dominassero incontrastate sulla pagina (Bruzzone, 1979: 21), tanto che in nessuno dei documenti a noi giunti (manoscritti e dattiloscritti) c'è traccia degli interventi da lei compiuti nel corso delle interviste. Si evidenzia in questo la prima forte difformità tra la trascrizione di Bruzzone e quella prodotta da noi: nel primo caso si è di fronte a un testo la cui sola voce narrante è quella del paziente; nel secondo caso le voci presenti sono tre: quella dell'autrice, quella

[5] Per fare in modo che i pazienti si aprissero con lei, l'autrice non si presentò ai degenti come un'estranea che voleva strappar loro un'intervista, ma cercò di integrarsi nella loro piccola comunità in modo da accorciare le distanze e consentire momenti di confidenza. Nel periodo trascorso nell'ospedale di Arezzo partecipò a tutte le assemblee generali e alle riunioni che si tenevano subito dopo per discutere degli argomenti trattati ed era quasi sempre presente nella struttura per trascorrere il proprio tempo con i ricoverati (Bruzzone, 1979: 17). Bruzzone si volle così allontanare dalla figura fredda e distaccata dell'intervistatrice, prendendo le distanze dalla concezione, diffusa all'epoca, che per raccogliere testimonianze obiettive era necessario un totale distacco emotivo (Bruzzone, 1979: 10).

del paziente intervistato e quella di un altro paziente[6], anch'egli presente alla registrazione.

Informazioni preziose sul contesto dell'intervista vengono anche dal materiale di corredo: nei quaderni manoscritti, sulle parti interne della copertina, sia anteriormente che posteriormente, sono state annotate dall'autrice informazioni addizionali, assenti nella registrazione, come il titolo di studi dell'intervistato, i nomi dei luoghi dove questi ha frequentato i corsi professionali di scuola alberghiera e la qualifica raggiunta. Il manoscritto rappresenta difatti una vera e propria prima stesura del testo in quanto è Bruzzone stessa a informarci di come lo stesso giorno dell'intervista, alla sera, lei riascoltasse e trascrivesse il materiale sonoro raccolto in modo da fissarsi nella mente un ricordo vivido del colloquio che le avrebbe poi permesso, anche a distanza di tempo, di comprendere il racconto (Bruzzone, 1979: 22). Non stupisce affatto che nel manoscritto si possano ancora trovare tutte quelle informazioni che sono poi state cancellate, come la data e i nomi veri dell'intervistato e dell'altro paziente presente alla registrazione, così come tutti i nomi delle persone nominate durante il colloquio[7].

Per quanto concerne i tagli di contenuto, le parti di testo espunto risultano già cancellate a pennarello nella seconda versione dattiloscritta: oltre agli interventi dell'autrice e del secondo paziente presente all'intervista, sono state eliminate anche alcune parti di ciò che viene dichiarato dall'intervistato, principalmente quelle in cui egli non parla di sé, ma di altri, sia che si tratti di Pirella e Basaglia, che di altri medici e pazienti. I criteri e le ragioni dei tagli effettuati sono descritti nell'introduzione del libro. Nella versione edita da Einaudi il materiale venne selezionato e scremato per favorirne la leggibilità da parte di un pubblico largo. Una scelta che non risultò facile per Bruzzone, perché rischiava di alterare la logica intrinseca del racconto. (Bruzzone, 1979: 25).

Sia nei manoscritti che nelle versioni dattiloscritte sono assenti annotazioni tecniche relative all'avvio e all'interruzione delle registrazioni e non è neanche indicato il luogo nel quale si è svolta l'intervista, che si può comunque ipotizzare essersi tenuta all'interno data l'assenza di rumori di ambiente. Questa ipotesi pare anche essere confermata dal fatto che Anna Maria Bruzzone, per non intralciare

---

[6] L'intervista a Flavio venne svolta in presenza di una terza persona: un altro ricoverato, anche questo intervistato da Bruzzone e che appare nel libro sotto il nome di Bruno. È l'autrice stessa, nell'introduzione del proprio libro, a informarci che era stato proprio Bruno a farla incontrare con Flavio nella Casa-famiglia Uomini, e che questi fu presente all'intervista dell'amico. Infatti, se in principio Bruzzone sentiva una sorta di impaccio a chiedere ai pazienti di poterli intervistare, successivamente erano loro stessi a portarli dai propri amici o a chiedere, in forma più o meno esplicita, di raccontare ciò che avevano da dire. Bruzzone non si era prefissata criteri troppo rigidi per scegliere i propri intervistati, l'ascolto non veniva negato a nessuno e le conversazioni procedevano liberamente secondo la proposta avanzata dai ricoverati all'autrice di non porre quesiti precisi ma di ascoltare quanto avrebbero voluto dirle (Bruzzone, 1979: 19-22).

[7] La trascrizione manoscritta e la prima versione dattiloscritta presentano ancora il nome reale dell'intervistato tra parentesi, mentre nella seconda versione dattiloscritta esso è cancellato con un pennarello; inoltre, già nel passaggio dai manoscritti alla versione dattiloscritta, i dati sensibili quali nomi di persone e luoghi sono stati modificati.

le voci dei narratori con rumori e suoni, preferisse svolgere i colloqui in biblioteca, luogo in cui il rivestimento in legno e i libri permettevano buone registrazioni (Bruzzone, 1979: 19-20).

Per quanto riguarda la resa del parlato, è utile ricordare che Bruzzone non opera una trascrizione linguistica *tout court* ed è quindi normale che nel suo testo molti fenomeni abbiano finito per andare persi. Infatti, seppure nella versione manoscritta e nel documento dattiloscritto si possono trovare casi, per quanto rari, in cui vengono indicate false partenze e ripetizioni, nella versione edita del testo – forse per privilegiare il senso più ampio del discorso – questi elementi linguistici risultano quasi del tutto assenti. L'autrice dichiara di aver provato a mantenere le parole e locuzioni toscane, trascrivendole così come erano state pronunciate, scegliendo grafie che non appesantissero la lettura (Bruzzone, 1979: 24). Ma, se si confrontano la trascrizione e il testo edito, si può notare come, quantomeno per quanto riguarda l'intervista di Flavio, molte volte le forme, soprattutto quelle tipicamente aretine, siano scomparse.

Le ragioni di questa "ripulitura" possono essere molteplici. La più evidente è la collocazione editoriale: Einaudi è infatti un editore di primissimo piano, perciò è plausibile ipotizzare che Bruzzone sia stata indirizzata verso una scelta più classica. Tuttavia, avranno senz'altro avuto un certo peso anche le finalità di denuncia sociale e quindi la volontà di raggiungere un pubblico quanto più possibile ampio, senza ostacolare la lettura con forme che per qualcuno sarebbero potute risultare oscure nella loro comprensione.

2.1 Esempi di comparazione

Gli esempi seguenti mostrano le diverse tipologie degli interventi compiuti dall'autrice. In prima battuta vediamo gli estratti della trascrizione svolta (in sigla, T), in cui sono messi in luce, tramite sottolineature ed evidenziazioni, tutti quegli aspetti che hanno subito modifiche durante il processo editoriale; poi compaiono gli estratti corrispondenti riscontrabili all'interno del libro *Ci chiamavano matti* (in sigla, TS), accompagnati da una spiegazione delle differenze tra i passi analizzati.

All'inizio dell'intervista l'intervento dell'autrice scompare: l'unica voce narrante nel testo edito è quella di Flavio. Trattandosi dell'inizio della conversazione tra intervistato e intervistatrice, c'è una presentazione del paziente in cui sono ovviamente presenti dati sensibili (evidenziati), modificati nel testo edito[8].

Le parti sottolineate sono fenomeni tipici del parlato: esitazioni, false partenze e sequenze di autocorrezione che vengono del tutto espunte nell'edizione finale del testo.

[8] L'entità della modifica è tuttavia solo parzialmente verificabile poiché, per ragioni di tutela della riservatezza nei confronti delle persone coinvolte, ci si è dovuti limitare al solo nome di fantasia attribuito al paziente nel libro, e per i luoghi citati soltanto alle iniziali.

> **T**[9]: F. – Sono nato il 30 aprile 1904, divid, quarant, quarantacinque. Nel... Se vuoi prendere dei dati...
> A.M.B. – Ora stanno per essere registrati i dati.
> F. – Sì, sì. Eeeeh... nel comune di B. e sono nato... Sono figlio di or, di or, di genitori sconosciuti... Bene. Io, diciamo, ho trasc, eeeh... ho trascorso, mmm... i, il limite della vita, dopo essere, dopo la nascita, nei vari istituti d'Italia, cioè, come B.
>
> **TS**: Io sono nato il 30 aprile 1945 nel comune di Montevarchi e sono figlio di genitori sconosciuti.
> Io ho trascorso il limite della vita dopo la nascita nei vari istituti d'Italia [...] (Bruzzone, 1979: 80).

Nell'estratto che segue, invece, il paziente racconta di come, almeno una volta nel corso della propria vita, si sia ritrovato a pensare al suicidio, ma abbia poi cambiato idea:

> **T**: A livello di questa, arivare a questo punto qui potrei farlo. Però dopo ci rifles, ci riflettii, cioè: ci riflessi sopra [...]
>
> **TS**: Arrivati a questo punto qui potrei farlo". Però dopo ci riflessi sopra [...] (Bruzzone, 1979: 82)

La falsa partenza e la sequenza di autocorrezione (sottolineate) vengono nuovamente neutralizzate nell'edizione finale. Si noti qui un primo esempio di una caratteristica dialettale tipicamente aretina (evidenziata): si tratta del processo fonologico per cui in alcuni contesti avviene lo scempiamento delle consonanti doppie (Nocentini, 1989: 35). Non si riscontra la presenza di queste varianti nel testo di Bruzzone.

Il passo successivo è parte di una riflessione di Flavio sulle reazioni della gente comune all'apertura dei manicomi e al conseguente contatto con i degenti:

> **T**: perché avvenendo tutto insieme senza aver incontrato la popolazione, senza aver conversato con loro, senza avé conosciuto chi eron loro, già nascevano dei pregiudizi.
>
> **TS**: perché avvenendo tutto insieme senza aver incontrato la popolazione, senza aver conversato con loro, senza aver conosciuto chi erano loro, già nascevano dei pregiudizi [...] (Bruzzone, 1979: 83)

Due caratteristiche tipiche del dialetto aretino – la forma di infinito con apocope dell'ultima sillaba: *avé*; e il verbo *essere* coniugato alla terza persona plurale dell'imperfetto indicativo con desinenza in *-ón*: *erón* – vengono sostituite nel testo edito dalle regolari forme italiane.

L'estratto seguente compare quasi alla fine dell'intervista, in un contesto in cui Flavio sta parzialmente criticando la linea di Pirella poiché ritiene che la sua équipe non sia all'altezza delle iniziative prese:

> **T**: a un certo punto 'n è che ecco io, io, sono molto tragico, tragico su, su una linea così del genere perché non, non puole essere così di, diffamata di, diciamo... perché

[9] Nell'esemplificazione, sottolineature ed evidenziature sono segnali di quanto indicato di volta in volta a testo: dati sensibili, fenomeni del parlato e tratti caratteristici dialettali.

> mi sa che piano piano le cose invece di migliorare si distruggono perché, perché… […] Tu cambia… puole essere anche quello, ma tu prova a cambiar l'équipe, prova a avere un'équipe come ha avuto Basaglia, vuoi scommettere che qualcosa viene fuori di reale?
>
> **TS**: A un certo punto, ecco, io non so, io sono molto tragico su una linea così del genere, perché non può essere così diffamata, diciamo, perché mi sa che piano piano le cose invece di migliorare si distruggono. […] tu cambia, pol essere anche quello, ma tu prova a cambiar l'équipe, prova a avere un'équipe come ha avuto Basaglia, vuoi scommettere che qualcosa viene fuori di reale? (Bruzzone, 1979: 88)

In questo caso, le false partenze (sottolineate) sono in parte riprodotte nel testo edito (ugualmente sottolineate). Per quanto riguarda invece la forma dialettale toscana *puol(e)* (evidenziata) alla terza persona singolare del presente indicativo del verbo *potere* in luogo di *può*, questa viene o riprodotta dall'intervistatrice nella versione edita con la forma *pol(e)* o con la regolare forma italiana standard. Si tratta dell'unica forma dialettale non del tutto neutralizzata nel testo.

Le modifiche apportare da Bruzzone al documento originale sembrano non rispettare integralmente le sue iniziali intenzioni di mantenere viva la parlata toscana: le caratteristiche tipiche del parlato sono difficilmente rintracciabili nel testo. Tuttavia, bisogna dire che era stata l'autrice stessa a rendersi conto di come il parlato, una volta messo su carta perdesse il calore dato dal ritmo, dal timbro, dalla voce, dalle pause e dal "bellissimo accento toscano" (Bruzzone, 1979: 22-24), tastando così con le proprie mani le difficoltà che comporta la resa grafica di un testo orale.

3. *Il punto di vista della storia orale: Ritrovare Roberto*

Fra le attività più significative svolte intorno all'archivio di Anna Maria Bruzzone vi è stata senz'altro la possibile ricostruzione dei rapporti con i soggetti che nel tempo avevano abitato o attraversato lo spazio manicomiale: ex pazienti, medici, operatori sanitari, famiglie, cittadini. Attivare, dunque, le memorie circostanti sia per ampliare la documentazione archivistica, sia per restituire alla cittadinanza questo patrimonio culturale di rilievo. È parso necessario cercare di rintracciare gli eredi degli intervistati e risolvere i problemi legali relativi all'ascolto e alla fruizione delle interviste di Bruzzone (Calamai, 2018). In questa direzione sono stati intensificati i rapporti con associazioni locali, quali ad esempio il Centro Franco Basaglia di Arezzo, e con numerosi operatori psichiatrici che, a vario titolo, hanno condiviso per questo progetto la loro esperienza. Le iniziative locali e gli articoli apparsi sulla stampa hanno permesso a molti altri cittadini di venire a conoscenza del lavoro e dei progetti sorti intorno all'archivio della studiosa torinese e più in generale a quello dell'ex ospedale e di mettersi in contatto con il gruppo di ricerca.

Grazie alla collaborazione di quanti avevano lavorato, a vario titolo, nell'ex ONP, è stato possibile rintracciare alcuni degenti che Bruzzone a suo tempo aveva intervistato: Roberto è uno di loro. Dopo aver felicemente comunicato la sua volontà e disponibilità a raccontare della sua esperienza e della sua attuale esistenza, non ha mostrato impaccio o remore nell'essere coinvolto nel progetto, rendendosi disponibile per una nuova intervista. A questo punto, era d'obbligo chiedersi come immaginare una nuova intervista e come riascoltare, insieme a Roberto, la sua giovane voce registrata tanti anni prima da Bruzzone. Quello che era importante sapere dall'intervistato non era solo cosa pensasse o ricordasse dell'apertura e della definitiva dismissione dell'ospedale psichiatrico, ma avere anche una sua testimonianza sulla vita di allora, durante e dopo l'esperienza del ricovero. Proprio la metodologia descritta dalla studiosa nell'introduzione del libro *Ci chiamavano matti* indica una strada possibile su come procedere. Bruzzone infatti dedica ampio spazio alla descrizione del rapporto con gli intervistati e proprio la relazione fra intervistato e intervistatore è al centro da anni delle riflessioni metodologiche relative alla storia orale (cfr. Bonomo, 2013: 24-25). La studiosa torinese evidenzia la rilevanza che ha avuto nel suo lavoro il fatto di aver trascorso larga parte del suo tempo all'interno dell'istituzione manicomiale: ha condiviso con loro momenti di vita collettiva, confidenze intime, dolori passati, ricordi "neri" del vecchio manicomio, emozioni, silenzi, sorrisi, costruendo quotidianamente un dialogo che non si riduceva solo al momento dell'intervista registrata.

Non potendo sperimentare una situazione analoga si è cercato di organizzare più incontri con Roberto, e in momenti diversi, per facilitare la relazione e stabilire una fiducia reciproca. Si è deciso così di non portare alcuna strumentazione al primo incontro, onde creare un'occasione di conoscenza facilitata dalla presenza dei mediatori. L'obiettivo è informare Roberto su come si intende raccogliere la sua testimonianza. L'incontro non va esattamente come previsto: la presenza dei mediatori, pur fondamentali per acquisire il contatto e rassicurare Roberto, risulta essere un po' troppo invadente. Mentre lo accompagnano all'appuntamento infatti, sono loro ad interrogare Roberto sulla sua esperienza in ospedale, aspettandosi che l'intervista venga svolta alla presenza di tutti. Chiarito l'equivoco, stabiliamo di proseguire il lavoro in un incontro successivo. È chiaro tuttavia che le sollecitazioni dei mediatori hanno fatto leva su un sentimento molto forte in Roberto: la nostalgia per la comunità terapeutica istituita da Agostino Pirella nel 1971 e in cui l'ex degente ricoverato a partire dal 1974 si sentiva protetto e valorizzato; non sarà facile ampliare il racconto anche ad altre fasi della vita di Roberto. Per il secondo incontro è stata preparata la strumentazione necessaria alla registrazione: gli audio-registratori e una telecamera, perché l'intervistato ha espresso la volontà di essere ripreso. Seguendo le indicazioni di Bruzzone, non è stato costruito un questionario e neanche una lista di temi da affrontare, così che sia Roberto a introdurre gli argomenti e a condurre il racconto, pur interrotto da eventuali domande e richieste di approfondimento.

L'incontro si è svolto al Parco del Pionta, che in precedenza ospitava l'ospedale e oggi accoglie l'università e l'archivio, precisamente nel giardino dove una volta si svolgevano le feste dell'amicizia per far incontrare la cittadinanza con i ricoverati. Roberto comincia subito a parlare, senza bisogno di sollecitazioni e prende la parola proprio a proposito dell'argomento che gli sta più a cuore, ovvero ricordare la comunità terapeutica istituita da Pirella, all'interno della quale lui si occupava del cinema, delle feste da ballo e dei tanti momenti ricreativi destinati ai ricoverati; nomina spesso molti di loro e le giornate trascorse insieme nel parco, al bar, ma soprattutto in assemblea generale dove veniva discusso tutto l'andamento dell'esperienza. Ci sono due elementi che intensificano la nostalgia che Roberto nutre per la comunità terapeutica: da un lato il fatto che vi trovasse protezione e tutela; dall'altro, l'opportunità che qui ebbe di vivere un protagonismo mai più sperimentato altrove, ovvero di essere ascoltato da una pluralità di persone, di prendere parte a un dibattito sulla salute mentale, così cruciale negli anni Settanta, anche in una cittadina come Arezzo (mentre oggi Roberto non ha la possibilità di confrontarsi con gli altri). Il primo incontro ruota tutto intorno a questo argomento, si decide così di rimandare ad un secondo incontro l'approfondimento di altri argomenti e di fargli ascoltare l'intervista di Anna Maria Bruzzone in cui un giovane Roberto raccontava le tristi vicende relative alla sua vita prima dell'ingresso in ONP. Quell'ascolto avrebbe di certo inciso sulle sue memorie più profonde.

4. *Riascoltare l'intervista*

L'ascolto si è svolto all'interno dell'archivio storico dell'ex ospedale psichiatrico, sul tavolo vi erano il computer collegato ad una cassa da stereo, la telecamera e gli audio-registratori per raccogliere le reazioni ed i commenti di Roberto. Egli si è seduto vicino alla cassa, e subito, dopo aver cominciato l'ascolto, ha chiesto di interrompere desideroso di sapere di quale intervista si trattava. Il suo ricordo più vivo era legato all'intervista rilasciata per la Rai che fa parte del filmato *Dietro l'alibi della follia*, girato per la trasmissione *Cronaca* negli anni Settanta. Ricorda anche di essere stato intervistato da Rai 2 e che il suo discorso venne mandato in onda al telegiornale. Non si ricorda di Bruzzone e della conversazione avuta con lei, ma dei momenti che testimoniano il protagonismo dei ricoverati all'interno della comunità terapeutica. In quest'organizzazione veniva infatti rovesciata la tradizionale gerarchia dell'istituzione che, invece, era concepita come un organismo complesso composto da medici, infermieri, degenti e personale ausiliario di cui era necessario sfruttare tutte le risorse in senso terapeutico. I momenti di confronto, le assemblee generali e di reparto, le riunioni di équipe introdotti in ospedale avevano come scopo l'agevolare la partecipazione di tutti, in particolare dei degenti, all'organizzazione della vita quotidiana, che non era più appannaggio della sola intelligenza medica (Basaglia, 1968; Foot, 2014). Alle assemblee generali aretine presero parte anche molti visitatori esterni: giornalisti, fotografi, volontari, studenti. Guardare e ascoltare il mondo "di dentro" per presentarlo, finalmente, "fuori". Questa presenza dimostrava a chi viveva in un manicomio che esisteva nella società un reale interesse per le loro condizioni. Insieme alla solidarietà e alla possibilità di rapporti reciproci che offriva la comunità terapeutica, Roberto sente la nostalgia di questi momenti di protagonismo e della possibilità di venire veramente ascoltato. Ricordi e nostalgia si intensificano in un contesto di vita in cui è difficile e frustrante farsi ascoltare anche da chi avrebbe il compito di occuparsi delle piccole cose che caratterizzano la vita quotidiana.

> **R**[10]: Dopo venne la legge per cui io passai alle assistenti sociali del comune e lì ho avuto assistenti sociali una peggio di quell'altra [...] Insomma è stata una trafila, per trovarne una ci vole lo scoppio, ultimamente ho chiamato i carabinieri, perché alla Croce Rossa mi hanno detto: Lei prende troppo, non va bene lei deve fare la richiesta al comune che lei ha bisogno del pacco, ti rendi conto che trafila? Allora spesi dieci euro di scheda del telefonino per trovare un'assistente sociale, o era in permesso, o era in ferie, o era a un congresso, alla fine una mattina chiamai i carabinieri dissi: È successo questo e questo come devo fare? Devo andà giù a trovà [...] Dissero: No no, ci si pensa noi. [...] Ragazzi ma qui siamo de fòri! La gente non se rendon conto che noi siamo messi male come tutti quell'altri. Ma perché si devon comportà così?! Han rubato i soldi, che poi quei soldi che han beccato a livello europeo, tutti soldi che avevano stanziato per qui, per noi, spariti! Fino a tre anni fa el comune – perché erano gestiti dal comune – ce davan duecento euro all'anno. **C**: Eh è finito il fondo sociale europeo. **R:** Che fine han fatto un si sa, l'han rubati loro! No, guarda, io sto laggiù, ci sto benissimo, è l'unico edificio a Arezzo che ciànno da restaurare esternamente e non ciànno i soldi, dice, allora un giorno andai alle case popolari, avevo da fà un piccolo richiamo e vedo che stanno restaurando gli uffici delle case popolari, dico: Mi oh! Restauran gli uffici con i soldi che ci rubano a noi! A voce alta no? Perché è vero, con i nostri soldi ve fate l'uffici novi, tutto bello el parquet, noi esternamente sembra un accampamento del mille e novecentoquaranta: Eh no no, è la regione che ci manda i soldi per farlo. **C:** Ma questo dove lo dici te, in comune? **R**: Alle case popolari, poi ora ciò da chiappare il sindaco ora glielo faccio vedere anche a lui perché: ma come, dico? Perché ci pensan, ma come – dico – laggiù dove sto io le piante non l'han mai potate, allora nel duemila me dan la casa [...] eran quattr'anni che avevo fatto la domanda, piglio vengo giù la numero 146, vado su dopo averla scelta e così dice poi le chiavi le voglio subito immediatamente, mi preparo, le piante mi danno sulla terrazza, mi dicono guarda il giardino un è nostro, è del comune, telefoni al comune e senti no? Telefono al comune e dico: Guardate io ciò questa terrazza devo buttare fuori le foglie. Eh ma il Comune un ha i soldi per potarle. Li richiamo dice: Guardi ancora non è stata stabilita una data per la potatura. E va bene, insomma più di un mese un mese e mezzo, una mattina, ora – senti – li pettino io! Li richiamo e dico: Allora qui che si fa? Io compro una motosega la butto giù dal fondo! Lei non puole. Non posso? Presi la moto, il quod lo misi vicino alla pianta, dissi: Guardi che metto in moto! Misi in moto il quod, la mattina alle sette eran lì a potarla, ha' capito? [...] ma ragazzi ma indó siamo arrivati? (Intervista a Roberto A., 23/07/2018, Archivio storico dell'ex ospedale psichiatrico di Arezzo)

Proprio come era abituato a fare negli anni Settanta discutendo di ogni aspetto della quotidianità con i medici, gli infermieri, le assistenti sociali e gli amministratori, Roberto cerca continuamente di interagire con gli enti pubblici e con le istituzioni, non ha remore a far presente quello che pensa a qualsiasi funzionario con cui entra in contatto. Infatti, per Roberto l'unico modo per far fronte all'emarginazione in cui si

[10] Per la trascrizione si è fatto riferimento alle norme elaborate da Luciano Giannelli e Valeria Di Piazza (Giannelli, Di Piazza, 1995). Con R si intendere l'intervistato, con C l'intervistatrice.

trova è denunciare la propria condizione e rianimare il dibattito sulla chiusura degli ospedali psichiatrici e sulle difficoltà da affrontare quando si è fuori da quel contesto.

> **R**: Eh il discorso è uno, noi e' si ritorna un attimino ar discorso che io ho fatto nell'intervista, quando si tocca certe strutture noi siamo eliminati, tu fai caso ad una cosa, il drogato ha tutto, te facci caso, gli danno perfino ciàn le comunità non pagan niente, fanno.. eh, hanno tutto! E i cosi? L'extracomunitari han più de noi! Te facci caso. [...] Io parlo di me stesso, noi che siamo stati in queste strutture, io parlo di questo, noi che siamo stati in queste strutture nell'ospedale psichiatrico come lo chiamano tutti noi per la società siamo pericolosi. **C:** Ma ancora Roberto secondo te? **R:** E ancora qualcuno c'è, ancora la società è a barzello perché se vede se uno è stato in manicomio capito? **C:** Ma di te un si vede mica che tu sei stato in manicomio. **R:** Io ho rischiato la patente per questo discorso dell'ospedale psichiatrico. A me non m'han dato la patente, io fino a ottobre posso viaggiare con la moto, dopo se rinnovo la patente che devo fare? Devo tornà al SIM e chiedere una visita psichiatrica che io non so più un malato mentale e comporta delle spese, vengo e trecento euro! Poi altre visite, siamo segati fòri, siamo oramai c'è gente che un ha le gambe e guidano – dico – però vedi tutto nell'insieme noi siamo emarginati! Gente che ha la patente che viaggia senza assicurazione, quelli che hanno ammazzato per omicidio stradale dopo un anno o due sono fòri e guidano, ti sembra giusto questo discorso? **C:** No. **R:** Ecco, perché dopo tutto noi siamo emarginati! La matematica non è un'opinione, te fai el conto! **C:** E secondo te Roberto come si può risolvere questa cosa? **R:** Questa cosa si può risolvere facendo degli incontri con le persone e trovarsi davanti a loro, guarda che quello che voi pensate non è vero, fargli vedé il filmato *Dietro l'alibi della follia*, fàgli vedé dell'interviste chiare di queste cose e portargli anche delle foto perché c'è alcune foto che non sono vere, sono montate perché alcune cose qui hanno esagerato de quello che era [...] era tutta una cosa un po' montata quel che era l'ospedale psichiatrico, per me, poi per gli altri non lo so, io parlo per me stesso non parlo per quell'altri però noi siamo stati tagliati fòri, la generazione una volta sparirà, questa generazione se sparirà tutti, allora la gente... mah, se ne fregavano! La generazione del vecchio manicomio è sparita, morta lì! Capito il discorso? Invece dovrebbero rivedere le cose reali per farli capire.

Prima di potersi concentrare sulle sensazioni che l'ascolto suscita, Roberto ha bisogno di ribadire l'importanza e il significato che per lui ha avuto fare parte della comunità terapeutica, occuparsi del cinema e dei momenti ricreativi, nominare gli altri pazienti, insieme agli infermieri, e ricordarli tutti come una grande famiglia. Fra le analogie più significative tra il racconto reso a Bruzzone e quello raccolto oggi resta la necessità di additare tra i motivi del ricovero la perdita della famiglia, la prematura morte del padre, l'alcolismo della madre e l'impossibilità di costruire una propria famiglia, che i suoi fratelli invece sono riusciti a costruirsi. Nell'ospedale psichiatrico diretto da Agostino Pirella Roberto aveva trovato quella tutela e quella protezione che non aveva sperimentato fra le mura domestiche. Per questo si rammaricava con Bruzzone della volontà dei medici di reinserire le persone in società e protestava per il timore di non saper affrontare in autonomia le difficoltà della vita "fuori". Per le stesse ragioni si dilunga sulle difficoltà che affronta nella vita di ogni giorno, mentre allora poteva contare su un'intera comunità di affini, in cui era possibile avere lo spazio per esprimere le proprie preoccupazioni e la propria opinione. Solo in parte l'intervista stimola il racconto di altre vicende della vita di Roberto: l'entusiasmo provato nel sapere che i nastri di Anna Maria Bruzzone sono stati ritrovati enfatizza ancora di più il bisogno di dichiarare qual è la propria posizione e il proprio vissuto relativamente a questa esperienza. Per lui, il ritrovamento di questo materiale apre la possibilità di riaccendere nell'opinione pubblica l'interesse per la condizione degli emarginati.

Dopo questa giornata dedicata a riascoltare insieme a Roberto l'intervista di Anna Maria Bruzzone è maturata all'interno del gruppo di ricerca la decisione di invitare Roberto alla Notte dei Ricercatori, iniziativa promossa dal dipartimento di Scienze della formazione, Scienze umane e della comunicazione interculturale e dedicata all'archivio sonoro della studiosa torinese[11]. È stata questa l'occasione in cui la propria voce, a distanza di quarant'anni, ha suscitato un enorme impatto emotivo in Roberto. A differenza che nel primo ascolto avvenuto in privato, la propria testimonianza era qui condivisa con un vero e proprio pubblico e con alcuni giornalisti. È stato così possibile per Roberto rivivere per un pomeriggio quel protagonismo sperimentato negli anni Settanta e far arrivare la propria voce e la propria esperienza alla città e alla comunità, il che rappresenta proprio una delle motivazioni che spinsero Anna Maria Bruzzone, nell'estate del 1977, ad accendere il registratore dentro un ospedale psichiatrico.

5. *Osservazioni conclusive*

Il lavoro avviato sull'Archivio mostra come le scelte di trascrizione siano indissolubilmente legate agli obiettivi generali di un lavoro con le fonti orali: cosa è rilevante per un linguista è spesso di scarso interesse per lo storico orale, che nel costruire archivi orali ha obiettivi e finalità diverse. L'analisi comparata delle diverse versioni delle interviste dell'Archivio di Anna Maria Bruzzone offre tuttavia un ambito ove le diverse discipline possono incontrarsi: cosa si espunge dalla versione 'pubblica' di un documento orale non è solo una falsa partenza o un dialettalismo, ma anche un riferimento a dati sensibili, o anche un cenno stonato e discordante rispetto a una ricostruzione storica e sociale che l'autore vuole raffigurare nel volume edito. Il ritrovamento di un paziente dell'epoca consente poi uno scavo nella diacronia del personaggio – la voce di allora e la voce di oggi – in un percorso di cui è possibile intravedere le potenzialità e anche l'eccezionalità.

Bibliografia

Agostiniani L., Giannelli L. (1990). *Considerazioni per un'analisi del parlato toscano.* In Cortelazzo, M.A., Mioni, A.M. (Eds.), *L'italiano regionale. Atti del XVIII congresso*

[11] http://www.rainews.it/dl/rainews/TGR/media/tos-manicomio-arezzo-d942e038-f797-4f96-9f6c-f63678fcbe72.html.

internazionale di studi della Società di Linguistica Italiana. Padova-Vicenza, 14-16 settembre 1984 Roma: Bulzoni, 219-237.

Art Bonus, *Interventi di manutenzione, protezione e restauro di beni culturali pubblici. Patrimonio documentario e bibliografico dell'Ex Ospedale Neuro Psichiatrico di Arezzo*, Arezzo, http://artbonus.gov.it/1398-patrimonio-documentario-e-bibliografico-dell'ex-ospedale-neuro-psichiatrico-di-arezzo.html.

Basaglia, F. (1968). *L'istituzione negata. Rapporto da un ospedale psichiatrico.* Torino: Einaudi.

Bianchi, C. (25/10/2017). *Donato all'Università l'archivio sonoro che conserva le voci dei ricoverati nel manicomio di Arezzo*, http://unisinforma.unisi.it/2017/10/25/donato-alluniversita-larchivio-sonoro-della-storica-bruzzone-che-conserva-le-voci-dei-ricoverati-nel-manicomio-di-arezzo/.

Bonomo, B. (2013). *Voci della memoria. L'uso delle fonti orali nella storia.* Roma: Carocci.

Bruzzone, A.M. (1979). *Ci chiamavano matti, voci da un ospedale psichiatrico.* Torino: Einaudi.

Calamai, S., Biliotti, F. (2017). *Le voci dei matti. Il ritrovamento dell'archivio sonoro di Anna Maria Bruzzone.* In Baioni, M., Setaro, M. (Eds.), *Asili della follia. Storie e pratiche di liberazione nei manicomi toscani.* Pisa: Pacini editore, 22-34.

Calamai, S., Kolletzek, C., Kelli, A. (2018). Towards a protocol for the curation and dissemination of vulnerable people archives. In *CLARIN2018 Book of Abstracts, Paper session 4*, https://www.clarin.eu/clarin-annual-conference-2018-abstracts.

Foot, J. (2014). *La "Repubblica dei matti". Franco Basaglia e la psichiatria radicale in Italia, 1961-1978.* Torino: Einaudi.

Giannelli, L. (1988). Introduzione alla lettura. Il testo come documento di lingua. Problemi di rappresentazione. In Di Piazza, V., Mugnaini, D. (Eds.), *Io so' nata a Santa Lucia. Il racconto autobiografico di una donna toscana tra mondo contadino e società d'oggi*. Firenze: Tip. Baccini & Baldi S.N.C., 43-62.

Giannelli, L., Di Piazza, V. (1995). Una proposta metodologica per l'edizione dei documenti orali del fondo Roberto Ferretti. In *Quaderni dell'Archivio delle Tradizioni Popolari della Maremma Grossetana*, 2, 51-71.

Montanari, L. (17/11/2017). *Matti da ascoltare*, Il venerdì di Repubblica.

Nocentini, A. (1989). *Il vocabolario di Francesco Redi*. Firenze: Elite.

Rastelli, A. (19/11/2017). *Il diario con dentro le voci dei matti*, Corriere della sera – La lettura.

Snow, E. (1965). *Stella rossa sulla Cina*. Torino: Einaudi.

INTERVENTI ALLA TAVOLA ROTONDA SUGLI ARCHIVI ORALI

DOI: 10.17469/O2106AISV000010

SIMONETTA BUTTÒ

Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale

La sempre crescente attenzione che il mondo della ricerca scientifica riserva al patrimonio immateriale ha contribuito a evidenziare la grande varietà di archivi sonori, realizzati nel corso degli anni, che riguardano molti campi delle scienze umanistiche e sociali: dalle discipline linguistiche (fonetica, fonologia, dialettologia, sociolinguistica) alle scienze che si occupano della voce, dell'elaborazione del linguaggio naturale, del riconoscimento automatico del discorso, fino alla storia orale e all'etnografia, alla sociologia, all'antropologia.

Questi archivi contengono da una parte una grande quantità di dati di ricerca, e rappresentano dunque fonti importanti per le scienze sociali e le discipline umanistiche, ma, allo stesso tempo, costituiscono ormai anche un settore ben definito del patrimonio culturale del Novecento.

Quello della loro descrizione, conservazione e valorizzazione, in Italia e all'estero, è un tema di grande attualità, che acquista dimensioni sempre più estese con il progredire delle soluzioni tecnologiche applicabili ai processi di acquisizione dei dati, alla loro consultazione e all'utilizzazione nei lavori di ricerca. Sono oggi sempre più numerosi gli istituti universitari, le biblioteche e i centri culturali che raccolgono e custodiscono documenti sonori, consentendone la consultazione nelle rispettive sedi o nella rete. Tuttavia, come accade per tanti altri tipi di risorse digitali prodotti dalle istituzioni, numerosi "giacimenti" di documenti sonori sono ancora da scoprire e molti di essi sono poco conosciuti e scarsamente utilizzati: il fenomeno della dispersione li riguarda molto da vicino, anche a causa della loro varietà tipologica e della delicatezza delle problematiche relative alla diffusione in rete di voci e di suoni.

Ai tanti interrogativi posti in questo Convegno in merito al censimento, alla descrizione, alla diffusione, all'uso e al riuso ai fini della ricerca di un tale patrimonio di conoscenze trasversali, è forse possibile fornire alcune risposte prendendo spunto dall'esperienza che l'Istituto Centrale per il Catalogo Unico delle biblioteche italiane-ICCU ha sviluppato all'interno delle reti istituzionali italiane ed europee, che coordina o alle quali ha partecipato fin dalla sua istituzione, con l'obiettivo di diffondere procedure condivise, strategie comuni, flussi di lavoro standardizzati.

Un quadro tanto eterogeneo non deve spaventare, deve semmai indurre a ragionare sulla necessità di adottare standard riconosciuti da un'ampia platea, anche internazionale, a garanzia della piena interoperabilità tramite protocolli di scambio.

Non è più tempo di andare ciascuno per conto proprio, ora meno che mai, perché è diventato indispensabile razionalizzare le risorse di personale, scarse ovunque, e quelle economiche, sempre incerte, ma anche perché le tecnologie che possediamo ce lo consentono con una facilità prima impensabile.

In questo contesto, dunque, la parola d'ordine non può che essere "condivisione", lo stesso principio fondante della comunità creata oltre trenta anni fa dal Servizio bibliotecario nazionale-SBN, che oggi accoglie oltre 6500 istituzioni culturali di ogni tipologia funzionale e di ogni appartenenza amministrativa che partecipano alla catalogazione collettiva del patrimonio bibliografico italiano[1].

Oggi SBN è ben di più di un base di dati bibliografica: nei suoi 18 milioni di registrazioni (con oltre 98 milioni di esemplari localizzati in Italia) sono presenti vari materiali non librari, sicuramente di largo interesse per il pubblico generale, ma anche per gli studiosi: ritratti, immagini di luoghi (mappe, stampe, fotografie, cartoline, ecc.), autografi, carteggi o documenti riferibili a persone, mostre virtuali, oggetti d'arte, cimeli e – ormai non più trascurabili per qualità e quantità – risorse audio o video relative sia alle persone (interviste, discorsi registrati) che a luoghi (documentari), oltre ai cosiddetti "materiali non pubblicati", come conferenze, appunti, dispense, esecuzioni musicali presentate nei festival o in piazza, oppure per esami di concorso.

SBN ha infatti alle spalle una lunga tradizione nell'acquisizione a catalogo di documenti sonori o musicali: l'interesse per la documentazione musicale è nato negli anni Ottanta del Novecento, con la creazione di una apposita *Base dati Musica*, e nel corso degli anni sono state prodotte line guida per la catalogazione in SBN della musica, dei libretti, delle registrazioni sonore, dei video e delle risorse elettroniche digitali[2], fino alle recenti *Norme di catalogazione delle risorse musicali non pubblicate* (maggio 2018), tutte consultabili su piattaforma wiki all'interno del sito dell'Istituto[3].

Un grande traguardo per l'implementazione dei beni sonori e audiovisivi nel catalogo, arricchiti da link alla corrispondente risorsa digitale, se disponibile, è stato raggiunto alla fine del 2015 grazie al riversamento dei record bibliografici dell'Istituto centrale beni sonori e audiovisivi (ICBSA): operazione alquanto complessa, per compiere la quale sono state messe in campo le migliori energie dei due istituti centrali, realizzando una sinergia tale da rendere sostenibili anche le operazioni di pulizia del catalogo dalle duplicazioni, inevitabili nei processi di importazione massiva di tale portata.

In SBN hanno dunque fatto il loro ingresso le registrazioni su supporto audio e video acquisite dall'ICBSA per deposito legale: circa 300.000 monografie (dischi in vinile, audio e videocassette, compact disc, DVD, nastri, ecc.), e 1.500.000 spogli, ovvero brani sonori (e relativi link alla copia digitale). In questo prezioso insieme sono incluse sia fonti musicali, sia fonti orali, come ad esempio l'Archivio Etnico Linguistico Musicale - AELM[4].

[1] https://opac.sbn.it/opacsbn/opac/iccu/free.jsp.

[2] https://www.iccu.sbn.it/export/sites/iccu/documenti/2012/Guida_alla_catalogazione_in_SBN_Musica.pdf.

[3] http://norme.iccu.sbn.it/index.php?title=Norme_musica_non_pubblicata.

[4] http://www.icbsa.it/index.php?it/116/aelm-archivio-etnico-linguistico-musicale.

Frutto dell'importazione di tali materiali è la possibilità, dal 2017, di visualizzare oltre 200.000 incipit musicali con il pentagramma e le note, che attraverso l'OPAC di Polo SBN Web possono essere anche ascoltati.

Per arricchire ulteriormente la presenza delle risorse musicali nel catalogo nazionale, nell'estate del 2016 è stato firmato un Accordo tra ICCU, Internationales Quellenlexikon der Musik e.V. (RISM) e Associazione Italiana delle Biblioteche, Archivi e Centri di documentazione musicale (IAML-Italia)[5], con l'obiettivo di individuare e importare nel catalogo SBN i record relativi a risorse conservate in biblioteche italiane oggi presenti solo nei cataloghi cartacei ed elettronici del RISM; rendere i record bibliografici e di autorità presenti in SBN ricercabili e indicizzabili anche nel catalogo online del RISM; individuare e risolvere le duplicazioni presenti nelle basi dati SBN e RISM; implementare nell'OPAC SBN la ricerca e la visualizzazione dell'incipit musicale.

Anche in *Internet Culturale*[6], la teca digitale delle biblioteche italiane, si possono consultare numerose risorse sonore tramite il *player* dedicato che consente di accedere direttamente alla risorsa sonora e ascoltare, per intero se fuori diritti, o per soli 30 secondi, le tracce fornite dai partner del portale. Attualmente si tratta di oltre 135.000 notizie bibliografiche che corrispondono a quasi 35.000 monografie e circa 100.000 spogli.

Le risorse digitali dell'ICBSA hanno dato origine a tre collezioni di *Internet culturale*: *Fondo storico cilindri fonografici*, *Fondo storico dischi* e *Fondo Compact Disc* dove sono consultabili tanto risorse musicali, quanto fonti orali (discorsi, brani teatrali e letterari, programmi radiofonici). Per quanto riguarda le fonti per la tradizione orale è anche significativa la presenza del *Fondo Majorano* che contiene canti della tradizione orale di Taranto e provincia, raccolti dall'etnografo Alfredo Majorano Nunziato[7].

Musica, voci, discorsi e canzoni sono anche presenti nel portale tematico *14-18: Documenti e immagini della Grande Guerra*[8], che raccoglie una documentazione quanto mai eterogenea sul primo conflitto mondiale, proveniente da 113 istituti di conservazione, per un totale di 645.000 risorse digitali, da anni in continua crescita.

Proprio a partire da queste risorse, nel novembre del 2016, in accordo con l'ICBSA e con Wikimedia Italia, è stata promossa una esperienza di riuso: l'*Edit-a-thon verdiano*, sulle opere di Giuseppe Verdi. In quella occasione, grazie alle tracce sonore, in versione integrale e libere dai diritti, concesse dall'ICBSA, gli esperti che hanno aderito e partecipato all'iniziativa hanno integrato, innalzandone il livello di autorevolezza, 19 voci esistenti in Wikipedia e ne hanno pubblicate 12 del tutto nuove.

Per consentire ad un pubblico, che ci auguriamo sempre più vasto, la consultazione integrata di tali risorse multidisciplinari presenti nelle varie basi dati gestite

---

[5] https://www.iccu.sbn.it/export/sites/iccu/documenti/2016/Convenzione_IALM_ICCU_RISM.pdf.
[6] http://www.internetculturale.it/.
[7] http://www.internetculturale.it/it/41/collezioni-digitali/26306/archivio-sonoro-alfredo-majorano.
[8] http://www.14-18.it/.

dall'ICCU (SBN, Internet culturale, il portale tematico 14-18, ma anche il sito delle mostre virtuali, MOVIO, il censimento dei manoscritti antichi e moderni *Manus OnLine* e quello delle edizioni italiane del XVI secolo EDIT 16), l'ICCU ha recentemente avviato un progetto, della durata di 24 mesi, che consentirà di navigare fra record di catalogo e risorse digitali a partire da un unico Portale delle biblioteche e degli istituti culturali italiani, una piattaforma di servizi informativi in grado di offrire al pubblico una nuova esperienza di ricerca basata sulla scoperta e la navigazione tra risorse di natura diversa ma riferibili a categorie di interesse come luoghi, persone, istituzioni[9].

Per quanto riguarda le basi dati appositamente create per lo sviluppo della ricerca scientifica in ambito linguistico, glottologico, fonologico, l'esperienza dell'ICCU si riferisce in primo luogo alla partecipazione, negli ultimi dodici anni, a una linea di progettualità europea dedicata specificamente alle *e-infrastructures*, piattaforme digitali per la ricerca, che costituiscono oggi un fattore unificante contro la frammentarietà esistente e una grande opportunità di condivisione e reciproco arricchimento offerta a università ed enti di ricerca che operano in tutti i paesi dell'Unione.

L'ICCU, in particolare, ha partecipato al progetto PARTHENOS[10], finanziato nell'ambito di Horizon 2020 e formato da un consorzio di 16 partner, che ha riunito le grandi infrastrutture europee nel settore del patrimonio culturale e delle *digital humanities*, come DARIAH, CLARIN, INRIA e importanti centri di ricerca come il DANS in Olanda e il PIN-Polo universitario della città di Prato, con l'obiettivo di armonizzare standard, modelli di gestione dei dati, protocolli di interoperabilità per aumentare la visibilità e la riusabilità dei dati della ricerca.

Le linee guida di PARTHENOS[11], recentemente pubblicate in inglese, sono indirizzate sia ai produttori di dati, e dunque ai ricercatori nell'ambito di discipline umanistiche come la storia, l'archeologia, gli studi linguistici e le scienze sociali, che alle infrastrutture di ricerca e agli istituti che gestiscono archivi e repository di dati.

Attraverso il paradigma FAIR (acronimo di *Findable*, *Accessible*, *Interoperable* and *Reusable*[12]) le raccomandazioni richiamano, in venti semplici punti, principi e metodi per migliorare l'accessibilità e il riuso dei dati[13]. Le linee guida sono state tradotte in italiano a cura dell'ICCU, che ne promuoverà in Italia la diffusione, ma anche l'adattamento alla realtà nazionale, perché è evidente che le soluzioni dei problemi di interoperabilità e di riusabilità non dipendono solo dalla tecnologia, ma riguardano le politiche nazionali, le licenze adottate, le restrizioni legali, gli aspetti etici collegati alla tutela della privacy e, non da ultimo, le figure professionali necessarie (su questo tema il progetto Parthenos è intervenuto offrendo una sezione di *training on-line* che presenta le migliori pratiche nella gestione dei dati di ricerca).

---

[9] https://www.iccu.sbn.it/it/eventi-novita/novita/Portale-delle-biblioteche-e-degli-istituti-culturali-italiani/.
[10] http://www.parthenos-project.eu.
[11] http://www.parthenos-project.eu/portal/policies_guidelines.
[12] FAIR, (2006). The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3.
[13] https://zenodo.org/record/2668479#.XOqn5CZ7m3A.

Per favorire la conservazione, la descrizione, il riutilizzo di risorse orali raccolte in ambito di progetti di ricerca è necessario dunque prevedere fin dall'inizio l'uso di standard che garantiscano la qualità dei dati e un buon livello FAIR e puntare sulla completezza dei metadati, che devono accogliere sia le informazioni descrittive della risorsa, che gli elementi tecnici e gestionali, necessari anche per la conservazione a lungo termine dell'oggetto digitale.

Anche *Europeana*[14], il portale del patrimonio culturale digitale della UE, ha definito un livello minimo di metadati (che può essere eventualmente arricchito con ulteriori informazioni), che consente a tutti gli aggregatori nazionali di inviare i dati. Questi – a loro volta – condividono con la rete nazionale degli istituti culturali gli standard per l'aggregazione dei metadati e la qualità degli oggetti digitali, oltre che le indicazioni sulle licenze da adottare.

Protocolli di scambio e standard descrittivi rappresentano dunque il presupposto per la costituzione di una rete tra associazioni, istituzioni, enti di ricerca, dedicate alla cura di dati trasversali a molte e diverse discipline.

Per quanto riguarda l'Italia, che vanta una collaborazione ultradecennale con *Europeana*, il ruolo chiave nel flusso di risorse digitali legate al patrimonio culturale italiano verso il portale europeo è svolto dall'aggregatore nazionale *CulturaItalia*[15], gestito dall'ICCU, che – insieme a *Internet culturale* – costituisce il ponte che collega le istituzioni culturali del nostro paese con la grande infrastruttura europea per il patrimonio cultural digitale, con cui condivide standard e *policy* di pubblicazione dei dati.

Così è avvenuto per il progetto tematico *Europeana Sounds* (2014-2017), guidato dalla British Library con l'intento di facilitare l'accesso e migliorare la visibilità dell'enorme patrimonio sonoro conservato presso le istituzioni Europee, nell'ambito del quale l'ICCU ha aggregato e reso disponibili 83.000 risorse sonore fornite dall'ICBSA.

A parte i notevoli risultati quantitativi (520.000 risorse provenienti da 24 istituzioni di 12 paesi europei) che hanno fatto di *Europeana Sounds* uno dei progetti tematici di maggiore successo nella grande famiglia di *Europeana*, un importante contributo allo studio delle politiche comuni per migliorare l'accesso al patrimonio sonoro digitalizzato è anche rappresentato dal documento *Europeana Sounds Policy Recommendations for improving online access to audio and audio-related heritage in Europe*[16] dove viene direttamente affrontata la delicata tematica del diritto d'autore, dal punto di vista etico, oltre che legale, che rappresenta una caratteristica peculiare delle fonti orali e richiede alla Commissione europea un impegno ulteriore per lo sviluppo di un bilanciamento tra la legislazione sulla protezione dei dati, la legge sulla privacy e la missione delle istituzioni del patrimonio culturale di rendere accessibili le proprie collezioni al pubblico.

---

[14] https://www.europeana.eu/portal/it.

[15] http://www.culturaitalia.it/.

[16] https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/europeanasounds-d3.4-policy-recommendations.pdf.

DOI: 10.17469/O2106AISV000011

SABINA MAGRINI, PIERO CAVALLARI

Voci di storia orale all'Istituto Centrale per i Beni Sonori ed Audiovisivi (ICBSA)[1]

Molti anni sono passati dal 1986, anno in cui l'allora Discoteca di Stato organizzò, in collaborazione con la Società di Storia Orale, il Convegno *L'intervista strumento di documentazione: giornalismo, antropologia, storia orale*[2]. Negli anni a seguire, in Italia sia la tradizione degli studi sia il legislatore hanno riconosciuto una ben diversa dignità alle risorse della storia orale e così nel 2009 l'Istituto Centrale per i Beni Sonori ed Audiovisivi – subentrato nel 2007 alla Discoteca di Stato già istituita nel 1928[3] – aveva organizzato un nuovo convegno dal titolo *L'intervista "fonte" di documentazione*. Con questa operazione si voleva evidenziare, come ha avuto modo di ricordare nella stessa premessa agli atti del convegno del 2009 il Direttore dell'Istituto Massimo Pistacchi, "l'evoluzione avvenuta negli ultimi venti anni in merito al concetto di intervista", passata per diversi settori di ricerca da mero strumento a vera e propria fonte[4].

Sempre nella sua premessa agli atti del Convegno del 2009, Pistacchi osservava a proposito dei problemi della conservazione e gestione delle raccolte di fonti orali poste da Nuto Revelli nel Convegno del 1986 che "Molti passi in avanti sono stati compiuti, l'attenzione del Ministero per i Beni e le Attività culturali attraverso i suoi più importanti organi è stata costante e concreta, ma certo esiste il problema di una forte dispersione che costituisce un ostacolo a quella che è la caratteristica principale della fonte storica: l'accessibilità e la verificabilità". Parimenti, poco dopo, Pistacchi sottolineava che "le parole di Revelli ci ricordano che in Italia ancora esistono problematiche irrisolte legate alla mancanza di un catalogo collettivo, di modalità e di metodologie standardizzate per la conservazione dei supporti, per i riversamenti, le trascrizioni in digitale di raccolte analogiche: di fatto in Italia non esiste alcuna banca dati collettiva di riferimento"[5].

E oggi, a circa dieci anni di distanza dal Convegno del 2009, la situazione è migliorata?

---

[1] L'articolo è a cura di Silvia Magrini. Le Appendici sono a cura di Piero Cavallari.

[2] Da cui la pubblicazione *L'intervista strumento di strumento di documentazione. Giornalismo, antropologia, storia* orale. Atti del Convegno, Roma 5-7 maggio 1986. Roma: Quaderno della rassegna degli Archivi di Stato (1987), 53.

[3] L'Istituto, nel subentrare alla Discoteca, ne ha acquisito le competenze, il personale, le risorse finanziarie e strumentali, le attrezzature e il materiale tecnico e documentario" (D.P.R. 233 del 26 novembre 2007 regolamentato dal D.M. 7 ottobre 2008).

[4] Pistacchi, M. (2010). *Vive voci. L'intervista come fonte di documentazione*. Roma: Saggi. Storia e scienze sociali, 7-13: 12.

[5] *Ibidem*, p. 12.

Non molto. Molte esperienze positive sono state condotte e hanno preso piede, ma ancora non hanno fatto 'sistema'. Nelle linee guida che Silvia Calamai ha predisposto per stimolare e orientare il dibattito di questa tavola rotonda ritorna quindi impietoso il tema delle difficoltà tuttora esistenti relative alla conservazione delle risorse orali in Italia.

Facendo riferimento agli esiti del censimento condotto da Vincenzo Galatà e da lei stessa nel 2018 e consultabile online[6], Calamai rileva che "Meno della metà delle risorse orali è accessibile. I problemi etici e legali sono affrontati dal 46% degli studiosi. Oltre il 50% delle persone che lavora con gli archivi sonori non ha una posizione a tempo indeterminato presso istituzioni di ricerca / università. In altri e più spicci termini si può concludere che nella maggior parte dei casi, le risorse orali si trovano dentro le case di chi le produce o presso archivi di enti di piccola dimensione, poco finanziati e poco attrezzati per rispondere alle necessità di conservazione e di accesso di questo tipo di fonti".

Questa tavola rotonda ha avuto dunque il compito di ripensare e progettare quello che potrebbe essere il panorama futuro della conservazione di tale specifica tipologia di materiale: alcuni quesiti circa il 'dove' e 'come' conservare e gestire in maniera sistematica le risorse della storia orale reclamano una risposta da più di un trentennio a questa parte.

Sempre per seguire il *vademecum* predisposto da Calamai, si proverà a rispondere ad alcune delle domande poste:

1. Esistono in Italia delle istituzioni attrezzate per accogliere questi archivi? È preferibile puntare su un archivio unico nazionale o su una rete di archivi locali o tematici? In questo secondo caso, come far sì che possano lavorare insieme, in maniera coordinata, seguendo procedure condivise, condividendone i rispettivi cataloghi?
2. È possibile creare una rete tra associazioni, istituzioni, enti di ricerca che si prenda cura di questo, trasversale a molte e diverse discipline?
3. Quali indicazioni di minima è possibile dare a queste istituzioni che in gran parte esistono, spesso attive su base locale?

In primo luogo, è opinione di chi scrive che non esista in Italia allo stato attuale una singola istituzione attrezzata per accogliere tutti questi archivi.

Se il progetto relativo alla istituzione della Discoteca di Stato-Museo dell'Audiovisivo presso il Palazzo della Civiltà Italiana nel quartiere EUR a Roma si fosse concretizzato, si sarebbero potute creare – forse – in quella sede le premesse per la costituzione di un archivio di tal sorta. Ai sensi della legge 237 del 12 luglio 1999 (art. 1 comma 3) venne difatti istituito il Museo dell'Audiovisivo, a proposito del quale Pistacchi scriveva nel 2006: "La realizzazione del progetto porterà alla costituzione di un polo di attrazione e di riferimento per la documentazione audiovisiva, nonché bibliografica, di settore e multimediale off- e on-line, capace di abbracciare tutto il complesso della comunicazione contemporanea. Obiettivo primario è di assicurare, nel tempo, il valore della testimonianza storica dei beni culturali nella loro interezza – nella forma scritta, su supporto multimediale, sonoro e audiovisivo

[6] https://www.clarin.eu/sites/default/files/CLARIN2018_Poster-3_Paper-12_Galata-Calamai.pdf.

– con il contestuale recupero architettonico e ripristino della destinazione originaria di un edificio (individuato nel Palazzo della Civiltà Italiana) che, negli anni, ha subito un degrado strutturale ed estetico"[7].

Il dettato normativo, tuttavia, è rimasto lettera morta per una serie contingente di motivazioni politico istituzionali. Oggi, vent'anni dopo, se ne può solo prendere atto e constatare che in effetti allo stato attuale è di gran lunga più sostenibile immaginare un sistema reticolato di archivi e di istituzioni operanti nel settore in maniera coordinata.

L'esperienza mostra che, per molti aspetti, forme di collaborazione esistono già, sia tra alcune realtà del territorio tra di loro sia tra alcune realtà locali e l'ICBSA: in entrambi i casi tali cooperazioni sono tutt'altro che di livello embrionale come si vedrà più sotto.

La questione, se mai, è l'individuazione delle modalità di coordinamento di siffatta rete di collaborazioni. Ragioni storiche e giuridiche, non di mera partigianeria, inducono chi scrive a ritenere che un ruolo di primo piano in tale coordinamento debba essere giocato dall'ICBSA.

È questo un onere/onore che deriva dalla storia dell'Istituto e dallo stesso apparato normativo italiano. Il 10 agosto del 1928 infatti nasceva l'Istituto "ritenuta la necessità assoluta e urgente di disciplinare e sviluppare mediante l'istituzione di una Discoteca di Stato la raccolta e la diffusione di dischi fonografici riproducenti la voce dei cittadini italiani benemeriti della Patria" e non molti anni dopo, con la legge 130 del 18 gennaio 1934, l'ambito di interesse e delle competenze della Discoteca si allargarono, comprendendo la cultura scientifica, artistica e letteraria compresi canti e dialetti d'Italia e gli studi di glottologia e storia.

È da allora che la Discoteca di Stato ha avviato la sua attività in questo ambito ottenendo alcuni dei più grandi risultati negli anni Cinquanta e Ottanta del secolo scorso in occasione delle campagne di registrazione sul campo e attività di catalogazione che sono sfociate nell'Archivio Etnico Linguistico-Musicale (AELM). Per una panoramica sulla nascita dell'Istituto e il contesto internazionale in cui esso ha mosso i suoi primi passi si veda l'*Appendice 1*.

L'ICBSA costituisce di fatto la più grande collezione pubblica di materiale sonoro e audiovisivo nazionale, implementato dall'istituto del Deposito legale ai sensi della L. 106 del 15 aprile 2004.

A proposito delle funzioni e del ruolo dell'Istituto è importante, anche, sottolineare come l'ICBSA sia in procinto di avviare una profonda riorganizzazione della filiera di lavorazione dei supporti (sonori e audiovisivi insieme) per una sua ottimizzazione e ripensamento ai sensi delle nuove normative vigenti per la conservazione e gestione dei dati (misure di sicurezza ICT emanate dall'AgID e GDPR) e quindi come sia prevista in questo contesto una profonda revisione e riflessione sulle attività in corso.

[7] Pistacchi, M., Aquilanti, F. & Baldi, F. (2006). Verba Manent. Teoria e prassi della digitalizzazione dei documenti sonori e vide della Discoteca di Stato-Museo dell'audiovisivo (prima parte). In *Digitalia*, 2, 131-148: 133.

La varietà del materiale che l'Istituto conserva e delle attività di conservazione, studio e valorizzazione che ha avviato negli anni sono tali che l'ICBSA ha potuto maturare nei decenni una esperienza significativa, riconosciutale sia a livello nazionale che internazionale.

Tale *expertise* si è tradotta nel tempo nella formulazione di standard e di linee guida nelle materie di competenza con particolare riferimento alle tematiche della digitalizzazione dei supporti analogici, della conservazione a lungo termine dei dati digitali e dei supporti originali e della catalogazione. Tutto ciò naturalmente è stato realizzato spesso in collaborazione e dietro sollecitazione di altre realtà del campo: istituzioni ministeriali 'sorelle', *in primis* l'Istituto Centrale per il Catalogo Unico e le Informazioni bibliografiche e la Direzione Generale Archivi, le Università (il Centro di Calcolo della Normale di Pisa, Facoltà di Ingegneria di Padova, l'Orientale di Napoli) enti di ricerca nazionali (CNR, sede di Roma) e internazionali (*Phonogrammarchiv* di Vienna, Fonoteca Nazionale Svizzera di Lugano) associazioni pubbliche e private tra le più varie (a mero titolo di esempio RAI Teche, USC Shoah Foundation).

In questo ambito, il ruolo giocato dall'ICBSA all'interno del Servizio Bibliotecario nazionale (SBN) ha indubbiamente particolare rilievo. ICBSA è entrato in SBN nel 2015 e di fatto è responsabile, coordina e gestisce in prima persona dallo stesso anno il Polo DDS (Discoteca di Stato): si tratta di un polo che, ad oggi (aprile 2019), riunisce 15 biblioteche di istituti del MiBAC, del Ministero per l'Istruzione Università e Ricerca o indipendenti che contribuiscono con i loro cataloghi al catalogo pubblico online del SBN nazionale. Tali istituti condividono gli standard fissati dall'ICCU per SBN e in particolare quelli proposti ed elaborati da ICBSA con ICCU stesso e accolti da ICCU a livello nazionale per il materiale librario, audio e audiovisivo.

Un esempio assai significativo a tale proposito è rappresentato dall'ingresso nell'Opac di SBN nel contesto del Polo DDS della biblioteca del Museo della Liberazione di Via Tasso, Roma (si veda *Appendice 2*).

L'ampliamento e la sistematizzazione di questa rete di collaborazioni già in atto, spesso nate sull'onda di singole iniziative di questo o di quel genere, è dunque un obiettivo prioritario in questo momento alla luce anche della nuova sensibilità e coscienza che si sta formando.

Un importante banco di prova del nuovo approccio potrebbe essere rappresentato dall'attività del costituendo gruppo di lavoro per la redazione di un vademecum per la conservazione delle fonti sonore con la partecipazione di rappresentanti della Soprintendenza Archivistica e Bibliografica della Toscana, dell'ISTORETO, dell'Università di Padova (Ingegneria del suono), dell'AISV, dell'AISO, di CLARIN-IT oltre che di ICCU ed ICBSA. Il gruppo lavorerebbe in sinergia con altri tavoli costituitisi per il censimento e la salvaguardia di archivi sonori facenti capo all'Istituto Parri di Bologna e alla rete archivi della CGIL e si prevede di presentare, discutere e 'suggellare' le conclusioni del lavoro svolto in occasione del convegno che ICBSA intende organizzare per la fine del 2019/inizi 2020 sul tema dell'uso e della gestione delle fonti orali.

L'impegno dell'ICBSA nei confronti delle fonti della storia orale comunque non si esaurisce sul fronte della catalogazione o della conservazione. Tutt'altro, come si potrà evincere dalla descrizione che segue di alcuni, fra i tanti, progetti in corso.

È degli ultimi anni ad esempio un grosso impegno nell'ambito educativo e specificatamente nel contesto dell'Alternanza Scuola Lavoro (ASL). È stato infatti avviato un progetto sperimentale ed assai innovativo denominato "laboratorio creativo suoni e immagini in movimento" riguardo alla alfabetizzazione sull'uso di strumenti di riproduzione audiovisiva volta al fine di sensibilizzare gli studenti sul panorama complesso dei media onde renderli soggetti attivi di una visione della realtà. Uno degli audiovisivi realizzati (dalla durata di circa 60 minuti) si intitola per l'appunto *L'intervista* ed ha permesso ai ragazzi coinvolti di comprendere sul campo cosa comporti entrare in relazione con un'altra persona su un tema 'caldo', intervistandola sul significato e il peso nella propria vita dell'esperienza lavorativa. Essi hanno imparato vedendo e ascoltando l'altra persona a vedere e ascoltare sé stessi, stimolando così un processo significativo di riflessione. La qualità personale dell'intervista di storia orale[8], in tutte le sue fasi di preparazione, realizzazione, documentazione e analisi, è stata per molti di loro una molla che ha fatto scattare un meccanismo di presa di coscienza non certo banale per un adolescente.

Per l'ICBSA, quindi, assume una valenza importante per la promozione e diffusione della conoscenza del patrimonio audiovisivo nazionale e per la realizzazione di attività inerenti l'utilizzo degli audiovisivi e delle apparecchiature di riproduzione sonora – entrambi mission istituzionali per altro – lavorare insieme al mondo della scuola per creare il fondamento di una società alfabetizzata secondo i meccanismi del linguaggio sonoro e delle immagini in movimento.

Per questo motivo, tra l'altro, l'Istituto aderisce dall'aprile 2019 al network italiano DiCultHer, la Scuola a rete "Digital Cultural Heritage, Arts and Humanities" nato nel febbraio 2015 per costruire e consolidare anche nella scuola le competenze digitali, tenendo ben presente però che la conoscenza del mondo digitale va intesa non solo quale conoscenza di un complesso di tecniche strumentali, ma anche del processo che permette di codificare e comunicare la realtà attraverso una grammatica e una sintassi logica ed emotiva che si può meglio comprendere grazie al confronto con la storia, dall'analogico al digitale, dei mezzi di riproduzione sonora ed audiovisiva.

Ancora, l'ICBSA è impegnato nella valorizzazione del suo patrimonio etnolinguistico e musicale promuovendone la conoscenza anche tra un pubblico di non specialisti, nel tentativo di stimolare una visione olistica del patrimonio culturale italiano. Grazie alla collaborazione con il Segretariato Regionale del MiBAC per l'Emilia-Romagna è stato infatti possibile realizzare all'interno della piattaforma Tourer.it un percorso georeferenziato e interattivo liberamente accessibile in inter-

[8] Sulla connotazione interpersonale dell'intervista vedi: Portelli, A. (2016). L'uso dell'intervista nella storia orale. In *Vive voci. L'intervista come fonte di documentazione*, op. cit., 3-12.

net (www.tourer.it – sezione itinerari) che porta alla scoperta delle tracce sonore in Emilia-Romagna. Si tratta di un itinerario virtuale che permette di ascoltare voci e canti che hanno animato i tanti beni architettonici e naturalistici della regione. Sono state collocate sulla mappa georeferenziata alcune significative testimonianze sonore registrate fra gli anni Cinquanta e Settanta del secolo scorso: un patrimonio che raccoglie racconti, indovinelli, filastrocche, ninnenanne e canti che ci restituisce integralmente le voci del territorio e dei suoi beni[9].

Da ultimo, a titolo di esempio, nel tentativo di riprendere la tradizione delle campagne di registrazione sul campo già promosse in passato l'Istituto sta ora avviando collaborazioni con il Museo dell'Emigrazione Italiana di Genova per l'acquisizione di voci del nostro tempo. Si tratta di un progetto che potrebbe partire nei prossimi anni e diventare di grande interesse per documentare la complessità del nostro contemporaneo.

Appendice 1. *Archivi di voci contenenti fonti di ricerca e di documentazione.*

Austria. Il più antico archivio sonoro è europeo, il *Phonogrammarchiv* di Vienna, fondato nel 1899 da alcuni membri della Imperiale Accademia austriaca delle Scienze: il nucleo storico delle collezioni si è costituito dalle campagne di registrazione sul campo avviate nei primi anni del Novecento, tra il 1901 e il 1903, in Croazia, Brasile, nell'isola di Lesbo.

Oltre a questo, rilevanti sono le registrazioni di decine di "Stimmporträt", ritratti vocali.

Le collezioni storiche di questo archivio nel 1999 sono state incluse nel programma dell'Unesco Memoria del mondo, per preservare e riscoprire l'eredità sopravvissuta (perché le collezioni e la cultura, come gli uomini – lo vediamo quotidianamente – sono esposti a sofferenza e distruzione).

I supporti commerciali saranno poi raccolti dalla Mediateca austriaca (*ÖsterreichischeMediathek*). Il *Phonogrammarchiv* di Vienna, inserito nell'Accademia delle scienze austriaca (PhA) è stato fondato il 27 aprile 1899 ed è come segnalato il più antico archivio sonoro del mondo. E proprio grazie alle attività del *Phonogrammarchiv*, che abbiamo fonti sonore relative anche alle prime testimonianze di musica/cultura popolare italiana. Dall'Opac ICBSA:

Livello bibliografico:	spoglio
Tipo di risorsa:	registrazione musicale
Autore:	Schüller, Dietrich
Titolo:	O tu stelle
Note:	Canto dei militari friulani. Cantato da Antonio Melinz, nato nel 1883. Registrazione Ph2583.
Formato distribuzione:	disco

[9] Hanno collaborato al progetto oltre a Ilaria Di Cocco (Segretariato Regionale Emilia-Romagna), Francesco Baldi e Bruno Quaresima (ICBSA) anche Giuseppina Colicci, Michele Melidoni e Giulio Tosti, attualmente impegnati nella catalogazione del materiale dell'AELM conservato presso ICBSA.

Velocità:	1.4 m. al secondo (compact disc)
Fa parte di:	Sound Documents from the Phonogrammarchiv of the Austrian Academy of Sciences. The Complete Historical Collections 1899-1950. Series 4: Soldatenlieder der k.u.k. Armee / General Editor: Dietrich Schüller; re-recording: Franz Lechleitner.
Nomi:	Hajek, Leo (autore dell'inchiesta)
Soggetti:	Friuli-Venezia Giulia
Lingua di pubblicazione:	Tedesco
Codice del documento:	DDS0221256

Germania. All'archivio di Vienna si aggiunse subito quello di Berlino, nel 1900, sorto anch'esso in ambito scientifico (un professore di psicologia e un fisico che iniziarono a registrare su cilindri di Edison un gruppo musicale che arrivò in Germania dalla Thailandia, allora Siam).

Francia. In Francia le origini di quello che è oggi il dipartimento dell'audiovisivo risalgono al 1911 quando nascono gli Archives de la Parole, creati dal linguista Ferdinand Brunot presso l'Institut de Phonétique della Sorbonne; nel 1928 nasce il *Musée de la parole et du geste* ribattezzato poi nel 1938 *Phonothèque Nationale*: questa nel 1977 diviene *Département de la Phonothèque et de l'Audiovisuel* della *Bibliothèque Nationale* e nel 1994, *Département de l'Audiovisuel*, come si chiama tuttora. Detiene il deposito legale di documentazione multimediale, documenti elettronici, etc.: archivi di parole, quindi, oltre che di suoni, archivi di ricerca e di documentazione.

Stati Uniti. Nel 1908 l'Università di Harvard aveva affidato a John Lomax – il fondatore della etnomusicologia – una ricerca sul campo in vista della costituzione di un Archivio del canto popolare americano ospitato a Washington dalla Library of Congress (diecimila reperti sonori). La Columbia University inaugurerà nel 1936 un Archivio del folklore e della musica primitiva, poi trasferito alla Indiana University dove sarebbe confluito in parte il *Phonogrammarchiv* di Berlino.

Italia. La Discoteca di Stato nasce nel 1928, a dieci anni dalla fine della Prima guerra mondiale. Una data non casuale: il regime fascista, che fece la sua fortuna anche con l'eredità della guerra, insieme con l'Anmig (Associazione nazionale mutilati e invalidi di guerra), aveva acquistato il fondo di incisioni frutto della intuizione del privato cittadino, artista teatrale, futurista, autore, cantante Rodolfo De Angelis, che incise nel 1924 le voci dello stato maggiore dell'esercito italiano all'epoca della Grande Guerra. Si istituisce così un archivio per acquisire le voci dei "grandi" della patria mettendolo alle dipendenze del capo dello stato. La collocazione nel tempo dell'Istituto è sintomatica: alle dipendenze del capo dello stato prima, poi a quelle del Ministero della Cultura Popolare nel 1937, al momento della istituzione, poi nel 1944, quando il Ministero viene soppresso, l'archivio passa alla Presidenza del

Consiglio dei Ministri e infine, nel 1974, al neonato Ministero dei Beni Culturali, ma permanendovi a lungo come divisione amministrativa, diretta da un funzionario.

L'eredità della Discoteca di Stato, un istituto allo stesso tempo biblioteca, archivio, museo, con le sue collezioni, edite e inedite, è confluita nell'ICBSA che nasce nel 2007 (l'anno dopo la pubblicazione del Regolamento delle norme sul deposito legale dei documenti di interesse culturale).

Il fondo di documentazione storica della Discoteca di Stato, rappresentato come nel caso del *Phonogrammarchiv* di Vienna in una sorta di ritratti sonori, è stato spesso fonte di ricerca, anche linguistica.

Alcuni anni fa la professoressa Anna Antonini nell'illustrare il progetto della Accademia della Crusca *I mutamenti del parlato attraverso un secolo di registrazioni*, affermava:

> [...] un progetto dell'Accademia della Crusca [...] con la collaborazione di varie istituzioni: la stessa Accademia, la Discoteca di Stato che ha fornito in particolar modo i primi materiali sonori, [...] che ha dato al lavoro un trattamento televisivo poiché lo ha ridotto in una trasmissione in quattro puntate andata in onda in concomitanza con la quarta "Settimana della lingua italiana nel mondo" (2004).
> Lo scopo del lavoro è stato quello di documentare, attraverso la registrazione di una serie di voci autentiche, l'evoluzione della lingua italiana, cioè come noi italiani abbiamo via via imparato a padroneggiare, soprattutto parlando, la nostra lingua; una lingua che abbiamo ereditato come lingua di cultura scritta e quindi di registro sostenuto. L'intero repertorio di voci copre un secolo e si configura come una colonna sonora del Novecento appena passato, un viaggio nella lingua italiana e nella sua voce a partire dalla prima registrazione disponibile, la voce di Giuseppe Giacosa (1902).
> Per misurare la lunghezza del cammino percorso e illustrare le tappe attraverso le quali la divaricazione dal parlato è stata a poco a poco moderata da un graduale assorbimento della lingua corrente, sono state scelte voci di personaggi di spicco, per posizione sociale, grado di istruzione, ruolo nella società; e voci magari anonime, ma che hanno fatto ugualmente storia, perché legate ad un annuncio, o anche solo alla funzione di quell'agente nuovo nella comunicazione umana che è l'"emittente'.
> In ogni caso, voci autorevoli o 'autorizzate', quelle che ci permettono di misurare il cammino dell'italiano parlato restando sullo stesso piano di responsabilità e di ufficialità; ma pur sempre voci in originale dei principali protagonisti del '900 che sintetizzano il secolo dell'italiano parlato, perché il Novecento è stato proprio il secolo della progressiva generalizzazione della capacità di parlare l'italiano.

Tenendo ben presente il vademecum elaborato per il convegno AISV, soprattutto per quella sorta di appello rivolto alle istituzioni operanti nell'ambito della conservazione e valorizzazione dei beni culturali sonori, si riporta di seguito un breve excursus storico legislativo della Discoteca di Stato che potrebbe essere utile a rispondere appunto al *grido d'aiuto* lanciato dagli organizzatori del convegno di Arezzo:

> *Regio decreto-legge 10 Agosto 1928, N. 2223 - Istituzione di una Discoteca di Stato in Roma*
>
> Visto l'art. 3, n. 2, della legge 31 gennaio 1926, n. 100; Ritenuta la necessità assoluta ed urgente di disciplinare e sviluppare, mediante la istituzione di una Discoteca di Stato, la raccolta e diffusione di dischi fonografici riproducenti la voce di citta-

> dini italiani benemeriti della Patria; Sentito il Consiglio dei Ministri; Sulla proposta del Capo del Governo, Primo Ministro Segretario di Stato, di concerto coi Ministri Segretari di Stato per le finanze, istituzione pubblica ed economia nazionale; Abbiamo decretato e decretiamo: *art. 1*: È istituita in Roma la Discoteca di Stato allo scopo di raccogliere e conservare per le future generazioni la viva voce dei cittadini italiani, che in tutti i campi abbiano illustrata la Patria e se ne siano resi benemeriti [...].

Ma, a parte la legge istitutiva della Discoteca, nel 1934 fu varato un provvedimento veramente interessante, anche pensando agli interrogativi sollevati ad Arezzo:

> *Legge 18 Gennaio 1934, N. 130 – Estensione dell'attività della Discoteca di Stato anche a manifestazioni interessanti la cultura nazionale, scientifica, letteraria e le tradizioni ed i costumi del Paese*
>
> *Art. 1*: La Discoteca di Stato, oltre ai fini previsti dall'art. 1 del R. decreto-legge 10 agosto 1928, n. 2223, convertito nella legge 3 gennaio 1929, n. 81, ha per iscopo di raccogliere ed ordinare sistematicamente, registrandolo in matrici, in dischi e con qualsiasi altro mezzo meccanico, tutto quanto nel campo dei suoni interessi la cultura scientifica, artistica e letteraria, ed in modo particolare: a) i canti e i dialetti di tutte le regioni e le colonie d'Italia; b) le interpretazioni "definitive" delle opere principali dei maggiori compositori e poeti viventi; c) ciò che possa interessare gli studi di glottologia, di zoologia, di fisiologia, di storia, ecc. [...].

Ecco quindi che già dal 1934 inizia per l'Istituto l'acquisizione "legale" dei documenti sonori, ma sempre su questo argomento vi sono stati successivamente altre tappe importanti:

> *Regio decreto legge 2 febbraio 1939*
>
> *Art. 6*: Ai fini di cui al n. 2 dell'art. 2. tutti gli editori fonografici e fonomeccanici italiani o rappresentati in Italia debbono inviare alla Discoteca di Stato, in duplice esemplare ed in porto affrancato, copia di quelle loro pubblicazioni discografiche che sono richieste dal Ministero della cultura popolare. Gli stessi editori predetti debbono anche rimettere alla Discoteca di Stato, oltre a tutte le loro pubblicazioni tipografiche, un elenco mensile di tutti i dischi editi, nel quale debbono essere riportati tutti i dati delle etichette apposte sui dischi stessi. II Ministero della cultura popolare può inoltre richiedere agli editori italiani quei dischi prodotti anteriormente alla presente legge, che la Discoteca, ritiene di conservare per i propri fini e gli editori stessi sono obbligati a fornirli in duplice copia gratuitamente.

Infine, per tornare all'attualità, la nascita dell'ICBSA:

> *Decreto 7 ottobre 2008 – Istituzione dell'Istituto centrale per i beni sonori e audiovisivi; art. 2: Compiti Istituzionali*
>
> 1. L'ICBSA svolge, con valenza sull'intero territorio nazionale, attività di documentazione, valorizzazione, restauro, conservazione, ricerca e consulenza sui documenti sonori e audio visivi appartenenti allo Stato e ad altri Enti pubblici, anche non territoriali, nonché alle persone giuridiche private senza fine di lucro.

> […] 3. L'ICBSA: […] b) svolge attività finalizzate alla collaborazione e alla partecipazione a programmi europei per la conservazione, la valorizzazione e la fruibilità del patrimonio audiovisivo; c) promuove e coordina iniziative nel settore della catalogazione, della valorizzazione, della fruizione e della conservazione della documentazione audiovisiva in collaborazione con organismi di ricerca nazionali e internazionali; d) raccoglie, conserva e rende fruibile la memoria delle tradizioni popolari nonché della società, della cultura e della lingua italiana anche attraverso l'acquisizione di registrazioni audio-visive […].

E la riforma del deposito legale:

> *A seguito dell'emanazione del regolamento di attuazione, con il D.P.R. 3 maggio 2006, n. 252 è entrata in vigore, dal 2 settembre, la nuova Legge sul deposito legale in Italia (L.106/2004).*
>
> Legge e relativo regolamento prevedono la consegna alla Discoteca di Stato e Museo dell'Audiovisivo (ora Istituto Centrale per i Beni Sonori ed Audiovisivi), entro i 60 giorni successivi alla prima distribuzione, di una copia dei "documenti sonori e video prodotti totalmente o parzialmente in Italia o offerti in vendita o distribuiti su licenza per il mercato italiano e comunque non diffusi in ambito privato".

Il principale obiettivo del deposito legale è la costituzione dell'Archivio nazionale e regionale della produzione editoriale e gli istituti depositari hanno l'obbligo di raccogliere, conservare e catalogare tutti i documenti depositati.

Il regolamento attuativo della legge ha individuato la Discoteca di Stato Museo dell'Audiovisivo (poi ICBSA) quale soggetto depositario di una copia di tutti i documenti sonori e video prodotti totalmente o parzialmente in Italia o distribuiti su licenza per il mercato italiano.

Un'ulteriore copia deve essere depositata, per l'archivio regionale, presso l'istituto dove ha sede il soggetto obbligato al deposito ai fini della costituzione degli archivi regionali. Ciascuna regione e ogni provincia autonoma, previa consultazione con le associazioni degli enti locali e con gli istituti interessati, entro nove mesi dalla data di entrata in vigore del regolamento, propone alla Conferenza Unificata l'elenco degli istituti depositari delle ulteriori copie dei documenti oggetto del deposito.

Ma una malcelata difficoltà delle regioni a ricevere la copia per costituire l'archivio regionale è riconducibile all'impegno di risorse umane, alla mancanza di spazi dove collocare il materiale e soprattutto alla tutela, visto che il bene deve essere anche valorizzato. Tale preoccupazione deriva anche dalla lettura dell'art. 1 che al secondo comma precisa che "*dalla predetta disposizione non devono derivare nuovi o maggiori oneri a carico della finanza pubblica*"; ma appare alquanto difficile pensare di raccogliere, conservare nel tempo e rendere fruibile tutta la produzione editoriale pubblicata e distribuita in Italia senza avere a disposizione adeguate risorse economiche per garantirne la corretta gestione.

Appendice 2. *L'ICBSA e le sue collaborazioni. Esempi concreti.*

Come una sorta di *cahier de doléances*, tra i punti interrogativi focalizzati dal convegno vi sono:

– Come favorire la conservazione, la descrizione, il riutilizzo di risorse orali raccolte in ambito di progetti di ricerca passati e presenti?
– Quali metadati usare? È possibile proporre un set di minima e uno di massima?
– È possibile creare una rete tra associazioni, istituzioni, enti di ricerca che si prenda cura di questo problema, trasversale a molte e diverse discipline?

Voglio qui riportare un esempio che potrebbe in parte rispondere a questi interrogativi. La quasi decennale collaborazione tra l'ICBSA e il Museo Storico della Liberazione di Via Tasso ha interessato anche la sistemazione della mediateca del Museo (dedicata a Gerardo De Angelis, ucciso alle Fosse Ardeatine il 24 marzo 1944, autore cinematografico, Medaglia d'argento al valor militare alla memoria).

Nella fattispecie, l'accordo si è focalizzato sull'attivazione in prima istanza di procedure di catalogazione, conservazione e digitalizzazione dei documenti sonori, audiovisivi e multimediali posseduti dal Museo, circa un migliaio di materiali tra quelli analogici (audiocassette, VHS) e digitali (CD, DVD, CD-rom). Una volta svolto questo primo e fondamentale flusso di lavoro, tutti i metadati prodotti (di catalogazione, tecnici, amministrativi) sono migrati nel giovane polo SBN DDS/IC. Di seguito, un esempio di scheda relative a quanto descritto.

Livello bibliografico:	monografia
Tipo di risorsa:	film
Autore:	Fiorentini, Mario (1918)
Titolo:	Ottobre Nero 1943, Via Tasso - 23 Marzo 1944: intervista a Mario Fiorentini
Descrizione fisica:	1 DVD (43 min., 30 sec.). Azione dal vivo, colore, suono sul supporto, Stereofonico, Digitale, 625 PAL
Note:	durata: 43 min., 30 sec.
Abstract:	La grande razzia nel vecchio ghetto di Roma. Il 16 ottobre 1943 le SS deportano 1024 persone e oltre 200 bambini al campo di concentramento di Auschwitz.
Tipo video:	videoregistrazione
Tecnica:	azione dal vivo
Forma di pubblicazione:	videodisco
Formato di presentazione:	DVD-video
Nomi:	Fiorentini, Mario (1918) [Intervistato]
Soggetti:	Italia, Roma, ebrei, 1940-1949, nazismo, deportazione, storia contemporanea, storia orale, interviste, nazismo, deportazione, resistenza, Roma
Supporto:	dvd
Paese di pubblicazione:	Italia
Lingua di pubblicazione:	Italiano
Codice del documento:	DDS1714684

PARTE II

SCIENZE DEL PARLATO

DOI: 10.17469/O2106AISV000012

FABIO ARDOLINO

Adapting the Implicit Association Test to speech: Issues and methodological proposals

This contribution aims to present and to discuss how speech sciences can fruitfully apply the IAT (Implicit Association Test) protocol, an experimental procedure developed in the psycho-sociological fields designed to reveal and to measure implicit attitudes toward social objects (e.g., speech varieties). Currently, few studies employ IAT to evaluate speech attitudes applying different protocols, especially in the production and treatment of stimuli. This proposal intends to outline a framework of IAT use in spoken language study and to discuss a number of features potentially capable of constituting a valid and reproducible standard for IAT use in speech: stimuli duration, content and familiarity.

Key words: sociolinguistics, attitudes, Implicit Association Test.

1. *Preliminary Remarks*

1.1 Attitudes in linguistics

The evaluation of attitudes in linguistics is supported by relevant literature (Cooper, Fishman, 1974; Giles, Billings, 2004), and individual attitudinal states prove to have a key role in many phenomenological aspects of speech. Interlinked as they are with different behavioral, psychological and social aspects, attitudes seem to play a crucial part in the regulation of socially-motivated sound change (Labov, 1963), in speakers' social networks constitution, cohesion (Labov, 1966) and hierarchization (Ryan, Bülik, 1982), as well as in Second Language Acquisition (Anisfeld, Lambert, 1961; Lambert, Gardner, Olton & Tunstall, 1968; Mariani, 2010), interlingual intelligibility (Wolff, 1959) and correctness/pleasantness judgment (Fridland, Bartlet, 2006). As this non-exhaustive review shows, attitudes can be involved in both active and passive linguistic processes, and it is important to highlight the two-ways connections they establish with other factors that prove significant to linguistics, like motivation, grouping, self-esteem and so on.

Circumscribing the notion of attitude is often complex. If it is true – and it is – that attitudes represent some "state[s] of readiness [...] exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related" (Allport, 1935: 798), it is also true that attitudes themselves are often manifested as behaviors, and therefore they appear to be dynamical rather than fixed states. Moreover, attitudes seem to vary considerably across time and situations (McGuire, Lindzey & Aronson, 1985): these features impose the utmost rigor during their experimental elicitation.

However, in very few cases linguistics research addresses the substantial difference between explicit (mainly conscious) and implicit (mainly unconscious) attitudes; a difference considered instead crucial within psychological and psycho-social sciences. Explicit and implicit attitudes show, in fact, profound differences in their characteristics and in their modes of action: firstly, explicit attitudes exist on the conscious level, and they are deliberately constituted and often overtly accepted by the social group; on the contrary, implicit attitudes are mostly unconscious: their formation is involuntary, and they frequently represent potential objects of stigmatization (cfr. Bassili, Brown, 2005). Moreover, combined enquiries reveal a general and substantial discrepancy between explicit and implicit attitudinal constructs in one and the same person (Dunham, Baron & Banaji, 2008). Nevertheless, implicit attitudes, as well as explicit, are formed as a result of punctual or lasting experiences that "mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald, Banaji, 1995). This definition embraces individual experiences (traumas, familiar issues and so on) as well as forms of "social experience" like, for example, the integration in a covertly xenophobic community or the ingroup constitutions mechanics (Tajfel, 1974). In this sense, the study of attitudes is often able to reveal existing sociological processes and trends, in addition to individuals' psychological states.

1.2 Eliciting Implicit Attitudes: the IAT protocol

On an experimental level, implicit attitudes are considerably harder to elicit and to quantify, while explicit attitudes are readily revealed by tested subjects and relatively easy to measure. Nevertheless, as suggested in the previous paragraph, to outline a comprehensive picture of the attitudinal construct in a subject or in a group, it is necessary to collect data referring to implicit attitudes. It is not surprising, therefore, that the Implicit Association Test protocol (IAT, Greenwald, McGhee & Schwartz, 1998) generates particular interest among researchers, given its experimental validity and its considerable design malleability.

The IAT is a widely-used cognitive-behavioral paradigm that measures the strength of automatic (i.e., implicit) associations between concepts in people's memory; during the test, the participant is asked to quickly classify a stimulus visualized on a computer screen. Stimuli are generally provided in the form of written words or images and can be classified in four categories: two of those categories represent concepts (e.g., normal weight and obese[1] people), the other two represent polar attributes (e.g., good and bad); the classification task consists in associating each stimulus with its reference category. A primary characteristic of the IAT is to equip the participant with only two distinct keys to operate the classification; as a consequence, two categories (a concept and an attribute) are associated to the same key.

The protocol is intended to create an overlap of elaborations that highlights the existence of an unconscious association between an attribute (e.g., bad) and a target (e.g., obese people). This results in a faster classification of the stimuli when pre-

[1] See, for exemple, the Project Implicit's "Fat/Thin IAT", https://implicit.harvard.edu/implicit/selectatest.html.

sented one after another. It is clear that IAT design can be laid out after a specific hypothesis definition or, in other words, in the assumption of the existence of a pairing between an attribute and a concept category (e.g., obese people and bad).

An IAT test is generally composed by a variable number of successive sets of stimuli called blocks; in this sense, initial blocks serve as practice and provide the subject's latency tare. For each block, the attribute-categories associated to keys are inverted: the same key is associated with the hypothesis-consistent pair in one block, and with the hypothesis-inconsistent in the following. The overall score of the test (the so-called *D-score*; Greenwald, Nosek & Banaji, 2003) is computed in four steps: a) the mean of latency of hypothesis-inconsistent practice blocks is subtracted from the mean of latency of hypothesis-consistent practice-blocks; b) the difference is divided by the pooled standard deviations of latency in both practice-blocks; c) the operations a) and b) are repeated for the experimental-blocks; finally, d) the average of the practice-blocks and experimental-blocks is calculated as *D*.

In general, the higher the *D* the stronger the association between the 'hypothesis-consistent' pairings. Conversely, low and negative *D-scores* suggest a stronger association between the 'hypothesis-inconsistent' pairings. IAT protocols and their derivatives have been extensively applied to investigate implicit phenomena like the in-group favoritism in racial prejudice development (Baron, Banaji, 2006), the incidence of sexual preferences in gender prejudice (Anselmi, Vianello, Voci & Robusto, 2013), but also the role of self-esteem in social anxiety (Egloff, Schmukle, 2002) and the insurgence of suicidal tendencies (Nock, Park, Finn, Deliberto, Dour & Banaji, 2010).

To complete the picture, it is useful to recall that the use of IAT, also in virtue of its quick and extensive spread, has raised several doubts and criticisms both in public and academic worlds. In particular, some scholars invite to a deeper reflection about the psychological object that is detected by the Implicit Association protocol. As observed by Ottaway, Hayden and Oakes (2001) fluctuations in latency can be motivated also by an influence of stimuli familiarity (see infra), or, as highlighted by Rothermund and Wentura (2004), of stimuli salience. It goes without saying that these issues should be considered in the definition of the test's hypothesis and during the IAT design planning.

1.3 IAT implementation in linguistics

The versatility showed by the IAT in the collection of data for the social prejudicial constructs justifies the interest in employing this protocol for sociolinguistic studies. The possibility of replacing visual with auditive stimuli was firstly considered in Vande Kamp (2002). Campbell-Kibler (2012) clearly proved the potential of IAT in assessing the social meaning of language areal variation (e.g., northern and southern American English). IAT protocols generally highlight a more positive implicit attitude towards in-group language varieties (non-accented vs. accented American English; Pantos, Perkins, 2012; Pantos, 2017; Roessel, Schoel & Zimmerman, 2017 and regional varieties of Dutch; Rosseel, Geeraerts & Speelman, 2018) and towards the historically more prestigious variety (Standard South Afrikaans English vs.

Afrikaans-accented English; Álvarez-Mosquera, 2017; Álvarez-Mosquera, Marìn-Gutierréz, 2018). All these studies employ speech as a stimulus, but with different operational strategies: from short words (e.g., Campbell-Kibler, 2012) to entire sentences (e.g., Álvarez-Mosquera, Marìn-Gutierréz, 2017). Mitchell (2018), uses the IAT to assess attitudes towards artificial speech, whereas the other ones presented studies involving natural human speech.

All these works share a common point: the substitution of the target visual stimuli (e.g., pictures, written words) with auditive linguistic stimuli (speech) in their IAT designs: however, as this quick review clearly shows, arrangements concerning the production of those stimuli are far from a homogeneous standard; moreover, they appear to underestimate some potentially influential issues. A number of these issues will be analyzed and discussed in the following paragraph.

2. *Adapting IAT for Speech: issues and proposals*

The Implicit Association protocol is organized around the participant's reaction to the presented stimuli: it is therefore natural that this aspect of the design requires to exercise caution during the adaptation of the IAT for speech sciences, and thus during the selection of the linguistic auditive stimuli, whereas the classical protocol only provides graphic elements.

As seen in the introductive paragraphs, a major issue regards the duration of stimuli, that can vary significantly from one study to another. In a similar way, the question of the stimuli content appears to be poorly considered in the IAT adaptation for speech, while it receives great attention in the classical IAT designs. The issue of stimuli content recalls the equally crucial element of stimuli familiarity, an attribute that a participant can assign (or not) to a presented element and that can exert its influence on latency times. These three issues will be discussed in detail in this section. Where possible, some practical examples will be provided: such examples are taken from two specific tests developed by the author and currently employed in two ongoing studies. The first test is designed to explore the attitudes showed by Italian teachers towards Chinese-accented Italian speech (Chinese-accented Italian IAT; Calamai, Ardolino, 2020); the second one assesses the attitudinal constructs of Italian migrants in France towards French language (Italian/French IAT; Ardolino, forthcoming). Both protocols have been developed using a specific psychometric software (Implicit 5): to visualize the test structure, Figure 2 shows different phases of the Italian/French IAT. The labels for target and attribute categories are displayed on the top of the screen, while all the relevant information (instructions, graphic stimuli, feedbacks) appears in the center. It is important to highlight that, although short (around 10 minutes), the completion of an IAT can be a stressful activity, because it requires high reactivity and a significant amount of attention; the use of a headset is necessary in order to isolate the subject from environmental noise. For the same reason, the test should be submitted in a quiet room.

Figure 1 - *Screens from the Italian/French IAT (English translation provided in Appendix)*

2.1 Stimuli duration

The duration of stimuli presentation in IAT is crucial, since the scoring algorithm for IAT developed in Greenwald et al. (2003) is based on the latency times showed by subjects in their response and measured from the end of the submission of a stimulus. Therefore, classical IAT's *D* considers the reaction time elicited in relation to a visual stimulus, which remains homogeneous during the prefixed amount of time. On the contrary, the auditive stimulus varies through time, and this feature is potentially able to distort the results of the test. The duration of an IAT stimulus is quantified in 150-750 ms (see Greenwald et al., 1998; Nosek, Greenwald & Banaji, 2005); a longer stimulus affects the automatic nature of responses.

It goes without saying that an auditive linguistic stimulus must also comply with these restrictions. However, as suggested, such a small amount of time can hardly contain sentences, even short, as well as multisyllabic words; for instance, to fit this measure an Italian word should be two-three syllables long as a maximum (depending on the speaker's individual variability).

The production of suitable speech stimuli must, above all, consider the main aim of the IAT implementation. A number of language features can be inferred just from voice quality (see Keller, 2005), and the use of grammatical, senseful words is not always strictly indispensable (see infra).

If the studies exclusively require the use of senseful words as stimuli, these need to be carefully selected to be similar in terms of structure, duration and content. Avoiding words that are excessively different in the syllable number and structure can help to set up rigorous experiments. Alternatively, supposing that the aims of the specific studies so allow, a speech sample can be artificially produced or accelerated (preserving its comprehensibility, see Foulkes, Sticht, 1969) in order to fit the time span. Nevertheless, the major stimuli issues refer often to their content: the next paragraph specifically addresses this aspect.

2.2 Stimuli content

The content of a stimulus comprehends all the elements available for the subject's elaboration. The number and type of these elements highly influences the response and therefore the shape of the experimental result (Bricker, Pruzansky, 1966).

A first crucial distinction must be considered between classic visual IAT stimuli (typically simple images, e.g. a photograph) and auditive stimuli (specifically, speech excerpts). The constituents of a visual stimulus (colors, shapes, objects) appear at the same time during the test and remain on the screen for the total time of the stimulus presentation. On the contrary, a speech sample represents a more complex stimulus, since it is composed of different kinds of information (phonetic-phonologic, morphologic, prosodic, syntactic, semantic, indexical, emotive) that vary throughout the time unit. A similar discontinuous overlap can potentially mask the target elicitation and affect the reaction time (Foroni, Bel-Bahar, 2010), so strategies must be put in place in order to limit and control such complexity.

To simplify input overlapping, employing as stimulus short words (words pronounced in a time span that is sufficiently short to avoid a deeper elaboration) can be useful on itself; however, other aspects of speech have to be homogenized so as to exclude their role on the subject's reaction.

First, it is important to notice that different grammatical categories seem to produce a different response in association-like test settings (see Bagger Nissen, Henriksen, 2006). If possible, words from the same word class should be chosen. The semantic domain exerts a powerful effect on word processing; the semantic priming effect must be considered in the stimuli production, since the sharing of a same domain shortens considerably the reaction time (and vice versa, different domains induce longer reaction times), beyond the actual aim of the test (Masson, 2001).

Semantics is also implicated in the emotional content of a word. The emotivity evoked by a word can sometimes be very strong, and such a connotation may complicate the input processing and make the elicited answer ambiguous. The emotional content of a word can be evaluated using a pre-test phase (e.g., using a Likert-like enquiry, see Alvaréz-Mosquera, Marín-Gutiérrez, 2017, Alvaréz-Mosquera 2017) or by reference to a specific database like EMOTE (see Grühn, 2016).

Studies aimed to reveal implicit attitudes towards sex, age, as well as other biological and physiological features can consider using nonwords, since target characteristics can be recognized only from the indexical details of the speaking voice. As abovementioned, voice quality (see Trask, 1996: 381) itself entails a series of processable indexical information. In order to circumscribe the effects of such information on the answer, if stimuli are produced by a speaker, it is important to ensure that the same subject produces all the stimuli, without relevant modification in voice quality; at least, and if the experimental design consents so, different speakers can be involved in the production of stimuli if they share the same bio-physical and linguistic characteristics: same sex, comparable age, similar accent and so on. In alternative, a speech production software can be considered in order to limit the amount of socio-indexical information extractable from the speech used as stimulus. Nonwords can also be used as target stimuli for studies involving attitudes towards different languages or varieties. In this case, a same nonword can be declined in two different ways in order to follow language- (or variety-) specific phonological forms.

The use of nonwords as auditive stimuli has been practically applied in the creation of the Italian/French IAT: a total of 13 nonwords, randomly created in order to respect a CVCV structure and to not exceed 600 ms length (see Table 1), have been synthetized as speech segments with a text-to-speech software that contains synthetic voices programmed on specific languages (DSpeech); through DSpeech, each nonword has been uttered in its Italian and French phonological forms. The Italian/French IAT protocol was then successfully completed by 15 bilingual participants: no one showed perplexity about the stimuli classification task and, furthermore, stimuli classification was operated with a relatively short latency (M = 829.2, SD = 252.4, in milliseconds). This leads to believe that spoken non-words are unambiguously and quickly recognized. The advantages of the so-constructed stimuli are multiple: the substantial homogeneity of indexical information (due to the involvement of a synthetic uncharacterized voice), the absence of a recognizable semantics (due to the use of non-words) and, consequently, the impossibility to extract a radically different emotive content from the stimulus, not to mention the fact that synthetic speech stimuli are considerably easier to produce.

Table 1 - *Nonwords selected as stimuli for the Italian/French IAT, with transcription, and duration*

Nonword	Voice	IPA	Msec
cosso	ITA	/ˈkos:o/	434
	FRA	/koˈs:o/	521
garo	ITA	/ˈgaro/	395
	FRA	/gaˈʁo/	558
mitta	ITA	/ˈmit:a/	487
	FRA	/miˈt:a/	513
motti	ITA	/ˈmot:i/	454
	FRA	/moˈt:i/	576
piro	ITA	/ˈpiro/	319
	FRA	/piˈʁo/	448
rini	ITA	/ˈrini/	336
	FRA	/ ʁiˈni/	543
roto	ITA	/ˈroto/	415
	FRA	/ʁoˈto/	644
sari	ITA	/ˈsari/	388
	FRA	/saˈʁi/	528

Nonword	Voice	IPA	Msec
saro	ITA	/ˈsaro/	400
	FRA	/saˈʁo/	537
siro	ITA	/ˈsari/	395
	FRA	/saˈʁi/	525
sitta	ITA	/ˈsit:a/	522
	FRA	/siˈt:a/	549
tesso	ITA	/ˈtes:o/	420
	FRA	/teˈs:o/	518
tollo	ITA	/ˈtol:o/	370
	FRA	/toˈl:o/	482

2.3 Stimuli familiarity

Considering that the IAT is specifically developed to capture implicit attitudes towards objects from the informants' social context, it not surprising that the target stimulus will mostly contain elements that are somehow familiar to the subject's social experience. However, the familiarity variable must be treated carefully, since this appears as one of the most influential factors on response latency times (see Ottaway, Hayden & Oakes, 2001; Dasgupta, McGhee, Greenwald & Banaji, 2000; Zogmaister, Castelli, 2006). Stimuli with a higher degree of familiarity are usually processed and categorized faster, regardless of their effective positive/negative value. In a visual IAT, the content of a graphic stimulus can be easily evaluated by the experimenter in terms of familiarity, on the basis of the social and personal characteristics of the informants. Experimental design itself led to a focused participants' selection depending on their overt social experience (e.g., inhabitants of multiethnic neighborhoods would be systematically familiar with a foreign ethnicity).

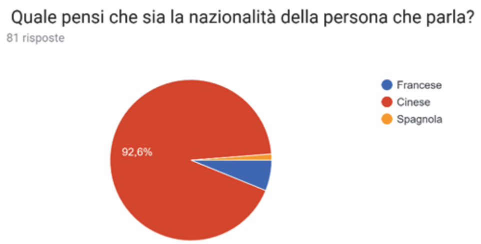
Assuming the familiarity of a sociolinguistic object is definitely more complex, because it presupposes a deeper knowledge of the subject's social behavior, of his linguistic use and exposition. To bridge this gap, it is useful to collect data within a representative population; in this respect, an enquiry can be conducted through the submission of a sociolinguistic questionnaire designed to collect exhaustive information on the linguistic behavior of the potential informant. Specific points should cover usual interlocutors, typical communicative context, and which languages/varieties are used or listened more often. If the stimulus comprehends senseful words, their frequency can be estimated through a lexical frequency corpus such as, for Italian, the CoLFIS (Bertinetto, Burani, Laudanna, Marconi, Ratti, Rolando & Thornton, 2005). In general, high-frequency words are easily processed (Brysbaert, Mandera & Keuleers, 2018) and require a shorter latency time; therefore, it is important for the stimuli words to have a comparable lexical frequency.

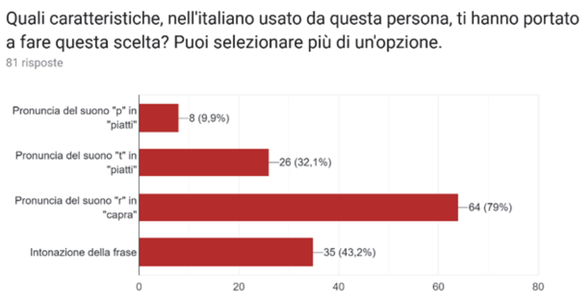
Partly connected to the stimulus' familiarity is the value of its recognizability. An IAT test produces valid results when the stimulus grants a quick recognition of the target element. Ethnicity (here intended as the physical and somatic characteristics due to a common origin), for example, is easily recognizable from a visual stimulus like the image of a face (see, e.g., Sporer, 2001). In the study of the attitudes towards linguistic varieties, ensuring the recognizability of the features of a speech sample requires caution, especially since the elements involved in variety recognition have not yet been clearly identified (see Clopper, Pisoni, 2007; van Bezooijen, Gooskens, 1999).

As for the familiarity variable, experimental validation for recognizability can be achieved through a wide-ranging perceptive pre-test, in which a population (possibly similar in composition to the expected informants) is invited to guess the origin of a speaker by listening a portion of accented speech.

A pre-test of this kind has been used to design the Chinese-accented Italian IAT. In this case, the enquiry is designed as an on-line quiz in order to simplify the approach and to maximize the spread. After hearing a short audio sample, in which brief Italian phrases are pronounced by three male speakers (native of China, France, and Spain), the participant is asked to guess the speakers' nationalities and to freely indicate which characteristics of the speech led him to that choice. The graphic in Figure 1 resumes the answers for one of the questions (a total of 82 people completed the survey randomly recruited on a popular Social Network). The pie chart clearly shows that the native Chinese speaker is easily recognized by the majority of the participants, whereas the bar graph indicates the substitution of the trill consonant as the most recognizable index of a native Chinese speaker: such pretest results constitute a solid speculative base to select IAT stimuli.

Figure 2 - *Answers summary for one of the pre-test questions. The bar chart resumes the main elements detected by speakers as nationality detectors (English translation provided in Appendix)*





3. *Conclusions*

This contribution aimed to present a protocol that, although disputed, allows a new outlook on a complex and crucial psycho-social aspect such as implicit attitudes. At the same time, as highlighted throughout the article, there is a general need to standardize practices and methods in order to fully integrate the IAT within the sociolinguist's toolbox.

Although IAT effectiveness may be questioned, the role of implicit attitudinal constructs cannot be ignored in the analysis of social and sociolinguistic processes. In this sense, it is interesting to note that the simultaneous application of implicit ad explicit protocols often reveals the substantial incongruity between revealed and captured attitudes (see, among others, Campbell-Kibler, 2012; Pantos, 2017; Pantos, Perkins, 2012; Calamai, Ardolino, 2020). These evidences put forth a series of experimental questions that represent the potential future steps of the implicit attitudes research in sociolinguistics: what is the effective role of implicit attitudes in the regulation of social-communicative behaviors? Do implicit biases play a role in linguistic acquisition and loss? What are the consequences of persistent implicit attitudinal constructs within the speech community? These are matters that are evidently in close connection with the issues of linguistic prejudices, namely linguicism (Skutnabb-Kangas, 2015) and glottophobia (see Blanchet, Clerc Conan, 2018), which are sociolinguistic phenomena that produce significant effects of social stigma and discrimination (e.g., Vanegas Rojas, Fernández Restrepo, González Zapata, Jaramillo Rodríguez, Muñoz Cardona & Ríos Muñoz, 2016).

Bibliography

Allport, G.W. (1935). Attitudes. In Murchinson, C. (Ed.), *A Handbook of Social Psychology*. Worcester: Clark University Press, 798-844.

Álvarez-Mosquera, P. (2017). The Use of the Implicit Association Test (IAT) for Sociolinguistic Purposes in South Africa. In *Language Matters*, 48(2), 69-90.

Álvarez-Mosquera, P., Marìn-Gutierréz, A. (2018), Implicit language attitudes toward historically white accents in the South African context. In *Journal of Language and Social Psychology*, 37, 238-248.

Anisfeld, M., Lambert, W.E. (1961). Social and psychological variables in learning hebrew. In *Journal of Abnormal and Social Psychology*, 63, 524-529.

Anselmi, P., Vianello, M., Voci, A. & Robusto, E. (2013). Implicit sexual attitude of heterosexual, gay and bisexual individuals: disentangling the contribution of specific associations to the overall measure. In *PLoS ONE*, 8, e78990.

Ardolino, F. (forthcoming). Recent Italian migrant communities in France: a sociolinguistic investigation. PhD Dissertation: Università di Pisa.

Bagger Nissen, H., Henriksen, B. (2006). Word class influence on word association test results. In *International Journal of Applied Linguistics*, 16, 389-403.

Baron, A.S, Banaji, M.R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. In *Psychological Science*, 15, 53-58.

Bassili, J.N., Brown, R.D. (2005). Implicit and Explicit Attitudes: research, challenges, and theory. In Albarracín, D., Johnson, B.T. & Zanna, M.P. (Eds.), *The handbook of attitudes*. Mahwan: Lawrence Erlbaum, 543-574.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C. & Thornton, A.M. (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*. http://linguistica.sns.it/CoLFIS/.

Blanchet, P., Clerc Conan, S. (2018). *Je n'ai plus osé ouvrir la bouche... Témoignages de glottophobie vécue et moyens de se defender*. Limoges: Lambert-Lucas.

Bricker, P.D., Pruzansky, S. (1966). Effects of Stimulus Content and Duration on Talker Identification. In *The Journal of the Acoustical Society of America*, 40, 1441-1449.

Brysbaert, M., Mandera, P. & Keuleers, E. (2018). The word frequency effect in word processing: A review update. In *Current Directions in Psychological Science*, 27, 45-50.

Calamai, S., Ardolino, F. (2020). Italian with an accent: the case of 'Chinese Italian' in Tuscan high schools. In *Journal of Language and Social Psychology*, 39, 132-147.

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic meaning. In *Lingua*, 122, 753-763.

Clopper, C., Pisoni, D.B. (2007). Perception of dialect variation. In Pisoni, D.B., Remez, R.E. (Eds.), *The handbook of speech perception*. Oxford: Blackwell, 313-337.

Cooper, R.L., Fishman, J.A. (1974). The study of language attitudes. In *International Journal of the Sociology of Language*, 3, 5-19.

Dasgupta, N., McGhee, D.E., Greenwald, A.G. & Banaji, M.R. (2000). Automatic white preference: eliminating the familiarity explanation. In *Journal of Experimental Social Psychology*, 36, 316-328.

Dunham, Y., Baron, A.S. & Banaji, M.R. (2008). The development of implicit intergroup cognition. In *Trends in Cognitive Sciences*, 12, 248-253.

Egloff, B., Schmukle, S.C. (2002), Predictive validity of an Implicit Association Test for assessing anxiety. In *Journal of Personality and Social Psychology*, 83, 1441-1455.

Foroni, F., Bel-Bahar, T. (2009). Picture-IAT versus Word-IAT: Level of stimulus representation influences on the the IAT. In *European Journal of Social Psychology*, 40, 321-337.

Foulkes, E., Sticht, T.G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. In *Psychological Bulletin*, 72, 50-62.

Fridland, V., Bartlett, K. (2006). Correctness, pleasantness and degree of difference ratings across regions. In *American Speech*, 81, 358-386.

Giles, H., Billings, A.C. (1974). Assessing language attitudes: speaker evaluation studies. In Davies, A., Elder, C. (Eds.), *The Handbook of Applied Linguistics*. Oxford: Blackwell, 187-209.

Greenwald, A.G., Banaji, M.R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. In *American Psychological Association*, 102, 4-27.

Greenwald, A.G., McGhee, D.E. & Schwartz, J.L.K. (1998). Measuring individual differences in Implicit Cognition: the Implicit Association Test. In *Journal of Personality and Social Psychology*, 74, 1464-1480.

Greenwald, A.G., Nosek, B.A. & Banaji, M.R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. In *Journal of Personality and Social Psychology*, 85, 197-216.

Grühn, D. (2016). An English Word Database of EMOtional TErms (EMOTE). In *Psychological Reports*, 119, 1-19.

Keller, E. (2005). The analysis of Voice Quality in Speech Processing. In Chollet, G., Esposito, A., Faundez-Zanuy, M. & Marinaro, M. (Eds.), *Nonlinear Speech Modeling and Applications*. Berlin: Springer, 54-73.

Labov, W. (1963). The social motivation of a Sound Change. In *Word*, 19, 273-309.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Lambert, W.E., Gardner, R.C., Olton, R. & Tunstall, K. (1968). A Study of the roles of attitudes and motivation in Second-Language Learning. In Fishman, J.A. (Ed.), *Readings in the Sociology of Language*. The Hague: Mouton, 473-491.

Mariani, L. (2010). Attitudini e atteggiamenti nell'apprendimento linguistico. In *Italiano LinguaDue*, 2, 253-270.

Masson, M.E.J. (2001). Priming, Cognitive Psychology of. In Smelser, N.J., Nalter, P.B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam: Elsevier, 12042-12046.

McGuire, W.J. (1985). Attitudes and attitude change. In Lindzey, G., Aronson, E. (Eds.), *Handbook of Social Psychology*. New York: John Wiley, 233-346.

McKenzie, R., Carrie, E. (2018). Implicit-explicit attitudinal discrepancy and the investigation of language attitude change in progress. In *Journal of Multilingual and Multicultural Development*, 39, 830-844.

Mitchell, B. (2009). The Audio Implicit Association Test: human preferences and implicit associations concerning machine voices. PhD Dissertation: Indiana University.

Nock, M.K., Park, J.M., Finn, C.T., Deliberto, T.L., Dour, H.J. & Banaji, M.R. (2010). Measuring the suicidal mind: implicit cognition predicts suicidal behavior. In *Psychological Science*, 21, 511-517.

Ottaway, S.A., Hayden, D.C. & Oakes, M.A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the Implicit Association Test. In *Social Cognition*, 19, 97-144.

Pantos, A.J., Perkins, A.W. (2012). Measuring implicit and explicit attitudes toward foreign accented speech. In *Journal of Language and Social Psychology,* 32, 27-32.

Pantos, A.J. (2017). Applying the IAT to language attitudes research. In Prikhodkine, A., Preston, D.R. (Eds.), *Responses to Language Varieties Variability, processes and outcomes*. Amsterdam: John Benjamins, 117-136.

Roessel, J., Schoel, C. & Sahlberg, D. (2018). What's in an accent? General spontaneous biases against nonnative accents: An investigation with conceptual and auditory IATs. In *European Journal of Social Psychology*, 48, 535-550.

Rosseel, L., Geeraerts, D. & Speelman, D. (2018). Measuring language attitudes in context: exploring the possibilities of the P-IAT, a case study on regional varieties of Dutch in Belgium. In *Journal of Linguistic Geography*, 6, 20-39.

Rothermund, K., Wentura, D. (2004). Underlying processes in the Implicit Association Test: dissociating salience from associations. In *Journal of Experimental Psychology*, 133, 138-165.

Ryan, E.B., Bülik, C.M. (1982). Evaluations of middle class and lower class speakers of Standard American and German-accented English. In *Journal of Language and Social Psychology*, 1, 51-61.

Skutnabb-Kangas, T. (2015). Linguicism. In Chapelle, C.A. (Ed.), *Encyclopedia of Applied Linguistics*. Hoboken, NJ: John Wiley.

Sporer, S.L. (2001). Recognizing faces of other ethnic groups: an integration of theories. In *Psychology Public Policy and Law*, 7, 36-97.

Tajfel, H. (1974). Social identity and intergroup behaviour. In *Social Science Information*, 13, 65-93.

Trask, R.L. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.

Van Bezooijen, R., Gooskens, C. (1999). Identification of Language Varieties: The Contribution of Different Linguistic Levels. In *Journal of Language and Social Psychology*, 18, 31-48.

Vande Kamp, M.E. (2002). Auditory Implicit Association Tests. PhD Dissertation, University of Washington.

Vanegas Rojas, M., Fernández Restrepo, J.J., González Zapata, Y.A., Jaramillo Rodríguez, G., Muñoz Cardona L.F. & Ríos Muñoz, C.M. (2016). Linguistic Discrimination in an English Language Teaching Program: Voices of the Invisible Others. In *Íkala, Reviste de Lenguaje y Cultura*, 24, 133-151.

Wolff, H. (1959). Intelligibility and Inter-ethnic Attitudes. In *Anthropological Linguistics*, 1, 34-41.

Zogmaister, C., Castelli, L. (2006). La misurazione di costrutti impliciti attraverso l'Implicit Association Test. In *Psicologia Sociale*, 1, 65-94.

Appendix

Figure 1, English translation:

- Place your index fingers on the keys (E) and (I). You're going to listen to a series of words. You'll have to decide if the speaking person is ITALIAN or FRENCH.
- To do it, classify each item in the LEFT category, by pressing the (E) key, or in the RIGHT category, by pressing the (I) key.
- If you fail, a red X will appear. Correct the mistake to continue.
- TRY TO ANSWER AS QUICKLY AS YOU CAN, WITHOUT MAKING MISTAKES
- Press START to continue.

Figure 2, English translation:

- What do you think is the nationality of the speaking person? (French - Chinese - Spanish)
- Which characteristic lead you to this choice? ("p" pronunciation - "t" pronunciation - "r" pronunciation - phrasal intonation).

DOI: 10.17469/O2106AISV000013

VIOLETTA CATALDO, LOREDANA SCHETTINO, RENATA SAVY, ISABELLA POGGI, ANTONIO ORIGLIA, ALESSANDRO ANSANI, ISORA SESSA, ALESSANDRA CHIERA

Phonetic and functional features of pauses, and concurrent gestures, in tourist guides' speech[1]

This study falls into the bigger framework of the CHROME project, addressing the definition and testing of a methodology of collecting, analyzing and modeling multimodal data for the design of virtual agents serving in museums. The paper analyses three tourist guides' speech, focusing on silent pauses and voiced pauses (filled pauses and segmental prolongations). In this regard, a description of phonetic-acoustic and functional features of pauses and a classification of concomitant gestures have been performed. Results show a) speakers' idiosyncratic linguistic behaviours; b) a clear distinction between silent pauses, mainly used for grammatical and intentional reasons, and voiced pauses that instead occur as ungrammatical and hesitation devices; c) such a distinction is confirmed by concomitant gestures, they are semantically loaded in silent pauses and semantically empty in voiced pauses.

Key words: Speech, Disfluencies, Pauses, Gestures, Virtual Agent.

1. *Theoretical background*

1.1 Disfluencies and pauses

Spontaneous human speech shows a widespread occurrence of a number of heterogeneous phenomena, which interrupt, suspend and/or delay the speech flow in the production of the intended message, affecting its fluency. For this reason, such phenomena are generally referred to as disfluencies.

Several studies, belonging to different investigation fields, have concerned disfluencies since the 1950s. Such studies start from different theoretical approaches, ranging from phonetics, speech pathology, psycholinguistics to automatic speech processing (Crocco, Savy, 2003); the coexistence of several theoretical perspectives have caused studies to use and provide different terminologies[2] and a variety of descriptions and classifications (Lickley, 2015)[3]. Despite this situation, some

[1] Authors' responsibilities – University of Salerno, speech analysis: *Violetta Cataldo*, linguistic analysis, writing; *Loredana Schettino*, linguistic analysis, related work; *Renata Savy*, study concept, supervision. University of Naples "Federico II": *Antonio Origlia*, automatic data analysis. Roma Tre University, gesture analysis: *Isabella Poggi*, writing, supervision; *Alessandro Ansani*, writing, data analysis; *Isora Sessa*, gesture coding, discussion; *Alessandra Chiera*, gesture coding, related work.

[2] Among others, "disturbances" (Mahl, 1956), "hesitations" (Maclay, Osgood, 1959), "hesitation phenomena" (Blankenship, Kay, 1964), "dysfluencies" (Johnson, 1961).

[3] A broad and highly regarded overview of disfluencies is provided by Eklund (2004).

agreement has been reached, as reference is made to Levelt's description of the general structure of disfluencies[4] (Levelt, 1983) and his model of speech production (Levelt, 1989), comprising the "double perceptual loop" theory of self-monitoring. Accordingly, speakers' perception of both the own produced speech (external loop) and speech plan (internal loop) enables them to detect and repair a problem before it is articulated. Hence, disfluencies in general trace back to either the external monitoring stage or the internal planning stage.

Disfluency phenomena expressly characterize oral texts as they are closely related to the speech modality; in fact, conditions of low degree of message design by the speaker, simultaneity of planning and production processes and regular concurrence of signals' emission and reception are likely to undermine the speakers' flow of speech (Voghera, 2017).

Fundamentally, the basic distinction regards two main categories[5]:

- **Hesitations**; phenomena of hesitations show a certain degree of speakers' uncertainty and are used as means for taking time to plan what follows in the message. Generally, this kind of disfluencies does not affect the verbal sequence of speakers' productions; instead, a range of phonetic cues, including silent pauses, filled pauses and prolongations, are employed.
- **Repairs**; speakers need to retrace something that has been already said and change it. It can happen for different reasons, all revealing a need for correction of errors that have occurred somewhere in the planning process. In this case, a change in the verbal sequence of the utterance occurs, since the speaker acts on something already uttered, repeating, replacing, adding, removing part of it; indeed, successful repair strategies consist of repetitions, substitutions, insertions, deletions.

Disfluency occurrences belonging to the hesitation category usually involve the temporary suspension of flowing speech (Lickley, 2015), which means that, in most cases, speakers hesitate by means of elements which enable a short-term delay to the production process; therefore, pauses are commonly employed as hesitation devices (Maclay, Osgood, 1959; O'Shaughnessy, 1992). Although pauses are part of disfluencies, as they affect to some extent the speech flow, a number of studies claim that these interruption phenomena need to be investigated from a "positive" approach. Starting by Chafe (1980), hesitations produced in spontaneous speech communication have not been considered anymore as merely disturbances or "errors" of human speech production and researchers in this field have increasingly acknowledged their role in dialogue. In fact, pauses represent the natural outcome of the condition of on-line planning process in human spontaneous speech (Giannini, 2003)

[4] On such account, Shriberg (1994) developed her own highly influential account on disfluencies, whose structure may consist of the following regions: Reparandum (RM), Interruption Point (IP), Editing Phase (EP), Repair (RR).

[5] Such distinction has been considered in a number of studies, to recall some of them: "stalls" (silent pauses, filled pauses, prospective repeats, syllabic prolongations) and "repairs" (false starts, retrospective repeats or bridging) in Hieke (1981); "forward looking disfluencies" and "backward looking disfluencies" in Ginzburg, Fernandez & Schlangen (2014); "hesitations" and "repairs" in Lickley (2015); "*disfluenze fonetiche*" (phonetic disfluencies) and "*disfluenze testuali*" (textual disfluencies) in Voghera (2017).

and a possible solution to in-time discourse planning (Clark, 2002). Using pauses, speakers manage to avoid producing performance errors and to achieve greater well-formedness in their speech (Hieke, 1981). More specifically, the insertion of such elements allows the speaker to gain extra time to retrieve content and provide the listener with valuable meta-information about the ongoing speech (Betz, Wagner & Voße, 2016; Betz, Carlmeyer, Wagner & Wrede, 2018).

In the literature, pauses have been classified and termed differently, according to different research approaches and criteria. The most recurring classification concerns unfilled (or silent) pauses and filled pauses.

Unfilled pauses consist in occurrences of silence of unusual length in the spontaneous speech flow (Maclay, Osgood, 1959); from the articulatory point of view, silent pauses can be produced together with other phenomena, comprising inspiration, swallowing, any laryngo-phonatory reflex, or a silent expiration (Zellner, 1994).

Filled pauses are usually subdivided into "unlexicalized" and "lexicalized" filled pauses; the former, e.g. "ah" or "uh", have been defined as seemingly meaningless words (Gabrea, O'Shaughnessy, 2000), as they are independent nonverbal elements, realized as vocalizations and/or nasalizations. Furthermore, several studies tend to include in this category phenomena of segmental prolongation (Giannini, 2003), while others consider them as a specific form of unfilled pauses (non-phonemic lengthening of phonemes; Maclay, Osgood, 1959). The category of lexicalized filled pauses refers to pauses that have lexical form, thus including different cases found in a number of languages, such as discourse markers (e.g. "you know"), repetitions, false starts (Zellner, 1994).

Researchers on conversational speech synthesis, looking for a model for synthetic disfluencies, considered "pauses" as one of the three micro-structural elements that may constitute disfluencies (Betz, Wagner & Schlangen, 2015): pre-disfluent syllable lengthening (L); cut-offs leading to word fragments (F); silent or filled pauses (P). According to their study, most speech disfluent occurrences are expressed via a combination of the above mentioned elements.

From a more closely phonetic perspective, researchers have expressed particular interest in identifying both segmental and suprasegmental features of pauses, especially of those vocalizations considered as filled pauses.

As regards segmental content, several studies concerning English spontaneous speech claim that filled pauses are typically produced as a steady mid-central vowel close to *schwa* (Maclay, Osgood, 1959; O'Shaughnessy, 1992; 1993; Gabrea, O'Shaughnessy, 2000; Schriberg, 2001).

In the literature, different studies have investigated the prosodic cues of filled pauses. Duration is one of the most noteworthy, as it succeeds in distinguishing filled pauses' segments from other similar vowels contained in unstressed words (such as "a" and "the") in spontaneous speech; in the first context, such vowels are regularly longer than in the second one (Shriberg, 2001). Additionally, specific intonational features have been identified as peculiar to filled pauses and the relationship with their prosodic surrounding (Shriberg, Lickley, 1993); filled pauses have been described as showing a low fundamental frequency F_0 compared to the adjacent context and displaying an ongoing F_0 fall (O'Shaughnessy, 1992; Gabrea, O'Shaughnessy, 2000).

Researchers investigating filled pauses, have also observed segmental prolongations. They were found to be often concurrent phenomena preceding filled pauses (Betz, Wagner, 2016); moreover, both phenomena, in contrast to other disfluency types, show features of vocalization and duration as means for expressing hesitation (Eklund, 2004).

In general, segmental prolongations that can be considered as pauses occur in disfluent contexts; in this way, it is possible to distinguish disfluent prolongations from other cases of lengthening, due to accentuation or in utterance-final or prepausal position. Recent studies have shown that the preferred segmental targets for disfluent prolongations are long vocalic nuclei and sonorant codas (Betz, Eklund & Wagner, 2017). On the intonation level, hesitant prolongations tend to show a specific flat pitch contour compared to non-disfluent prolongations, which are realized with a higher pitch range (for example, due to accentuation; Betz et al., 2017). As regards duration, segmental prolongations are commonly shorter than filled pauses (Eklund 2001; 2004; Betz et al., 2017).

1.2 Gestures

Communication entails a complex interplay between speech and gesture. The study of their relationship is characterized by a debate between two competing hypotheses: according to the Lexical Retrieval Hypothesis (e.g., Krauss, Hadar, 1999; Krauss, Chen & Gottesman, 2000; Morsella, Krauss, 2005) they fulfil different functions, with gestures mainly supporting the message encoding or having a compensatory role; a second hypothesis ascribes to gestures functions similar to those of speech (e.g., Kita, Özyürek, 2003; Kendon, 2004; McNeill, 2005). Support to the first view is given by a more frequent occurrence of gestures in disfluencies, their preparing language in infants (Liszkowski, 2008) and their helping word retrieval (Pine, Bird & Kirk, 2007).

Conversely, some models observe that gestures occur more often in absence of disfluencies (Christenfeld, Schachter & Bilous, 1991) and their strokes co-occur with prosodic peaks (Nobe, 2000), suggesting that speech and gesture are integrated systems playing similar pragmatic functions. For example, the Information Packaging Hypothesis (e.g., Alibali, Kita & Young, 2000; Kita, 2000) holds that gesture and speech are intertwined from early conceptual elaboration, and gestures provide help for thinking and then speaking; indeed, people gesture more frequently when facing a high conceptualization load (Kita, Davies, 2009; Kita, Alibali & Chu, 2017). Other proposals, inspired by McNeill (1992, 2005), argue that speech and gesture form a unique system where the propositional and the mental-image aspects of thought are linked together: they have a similar pattern of development in childhood (e.g., Capirci, Volterra, 2008) and different languages (Kita, 2009), they are synchronized in a semantic harmony in both production and comprehension (e.g., Kendon, 2004; Holler, Schubotz, Kelly, Hagoort, Schuetze & Özyürek, 2014), and are similarly affected by neurocognitive impairments (Duncan, Pedelty, 2007).

Since assumptions about the timing and functions of gesture in pauses (a crucial topic of this work) are opposite in the two frameworks above, their plausibility may be tested exploring the relationship between gesture production and disfluencies.

In the few studies exploring gestures accompanying speech pauses (e.g., Esposito, Marinaro, 2007; Stam, Tellier, 2017; Graziano, Gullberg, 2018; Krauss et al., 2000; Morsella, Krauss, 2005), data are inconsistent. As to their temporal relation, for some authors gestures occur just before or at the same time as disfluencies (Ragsdale, Fry Silvia, 1982), while for others speech and gesture are interrupted simultaneously (Mayberry, Jacques, 2000) or gesture stops before speech stops (Seyfeddinipur, 2006). Some find that disfluencies are specifically synchronized with gestures holds, i.e., the momentary suspension of movement, in both children and adults (Cibulka, 2016; Esposito, Marinaro, 2007): like speech pauses may be involved in the processes necessary to repair a problem in speaking, holds too may signal the activation processes to handle the same problem and re-plan a new message.

As for the function of gestures in pauses, along with a production-oriented function of lexical retrieval, an interactive role of managing turn-taking (Mondada, 2007) and a comprehension-oriented function of adding useful information for the interlocutor mainly in asymmetrical interactions (i.e., doctor-patient) are highlighted (Stam, Tellier, 2017). Graziano and Gullberg (2018) provide further evidence that speech and gesture form an integrated system by examining adult native speakers of two languages and language learners: gestures occur more in fluent than disfluent speech and, in the rare cases of strokes found in pauses, they belong not only to referential but also pragmatic gestures with an interactive function. Moreover, all participants tend to suspend or hold gestures in disfluency, showing that when speech stops, so does gesture (Yasinnik, Shattuck-Hufnagel & Veilleux, 2005).

2. *Object of the study*

The present study is part of the CHROME project – Cultural Heritage Resources Orienting Multimodal Experience, which is aimed at defining and testing a methodology of multimodal data collection, analysis and modeling for the development of a Virtual Agent (VA); such a VA should be able to serve in museums and present cultural sites using an "anthropomorphic" human-machine dialog system. The purpose of the research focuses on this last feature, providing an accurate description and modeling of a number of specific features of verbal behavior, which contribute to speech naturalness and adequacy to its situational context. In fact, VAs usually speak in a fully fluent way, giving the interlocutors a feeling of artificiality and distance; actually, VAs are expected to emulate human-human communication, which, however, shows a series of phonetic and prosodic characteristics that cannot be neglected. In this perspective, phenomena of hypospecification, coarticulation and disfluency need to be investigated on the positive side, looking for characteristics and regularities, which can be effectively implemented in "text to speech" (TTS) synthesis in order to improve its performances. In the framework of the CHROME project, this work aims at: a) describing and analyzing a set of speech disfluency phenomena occurring in tourist guides' speech, focusing on three different types of pauses: silent pauses, filled pauses, vocalized filled pauses; b) researching patterns and regularities of such phenomena to be implemented in TTS system; c) investigating the existing relationship between pauses and concurrent gestures.

3. *Methodology*

3.1 Corpus and dataset

The present research focuses on a limited dataset drawn from the whole corpus collected for the CHROME project (Origlia, Savy, Poggi, Cutugno, Alfano, D'Errico, Vincze & Cataldo, 2018). The corpus consists of audiovisual recordings of guided tours led in Italian at the San Martino Charterhouse in Naples. Three female expert tourist guides accompany small groups of visitors in four guided tours, each of approximately an hour, for a total amount of about 3h30' of speech for each guide. The guided tours take place in six "points of interest" (POIs) of the Charterhouse: *pronaos*, great cloister, parlor, chapter hall, wooden choir, treasure hall[6]. The dataset of this study takes into account one POI of one visit led by each of the three expert guides, amounting to 36'88" of speech; the first POI, namely the *pronaos*, represents the opening moment of the guided tour, which starts at the doorway to the church of the Charterhouse.

From a linguistic point of view, the conditions set by the corpus identify a particular kind of oral texts, typical of tourist guides' oral performances. The guides' speech is characterised by a) a high degree of discourse planning, in fact, tourist guides make regular use of descriptive texts, which are partially pre-structured and repetitive; b) high selective attention of the participants, because of the hierarchical relationship between the guide and the audience, due to the guide's professional and linguistic competence in the topic; c) a resulting low degree of interlocutors' dialogic interaction and participation in the discourse construction; d) a close integration between verbal and non-verbal elements, due to the spatial context which, for example, makes it necessary to use spatial, verbal and gestural, deixis. Given the above features, it can be assumed that this study deals with semi-spontaneous and semi-monological speech.

3.2 Method

3.2.1 Disfluency annotation system

The three selected audiovisual recordings have been annotated on different linguistic levels.

[6] Detailed data collection protocols are provided by Origlia et al. (2018).

Firstly, orthographic transcription and phonetic and syllabic annotation have been carried out (see Origlia et al., 2018) using the software Praat (Boersma, Weenink, 2018).

As concerns disfluency phenomena, an *ad hoc* encoding system has been adopted, based on the disfluency modeling works of Hieke (1981), Shriberg (1994) and Lickley (1998). The system consists of four annotation tiers, which, although parallel to each other, refer to different occurrence domains. For this reason, disfluency phenomena have been annotated using the ELAN software (2018), which allows to carry out a multilevel annotation. A comprehensive description of the disfluencies' four domains and their relative annotation tiers follows.

- **Disfluency Type;** this first annotation level specifies the category of the occurring disfluent phenomenon. A finite number of categories has been selected, in the attempt to cover all the possible occurrences:
  - *Fresh Start*. It involves cases of false starts, when the speaker stops and rephrases a new utterance with no morphosyntactic and/or semantic relationships to the previous one.
  - *Repeated Start*. In the utterance, the speaker stops and exactly repeats something s/he has already uttered; such category deals with repetitions of single words, fragments of words, as well as whole utterances.
  - *Edited Start*. This category is divided into two subcategories; in both cases the speaker rephrases part of the string, either adding one or more elements (Addition subcategory), or substituting an element with another syntactically and/or semantically equivalent one (Substitution subcategory).
  - *Hesitative Start*. In this case, the speaker shows and produces a hesitation without repeating, substituting or abandoning any part of the utterance in production.
- **Disfluency Function;** this level provides information about the pragmatic function performed by each disfluent phenomenon. Such a pragmatic description specifically refers to Hieke's classification (1981) into retrospective and prospective disfluency phenomena. Accordingly, each disfluency type is associated with either a retrospective function, which plays a corrective role when the speaker "corrects" something s/he has already uttered, or a prospective function, which plays a control role for the speaker's production to avoid errors.
- **Disfluency Model;** this annotation level describes the occurrence model of disfluencies, based on Shriberg (1994) and Lickley (1998). Different regions are identified and annotated, namely:
  - *Reparandum*, the region where the speaker encounters difficulties and that will be later "repaired";
  - *Interruption Point*, a temporary suspension, where the speaker stops as s/he realizes to have difficulties;
  - *Interregnum*, a moment of transition where the speaker shows his/her hesitation;
  - *Repair*, the region that "repairs" to the Reparandum;

  - *Original Utterance*, the region that precedes the Interregnum in the case of the disfluencies that have no Reparandum (such as Hesitative Start);
  - *Continuation*, the region that follows the Interregnum in the case of the disfluencies that have no Repair (again, in the case of Hesitative Starts).
- **Disfluency Components;** based on the regions' identification in the previous level, here information about the inner phonic and/or linguistic components of such regions are provided.

A summary table for the disfluencies' annotation levels and the relative categories and labels is provided below (Table 1).

Table 1 - *Disfluency annotation system: levels of annotation and categories*

Levels of annotation	Categories
Disfluency Type	Fresh Start, Repeated Start, Edited Start Addition, Edited Start Substitution, Hesitative Start
Disfluency Function	Retrospective function, Prospective function
Disfluency Model	Reparandum, Interruption Point, Interregnum, Repair, Original Utterance, Continuation
Disfluency Component	Word, Word Fragment, Filled Pause, Vocalized Filled Pause, Discourse Marker Filled Pause, Silent Pause

3.2.2 Classification and analysis parameters of pauses

The present research primarily focuses on the analysis of pauses, classified as follows: a) Silent Pauses (SPs), pauses of silence; b) Filled Pauses (FPs), regarded as vocalizations and/or nasalizations, i.e. "eeh", "ehm", "mhh"; c) Vocalized Filled Pauses (VFPs), resulting from word-final segmental prolongation of lexical elements.

Pauses have been investigated from both a form and function perspective, in order to provide both a phonetic-acoustic description and a functional classification.

As concerns the phonetic-acoustic description, pauses' analysis has been conducted according to the following parameters:

- *Duration*; measured in milliseconds (ms).
- *Segmental content*; pauses' phonetic realizations and their characteristics have been investigated, considering their vowel quality and phenomena of diphthongization, triphthongization, nasalization, devocalization.
- *Pitch profile*; it provides a description of pauses' intonation patterns, classified as rising, falling, flat or valley. Additionally, their relative range level has been calculated in Herz (Hz) and semitones (ST). Both parameters refer to pauses compared to their cotextual sequence.

The occurrences of the three pause types have been extracted and classified according to the following primary functions:

- *Physiological function* (PHYS): it follows the "respiratory" function of Viola and Madureira (2008), which reflects the speaker's physiological need to pause and take a breath;
- *Demarcative function* (DEM): demarcative pauses play a grammatical role in structuring the discourse at different linguistic levels, such as intonation, syntax, information structure (Swerts, 1998);
- *Programmative function* (PROG): these pauses show the speaker's on-line process of planning and his/her difficulty in retrieving specific lexical elements (see Schnadt, Corley, 2006; Hartsuiker, Notebaert, 2009);
- *Hesitative function* (HES): pauses are widely employed to express the speaker's uncertainty over the general planning of the content as well as the form of his/her message ("unintended hesitation pauses", O'Shaughnessy 1992; "means of hesitation", Eklund 2001);
- *Strategic-rhetorical function* (STR-R): the speaker makes use of pauses in a partially conscious and deliberate way, in order to attract the audience's attention (Betz et al., 2016), introduce or underline key concepts (Duez, 1997), give particular emphasis on specific words (Strangert, 2003).

3.2.3 Gesture annotation system

To assess the relationships between gestures and pauses, among annotation systems proposed by previous works (see for instance Kong, Law, Kwanm Lai & Lam, 2015) the guides' gestures concomitant to pauses were annotated in terms of the following categories (Table 2), partly drawn from Colletta, Kunene, Venouil, Kaufmann & Simon (2009), but adapted according to Poggi (2007).

- **Gesture functions:** each gesture was classified in terms of these categories: *deictic*; extended index, thumb or the whole hand point at some place in the physical context where the referent of discourse presently is or can be connected to; e.g., while saying "*come dicee... la, laa... dicitura stessa del museo*" (as said by theeee... very wording of the museum), during the VFP *"laa..."*, the guide *moves both hands, with palms up, downward and rightward*, pointing at words written on the Museum entrance.
  - *Iconic;* the shape or movements of the hand(s) imitates the shape or movements of the referent. During a pause, before saying "*si prolunga*" (it is prolonged), the guide *moves her left hand, palm down, with almost closed fingers from right to left* as if picking and dragging something in a long fluctuating line, to represent something going through a long path.
  - *Metaphoric*; an iconic gesture refers to some abstract concept, or some inference is required to go from the bare imitation of a shape or movement to the intended meaning. While saying "*il certosino doveva preservare la sua vita isolata, contemplativa*" (the Chartusian had to preserve his isolated, contemplative life), the guide *pulls back her hands with palms forward*, to represent someone's withdrawing from something, which metaphorically means to retreat from life in the outside world.

  - *Coded*; this category includes not only symbolic gestures, those with a codified meaning and a shared verbal translation in a given culture (like *thumb and index making a ring* for "ok"), but also other gestures, without a straightforward verbal paraphrase, which yet do convey a shared meaning (Müller, 2004; Kendon, 2004; Nobili, 2019). The guide, while saying *"perché abbiamo... a cuore la conservazione del pavimento"* (because we care about the conservation of the floor) during the disfluency makes the *Palm Up Open Hands* gesture, which means "this is self-evident".
  - *Beat*; rhythmic gestures, with hands generally with a up-down movement, that emphasize a word in a sentence or a syllable in a word in order to highlight its importance, or scan the rhythmic structure of words for clearer articulation.
  - *Manipulator* (Ekman, Friesen, 1969); hands smoothing or rubbing between themselves or with other parts of the speaker's body, generally aimed at a reassuring self-contact, hence only indirectly being a cue to embarrassment or discomfort;
  - *Idle*: a "non-gesture", with hands not at rest.
- **Gesture phases**: following Kendon (1980) and Kita (1990), the phases of the gestures were annotated as:
  - *Preparation*; when the hand starts from its resting position;
  - *Stroke*: when it reaches its farthest point from the resting position, after which it starts to go back to it;
  - *Hold*: the hands remain on the stroke point before going back to the resting position;
  - *Chain*: the hand, after reaching the stroke point, starts retracting but then repeats its run to the stroke and back;
  - *Return*: after reaching the farthest point the hand goes back to the resting position.
- **Gesture meaning**: for communicative gestures (e.g., not idles), a verbal paraphrase of it is annotated: e.g. *Palm Up Open Hand* = "this is self-evident".
- **Synchrony**: based on the meaning attributed to the gesture, its semantic relationship is annotated with the co-occurring words or pause, with respect to which it can be anticipating, following, or synchronous. When the guide says *"come dicee... la, laa... dicitura stessa del museo"* (as said by theeee... very wording of the museum), although referring to the wording (*dicitura*), her deictic gesture falls during the VFP following "laa...", thus anticipating the referent it points at.

Table 2 - *Gesture annotation system: levels of annotation and categories*

Levels of annotation	**Categories**
Gesture Function (RHGF)	Deictic, Iconic, Metaphoric, Coded, Beat, Manipulator, Idle
Gesture Phases (RHGP)	Preparation, Stroke, Hold, Chain, Return
Gesture Meaning (Meaning)	"………"
Synchrony (RHGS)	Anticipating, following, Synchronous

The annotations were performed by two expert judges for pauses (Cohen's *k* of 0,7) and two for gestures (Cohen's *k* of 0,7).

4. *Results*

4.1 Linguistic analysis

The following section presents the research results, which emerged from the speech analysis. Firstly, general data on pauses' occurrence (Table 3) and incidence in the total speech of the dataset (36'88" of speech) are reported.

On a total of 384 pauses, occurrences of SPs exceed those of both FPs and VFPs, which present respectively 103 and 101 occurrences.

As regards incidence data, pauses, independently of the type of pause, register a per-word rate of 0,07 and a per-minute rate of 10,4, meaning that speakers produce a pause per about 14 words. More specifically, SPs register a per-word rate of 0,03 (a pause per about 30 words), FPs of 0,02 (a pause per about 53 words), and VFPs of 0,02 (a pause per about 54 words).

Table 3 - *Number of pauses' occurrences per pause type (SP, FP, VFP) and per speaker (G01: first guide; G02: second guide; G03: third guide)*

	G01	G02	G03	tot
SP	129	28	23	180
FP	82	14	7	103
VFP	72	27	2	101
tot	283	69	32	**384**

4.1.1 Idiosyncratic linguistic behaviour

Once reported the overall occurrence and incidence data of the three speakers of the dataset, it seems particularly noteworthy to examine speakers' individual speech in greater detail. In fact, it emerged that the three guides adopt different linguistic behaviors, as it can be seen from the different number of pauses uttered by each speaker (Table 3).

The first guide (G01) makes use of a higher number of pauses compared with the other two guides. She reports a per-word rate of 0,19 and a per-minute rate of 24,5, i.e., a pause per only about 5 words. Differently, the second guide (G02) produces pauses four times less often than G01, with a per-word rate of 0,03 (a pause per about 33 words) and a per-minute rate of 5,1. Ultimately, pauses in the third guide's (G03) speech register a definitely lower incidence: a per-word rate of 0,02 and a per-minute rate of 2,7.

4.1.2 Phonetic-acoustic features

Firstly, duration values of the three types of pauses are presented (Table 4).

Table 4 - *Duration values (ms) per pause type (SP, FP, VFP) and per speaker (G01, G02, G03)*

	G01		G02		G03		all speakers	
	mean	st.dev.	mean	st.dev.	mean	st.dev.	mean	st.dev.
SP	371	0,41	750	0,88	285	0,19	469	0,49
FP	486	0,33	250	0,15	360	0,23	365	0,23
VFP	253	0,12	260	0,08	93	0	202	0,07

SPs are the type of pauses which have the longest mean duration (469 ms); SPs show quite high deviation standard values among the three speakers; this can be explained by the fact that very variable lengths have been found, as SPs can be stretched in time without appearing detrimental for the conversation. FPs show a mean duration of 365 ms, with lower values of intra- and inter-speaker standard deviation. VFPs have a mean duration of 202 ms, resulting in a less variable value (standard deviation of only 0,07). This seems to happen for two main reasons. On the one hand, G02 and G03 realize very few VFP occurrences compared to G01; consequently, duration values appear more homogenous, as they specifically refer to VFP occurrences in almost one speaker's speech, having an effect on the overall standard deviation value. On the other hand, it seems that VFPs, unlike SPs and FPs, cannot be excessively stretched in time, without risking to appear deviant for the conversation (d'Urso, Zammuner, 1990). The overall duration values of our study follow the same pattern found by Eklund (1999), namely VFPs < FPs < SPs[7].

The parameter of segmental content is obviously considered only for the two types of voiced pauses, namely FPs and VFPs (Table 5).

[7] In Eklund (1999): PRs < FPs < UPs (prolongations < filled pauses < unfilled pauses).

Table 5 - *Segmental content per pause type (FP, VFP)*

	FP		VFP	
	n.occ.	% occ.	n.occ.	% occ.
vowel	77	75%	95	94%
consonant	11	11%	1	1%
vowel + consonant	15	15%	1	1%
consonant + vowel	0	0%	4	4%
TOT	103	100%	101	100%

FPs and VFPs occur with almost the same number of occurrences in the dataset. With regard to their segmental content, both types of pauses are realized with an entirely vocalic content (75% of FPs and 94% of VFPs).

More specifically, more than half of FP occurrences (62%) is realized by a *schwa* (in a few cases, a feature of nasalization has been found, 8 cases). The other cases of fully vocalic FPs are realized by alternatives of mid-frontal vowels, namely [e] or [ɛ], or diphthongs in which one of the two vowels is a *schwa* (however, it happens in very few cases, 7 occurrences). As to other segmental realizations of FPs, all consonant cases present the nasal [m] (11 occurrences), whereas the cases of vowel-consonant sequence are realized by a *schwa* followed by the nasal [m] (15 occurrences). No cases of FPs expressed by a consonant-vowel sequence have been found.

Concerning voice quality, it is worth noticing that 67% of FPs is realized with creaky phonation, regardless of FP segmental content. This seems to be a regular feature of FPs (see also Schriberg, 2001), probably caused by a lesser articulatory effort by the speaker in producing non-lexical elements.

Detailed data on VFP segmental content are reported below. As mentioned, the vast majority of VFPs has vocalic content. As opposed to FPs, VFPs occurrences present a wider range of vocalic alternatives, in order of frequency: 27 cases of [a]; 16 cases of [ɛ]; 14 cases of [ɔ]; 10 cases of [i]; 4 cases of [e]; 1 case of [o]; 1 case of [u]; no cases of [ə]. Moreover, 21 occurrences of diphthongs and one occurrence of triphthong have been found. Such a heterogeneity can be explained considering that VFPs are vocal prolongations of the word-final sound; as Italian syllables show a consonant-vowel (CV) canonical structure (Berruto, Cerruti, 2011), VFPs mainly deal with prolongations of different vowels. Other possible VFP realizations concern consonant lengthening: such prolongations are realized producing a final *schwa* in order to lengthen the consonant, except for nasal prolongations (1 occurrence of prolongation of [m]). As an outcome of coarticulation, 37 vocalic VFPs out of 95 are affected by nasalization due to their nasal acoustic context; for instance:

tradizione<ee> napoletana

[traditˌtsjonẽːnapoleˈtana]

As to intonation features of voiced pauses, fundamental frequency (F_0) profile and range level have been considered. Table 6 shows the occurrences of voiced pauses according to their F_0 profiles. One occurrence of both FP and VFP, totally realized with creaky voice, did not allow to clearly identify the actual intonation pattern; such cases are ignored in the analysis and reported as "not categorized".

Table 6 - *Pitch profiles per pause type (FP, VFP)*

	flat	rising	falling	valley	n.c.	tot
FP	43	20	8	31	1	103
VFP	68	7	11	14	1	101

As it can be observed, the two types of pauses show similar pattern behaviors: the greater part is realized with a flat pitch profile with respect to the preceding and following strings; in particular, VFPs register a higher number of flat occurrences (FPs: 43 cases; VFPs: 68 cases). Moreover, in both cases, few occurrences of rising and falling profiles have been found. Despite such similarities, what emerges is that FPs show a considerable amount of valley profiles (31 occurrences). In these cases, the pauses' intonation pattern is steadily overall flat, but the pitch is lower than the global F_0 pattern of its co-text.

Such a difference might be explained by the pauses' intrinsic characteristics. The majority of steady and flat F_0 profiles of VFPs suggests that such pauses, being word prolongations, are naturally performed as continuations of the lexical elements they lengthen on the segmental level. Conversely, FPs are not integral part of words, and behave as separate elements; hence, FPs are more likely to be produced with freestanding realizations, often showing a tonal trough relative to the co-text and appearing more embedded in it at the intonation level.

Range level values seem to confirm the difference between the two types of pauses (Table 7). In fact, VFPs show a mean F_0 value of 189,4 Hz, only 0,8 ST lower than their mean co-text (198,9 Hz) against the 1,4 ST of FPs. Such a higher difference reflects the number of valley pitch profiles. In this regard, a statistical analysis of range values of FPs and VFPs has been carried out; although it did not result significant, it indicated a general tendency towards a lower and more embedded F_0 for the FPs.

Table 7 - *F0 values of FPs and VFPs (per speaker: G01, G02, G03) and of their relative co-texts*

	FPs				VFPs			
	G01	G02	G03	mean	G01	G02	G03	mean
pause F_0 (Hz)	196,2	196,6	184,6	192,5	194,0	192,6	181,5	189,4
cotext F_0 (Hz)	209,9	211,3	206,7	209,3	199,7	201,9	195,0	198,9
difference (Hz)	13,7	14,7	22,1	16,8	5,7	9,4	13,5	9,5
difference (ST)	1,2	1,2	2,0	1,4	0,5	0,8	1,2	0,8

4.1.3 Functions

The functional classification aims at identifying the primary functions (described in § 3.2.2) performed by each pause occurrence. It is important to highlight that pauses have been considered as multifunctional elements, in fact, each pause occurrence is likely to simultaneously carry out more than one function. This allows for cases of overlapping of two or more functions fulfilled by a single pause. By implication, Table 8 shows the actual number of functions' occurrences per pause type.

Table 8 - *Occurrence number of functions per pause type*

	PHYS	DEM	PROG	HES	STR-R
SP	9	95	0	22	114
FP	0	18	52	99	3
VFP	0	11	66	99	8

An overall look at the pauses' functional behavior suggests that the three pause types can be considered as two macro categories: in fact, SPs systematically behave differently from voiced pauses, namely both FPs and VFPs.

Firstly, SPs are the only pause type to perform a PHYS function, fulfilling the speaker's respiratory need (9 occurrences). In this regard, it has to be noted that cases of breaths have not been taken into account as SP occurrences. On the contrary, no SP occurrence performing a PROG function has been found; indeed, such a function seems to be properly performed by voiced pauses. What appears quite remarkable data is that SPs show a considerable number of DEM e STR-R functions (95 and 114 occurrences, respectively). Pauses performing DEM functions assume the grammatical meaning of sectioning utterances at different linguistic levels; the majority of DEM SPs (65 out of 95) serves simultaneously as grammatical boundaries of constituents of syntactic structure, intonation units, and information structure. The following example shows a SP occurrence performing a DEM function according to the three linguistic levels mentioned above:

> *la Certosa di San Martino qui a Napoli ha almeno due anime* <SP> *una racconta la storia dei certosini.*
>
> the San Martino Charterhouse here in Naples has two souls <SP> one tells the story of the Carthusian monks.

SPs result to be extensively used for strategic goals (63%).

As already said, voiced pauses exhibit a similar behavior, as FPs and VFPs tend to share the same distribution of functions: no occurrences of PHYS function, rare occurrences of STR-R function (3 and 8 cases respectively), few occurrences of DEM function (18 and 11 respectively), common occurrences of PROG function (52 and 66 respectively) and a considerable number of HES occurrences (99 in both cases). More specifically, almost all occurrences of DEM and PROG functions overlap with the most performed HES function. Such a situation might be explained by the fact that both types of voiced pauses are widely used by the speakers as a means of expressing their indecisiveness about the general message to convey. Both FPs and VFPs register almost the totality of HES function, although showing high multi-functionality.

What is particularly worth noticing is the total overlapping between PROG and HES functions; more specifically, all the pauses carrying out a PROG function have been at the same time considered as performing a HES function. In general, HES pauses reflect the speaker's planning phase and the possible difficulties in dealing with general planning demands. In almost half of these cases, corresponding to PROG pauses, such a planning phase concerns expressly lexical elements. In this perspective, the occurrence of PROG pauses reveal the ongoing of a Word Searching (WS) process: the speaker makes use of pauses when she is struggling with lexical retrieval. WS phenomenon is more likely to occur when what the speaker is trying to retrieve are lexical elements with lower frequency (Lickley, 2015). The non-linguistic task requested to the speakers of the present study affects in this regard their lexical choices: the three tourist guides resort to a great number of technical terms belonging primarily to architectural, artistic, religious specialized languages in order to describe the Charterhouse and its history[8]. Such technical terms are lexical items, which show lower frequency than others, including features of lower name agreement and familiarity in the communicative exchange between the speaker and her audience. These characteristics entail a greater deal of effort for accessing to these words in both production and perception. The following example shows the presence of a FP just preceding the technical term *cappella* ("chapel"):

poi c'è anche una <FP> *cappella dedicata a San Giovanni Battista*

furthermore, there is a <FP> chapel consecrated to St. John the Baptist

The FP performs simultaneously a HES function in a broad sense and a PROG function in a narrow sense, working as a WS tool.

4.2 Correlation

4.2.1 Correlation between phonetic-acoustic features and functions of pauses

This part of results mainly concerns FPs and the functional distinction between mere HES FPs and HES FPs working as WS devices (WS FPs). This is supported by the parameters taken into account for the phonetic-acoustic analysis; the parameters of duration and range level seem the most interesting. It should be considered that statistical analysis concerned data relative to the three guides; however, most

[8] An analysis of the lexicon of the whole oral corpus collected in the CHROME project has been carried out (Senigalliesi, Sparano, Schettino, Savy, Dell'Orletta, Lubello & Basile, 2018). The lexicon of the oral texts results to be composed as follows, according to the GRADIT (De Mauro, 1999): 41% of *lessico fondamentale*, 16% of *alto uso*, 4% of *alta disponibilità*, 39% of *tecnicismi* (technical terms). More than half of the technical terms belong to the architectural field.

FP occurrences were produced by G01 (§ 4.1.1). Hence, results regarding the phonetic-functional correlation mainly trace back to one guide's linguistic behavior.

As concerns the first parameter, duration values correlate with the function carried out by FPs. WS FPs are twice longer than HES FPs (mean values: 590 ms and 300 ms, respectively); such a difference results to have statistical significance ($p < 0.001$)[9].

With regard to range level, FPs show a number of flat or valley intonation realizations and low fundamental frequency, thus occurring as tonal troughs within the neighboring speech (§ 4.2.1). On closer inspection, the range level parameter emphasizes the functional distinction. In fact, WS FPs are realized with even lower range values compared to their co-text (WS FPs: 1,6 ST vs. HES FPs: 0,9 ST). By implication, FPs carrying out a HES function tend to preserve the mean range level of the surrounding strings; on the other hand, WS FPs are more embedded within the co-text.

4.2.2 Frequency of gestures during pauses

Concerning the gestures co-occurring with pauses, we only considered the pauses longer than .20 sec. The remaining pauses were 223: 101 SPs (45.29%), 57 VFPs (25.56%), and 65 FPs (29.15%). A chi-square test for the distribution of gesture types among the pause types is significant for $p < .10$ (Table 9, 10).

Table 9 - *Chi square test (gestures types per pause type)*

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	18,656a	12	,097
Likelihood Ratio	21,337	12	,046
N of Valid Cases	223		

a. 8 cells (38,1%) have expected count less than 5. The minimum expected count is ,77.

Table 10 - *Distribution of gestures among pauses*

		Coded	**Deictic**	**Iconic**	**Metaphoric**	**Beat**	**Manipulator**	**Idle**
SP	n.	32	10	5	2	14	6	32
	%	31,68	9,9	4,95	1,98	13,86	5,94	31,68
VFP	n.	20	2	1,98	0	4	7	24
	%	35,09	3,51	13,86	0	7,02	12,28	42,11
FP	n.	20	3	5,94	1	6	12	23
	%	30,77	4,92	31,68	1,54	9,23	18,46	35,38

As already found by Graziano and Gullberg (2018), "when speech stops, gesture stops"; our first robust result is that the idle gestures are the most frequent in all

[9] Linear regression: dependent variable=duration; independent variable: function (2 levels: WS, HES). FuncWS_t-value: 5.21; p-value < 0.001.

three pause types, with the highest value in correspondence with VFPs (42.11%). Moreover, a clear difference emerges between the frequency patterns of gesture categories in SPs, on the one side, vs. FPs and VFPs on the other. The respective patterns are represented in Table 11.

Table 11 - *Ranking of gesture functions in pauses*

	1	2	3	4	5	6
SP	idle & coded	beat	deictic	manipulator	iconic	metaphoric
VFP	idle	coded	manipulator	beat	deictic	
FP	idle	coded	manipulator	beat	deictic	metaphoric

While *idles* and *coded* have the same frequency in SPs, in both VFPs and FPs *idles* are slightly more frequent than *coded*, while *manipulators* are more frequent than *beats* and *deictics* in Voiced Pauses (FPs and VFPs), and the opposite is the case in Silent pauses.

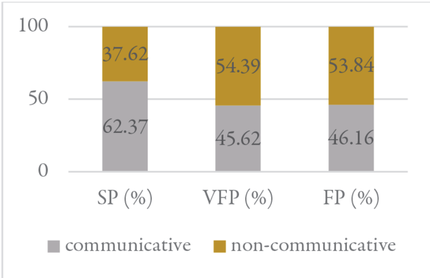
Starting from gesture categories, the percentage of *coded* gestures is almost similar across pause types; *deictics* mostly occur in correspondence with SPs, probably because they are usually performed immediately after uttering a locative adverb or a deictic pronoun, in silence, by pointing to a given object. *Iconic* gestures only show during SPs, perhaps because in these cases the speaker helps herself to recall a specific word to describe an item by imitating the item shape. As for *metaphoric* gestures, the occurrences are too poor to be commented upon. *Beat* gestures are much more frequent in SPs than FPs and VFPs.

The remaining two categories show an inverse pattern of frequency: *manipulators* are more frequent in FPs, followed by VFPs, maybe because they are performed, at a low level of consciousness, in order to self-assure oneself; and *idles* are more frequent in VFPs than FPs, only finally followed by SPs.

How can we explain these opposite patterns for the first five categories as opposed to the last two? As argued in Origlia, Savy, Cataldo, Schettino, Ansani, Sessa, Chiera & Poggi (2019), these categories differ as to their communicative import. Four of them – *deictic, iconic, metaphoric, coded* – are definitely communicative, since in performing them the speaker has a conscious intention to convey some meaning, while the last two – *idle* and *manipulator* – are non-communicative: an *idle* is a non-gesture, simply resting hands, while a *manipulator* may sometimes provide information (embarrassment, anxiety) to an observer, but the Sender does not have the goal for it to leak out: so it may be informative but not communicative. Finally, a *beat* may be communicative if governed by a specific goal of emphasizing something, but sometimes it simply accompanies the rhythm of speech by synchronous body movements, thus making it easier for the Speaker to impress it the right temporal structure.

Based on a chi squared test, the distribution of communicative and non-communicative gestures across pauses is $c^2(2, N = 223) = 6.03, p = .049$ (Graph 1). Such a distinction between communicative and non-communicative movements might account for the opposite pattern of Silent versus Voiced Pauses, with the communicative ones more often co-occuring with the former, and the non-communicative with the latter. Silent Pauses thus look as a moment of higher communicativeness, while Voiced ones as mainly of use for the Sender, less Addressee-oriented.

Graph 1 - *Communicative and Non-communicative gestures across pauses*



4.2.3 Relation between gestures and pauses functions

Three more analyses of HES, STR-R and PROG (WS) can deepen the relationship between pauses' functions and co-occurrent gestures' communicativeness (Table 12). Three 2-sided Fisher's Exact tests were run using a 2 x 2 contingency table. One for *HES x Communicativeness* (p = .007) shows that communicative gestures are less frequent, while non-communicative ones more frequent during hesitative pauses; one for *STR-R x Communicativeness* (p = .015) shows an opposite distribution: communicative gestures co-occur more with Strategic-Rhetorical pauses, non-communicative ones with non-strategic pauses. For *WS x Communicativeness* (p = .029), like with HES pauses, no-WS pauses present a higher number of communicative gestures, whereas those during WS pauses in the vast majority do not have a communicative function.

Table 12 - *Distribution of communicative and non-communicative gestures across pauses' functions*

	No-HES	HES	No-STR-R	STR-R	No-WS	WS
communicative	61,24%	42,55%	48,85%	69,39%	57.74%	40.00%
non-communicative	38,76%	57,45%	51,15%	30,61%	42.26%	60.00%

4.2.4 Amount of Movement (AoM)

To assess if the occurrence of pauses is associated with stops in hand movements, we computed the Amount of Movement (AoM) in the presenter's speech through automatic tracking of her hands: frame-by-frame changes in hands positions provide estimate of hands activity. In correspondence with pauses, two groups were considered: AoM equal or higher than a fixed threshold (0.2) at the beginning of FPs/VFPs vs. AoM lower than the threshold at the beginning of the FPs/VFPs. For each group, the rate of change was computed as the difference between starting AoM and ending AoM, representing how much AoM changed during the occurrence of a pause. Since the Shapiro test confirmed the normality of the two distributions for Group 2 only, the Wilcoxon rank-sum test was used for the comparison. Group 1 has a median AoM rate of change significantly lower than Group 2 ($p<0.01$, ES= 0.4). Moreover, while the median AoM of Group 1 is significantly lower than 0 ($p<0.01$), the median AoM of Group 2 is not significantly different ($p>0.7$). This appears to indicate that, if the considered guide is moving, an FP/VFP occurrence is usually accompanied by a drop in AoM while, if the guide is not moving, AoM does not change across the FP/VFP occurrence.

5. *Discussion*

The results of the present study can be divided into three main parts. The first part that appears to be particularly noteworthy concerns the differences among the three tourist guides' individual speech. Data on pauses' occurrence and incidence presented in § 4.1 reflect the idiosyncratic linguistic behavior displayed by the tourist guides; in fact, they employ different speech strategies aimed at performing their non-linguistic task of leading a tourist tour.

The high incidence of pauses' occurrences registered in the first guide (G01) reveals that this speaker adopts what has been termed an "on the fly" strategy of formulation (Ferreira, Lau & Bailey, 2004): she manages to sound as spontaneous as possible to her audience and to produce an error-free speech using pauses as devices of wellformedness.

In contrast, both the second guide (G02) and the third guide (G03) make a more limited use of pauses, although not employing the same strategy.

G02 tends to avoid both silent and voiced pauses, which could be perceived as signals of hesitation by the listeners, filling the moments of hesitation with a strategy of "juxtaposition" of utterances, clauses, sentences. Such a resulting speech turns out to be highly error-full, mainly at the morphosyntactic level: some of the errors are corrected by means of retrospective disfluency phenomena (see § 3.2.1), while others are simply ignored or passed by.

On the other hand, G03 plans her speech very carefully and adopts a strategy of "rhetorical control". She resorts almost exclusively to SPs and avoids both FPs and VFPs, producing a more error-free and high-quality speech.

The second part of results provides a general comment about the employment of pauses in the tourist guides' speech. What systematically emerges is the recurring distinction between silent (SPs) and voiced (FPs and VFPs) pauses. In fact, from the functional perspective, SPs efficiently perform both the demarcative (DEM) and the strategic-rhetorical (STR-R) functions, showing a certain degree of awareness of the role of such pauses in emphasizing key words or concepts and attracting the audience's attention. The resulting overlapping between these two functions confirms the distinction provided by O' Shaughnessy (1992) between grammatical and intentional pauses on one side and ungrammatical and hesitation pauses on the other, strengthening the link between grammaticality and willfulness. By contrast, both types of voiced pauses turned out to be rarely used for grammatical (DEM) and rhetorical (STR-R) goals but widely used to express speakers' indecisiveness while planning the discourse. Indeed, almost the totality of voiced pauses performs a hesitative (HES) function. What appears more interesting are the resulting overlapping and relationship between HES and PROG (programmative) functions. Firstly, both functions share the same characteristics of lack of grammaticality and intentionality (ungrammatical and unintended pauses, see O' Shaughnessy, 1992). Secondly, PROG is always included into the HES function; in other words, all pauses classified as PROG are at the same time classified as HES, but not vice versa. Hence, HES can be thought of as a "macro function" carried out by unintentional pauses; it assumes an added feature when it includes the PROG "micro function". Indeed, PROG pauses convey a particular subtype of hesitancy, which turned out to be strongly linked to the punctual search for lexical items, whose access is highly demanding for the speakers, termed Word Searching (WS). As concerns FPs, specific phonetic-acoustic features were found to correlate with the functional distinction between mere HES pauses and combination of PROG (WS) and HES pauses. WS FPs appear to be more disrupting: on the temporal level, they are twice longer than HES FPs; on the intonation level, they result embedded within their co-text (flat and valley F_0 patterns). On the contrary, realizations of HES FPs cause neither temporal nor melodic breaks within the utterance. In conclusion, duration and range level work as relevant phonetic-acoustic cues of an ongoing WS process.

Other results concern the relationship between pauses and gestures. The data on the gestures co-occurring with pauses definitely confirm the deep difference between Silent (SPs) and Voiced pauses (FPs and VFPs taken together). The gestures having a higher and lower communicative import, respectively, show a complementary distribution with the two types of Pauses.

As already hypothesized, VFPs, and even more FPs, may be seen as a cue to an underlying high cognitive effort: when a FP occurs, the speaker is struggling to plan her speech or committed to a word search activity; this is why she needs to freeze in an idle position or to self-assure herself, performing manipulators.

Such freezing of the body in voiced pauses is corroborated by the drop in Amount of Motion described in § 4.3.4.

One more reason for such a cognitive effort may be that, given the asymmetrical communication between the guide and her audience, she may deeply feel the need to be seen as trustworthy, reliable and confident; a FP or a VFP collides with such an image and conversely stands for an uncertain epistemic stance, that is what a guide needs to avoid the most.

On the contrary, SPs are cases in which the Speaker intentionally stops due to communicatively strategic reasons: she gives the time to the Addressee to process what she has just said, which is very clear to herself, and possibly she may perform semantically loaded gestures to make her message clearer or to reinforce the meaning she intends to convey. In doing so, her meanings flow out easily, no cognitive effort is present, and she makes deictic, coded, iconic or metaphoric gestures to clarify her semantic content, and beats to emphasize it by asking for attention.

To sum up, gestures may be seen as a cue of different mental states underlying voiced and silent pauses, respectively: they tend to be semantically empty in voiced pauses, but semantically loaded in SPs.

Overall, this account of the comparison between gestures and pauses might also shed light on the more general issue of the relationship between gesture and speech. Our view seems to support the idea of gesture and speech as an integrated system (McNeill, 1992; Kendon, 2004) in which they are but two different yet related routes for meanings to be expressed: if the mind is engaged in a high cognitive load, both speech and gesture suffer from this and they both freeze, stop, withdraw.

Future works might deepen subtler aspects of the temporal and semantic relations between gestures and pauses, e.g., by taking into account gesture's segmentation into phases and semantic relationship with speech, respectively.

6. *Conclusions*

This work has analyzed the phenomena of disfluency in a particular setting: the speech of tourist guides. In a corpus of guided tours, we have analyzed the silent and unlexicalized filled pauses (voiced pauses) performed by the guides, and the co-occurring gestures. The distribution of pauses and corresponding gestures, and the relations among types of pauses, their functions, and the types of gestures produced reveal that while in voiced pauses the speaker finds herself in a communicative *impasse*, and even gestures are less communicative, in silent pauses the speaker is in total control of her communication, and her gestures too are fully intentional and meaningful. Beside providing theoretical insights about speech disfluencies and concerning the relationship between speech and gesture, the proposed characterization of pauses and concurrent gestures and their functional role in spontaneous speech can be used to build a computational model predicting the occurrence of such elements, given a target text. Such a model can be used to control the synthesis process in text-to-speech systems: this will allow to investigate if the introduction of voiced pauses has an impact on the perceived naturalness of synthetic speech, in the cultural heritage presentation domain. In general, these data will support the development of interactive 3D avatars generating presentations of cultural heritage material on-the-fly.

Bibliography

Alibali, M.W., Kita, S. & Young, A.J. (2000). Gesture and the process of speech production: we think, therefore we gesture. In *Language and Cognitive Processes*, 15(6), 593-613.

Berruto, G., Cerruti, M.S. (2011). *La linguistica. Un corso introduttivo*. Novara: UTET De Agostini.

Betz, S., Carlmeyer, B., Wagner, P. & Wrede, B. (2018). Interactive Hesitation Synthesis: Modelling and Evaluation. In *Multimodal Technologies and Interaction*, 2(1), 9.

Betz, S., Eklund, R. & Wagner, P. (2017). Prolongation in German. In *Proceedings of DiSS 2017 The 8th Workshop on Disfluency in Spontaneous Speech*, Stockholm, Sweden, 18-19 August 2017, 13-16.

Betz, S., Wagner, P. (2016). Disfluent lengthening in spontaneous speech. In *Elektronische Sprachsignalverarbeitung (ESSV) 2016*, Leipzig, Germany, 2-4 March 2016.

Betz, S., Wagner, P. & Schlangen, D. (2015). Micro-structure of disfluencies: Basics for conversational speech synthesis. In *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, 6-10 September 2015, 2222-2226.

Betz, S., Wagner, P. & Vosse, J. (2016). Deriving a strategy for synthesizing lengthening disfluencies based on spontaneous conversational speech data. In *Phonetik und Phonologie*, 12, 19-22.

Blankenship, J., Kay, C. (1964). Hesitation phenomena in English speech: A study in distribution. In *Word*, 20(3), 360-372.

Boersma, P., Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0. 44) [Computer software]. Amsterdam: Institute of Phonetic Sciences.

Capirci, O., Volterra, V. (2008). Gesture and speech. The emergence and development of a strong and changing partnership. In *Gesture*, 8(1), 22-44.

Chafe, W. (1980). Some reasons for hesitating. In Dechert, H.W., Raupach, M. (Eds), *Temporal variables in speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton, 169-180.

Christenfeld, N., Schachter, S. & Bilous, F. (1991). Filled pauses and gestures: it's not coincidence. In *Journal of Psycholinguistic Research*, 20(1), 1-10.

Cibulka, P. (2016). On how to do things with holds: Manual movement phases as part of interactional practices in signed conversation. In *Sign Language Studies*, 16(4), 447-472.

Clark, H.H. (2002). Speaking in time. In *Speech Communication*, 36:1-2, 5-13.

Colletta, J.M., Kunene, R.N., Venouil, A., Kaufmann, V. & Simon, J.P. (2009). Multi-track annotation of child language and gestures. In Kipp, M., Martin, J., Paggio, P., & Heylen, D. (Eds.), *Multimodal corpora*. Berlin/Heidelberg: Springer, 54-72.

Crocco, C., Savy, R. (2003). Fenomeni di esitazione e dintorni: una rassegna bibliografica. In Crocco, C., Savy, R. & Cutugno, F. (Eds.), *API. Archivio di Parlato Italiano*, DVD.

d'Urso, V., Zammuner, V. (1990). The perception of pause in question-answer pairs. In *Bulletin of the Psychonomic Society*, 28, 41-43.

De Mauro, T. (1999). *Gradit*. Torino: UTET.

Duez, D. (1997). Acoustic markers of political power. In *Journal of Psycholinguistic Research*, 26(6), 641-654.

Duncan, S., Pedelty, L. (2007). Discourse focus, gesture, and disfluent aphasia. In Duncan, S., Cassell, J. & Levy, L.T. (Eds.), *Gesture and the dynamic dimension of language*. Amsterdam / Philadelphia: John Benjamins, 269-284.

Eklund, R. (1999). A comparative study of disfluencies in four Swedish travel dialogue corpora. In *Disfluency in Spontaneous Speech Workshop*, Berkeley, California, 1 July 1999, 3-6.

Eklund, R. (2001). Prolongations: A dark horse in the disfluency stable. In *Proceedings of DiSS 2001 Disfluency in Spontaneous Speech,* Edinburgh, Scotland, UK, 29-31 August 2001, 5-8.

Eklund, R. (2004). Disfluency in Swedish human–human and human–machine travel booking dialogues. PhD dissertation, Linköping University Electronic Press.

Ekman, P., Friesen, W.V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. In *Semiotica*, 1,1, 49-98.

ELAN (Version 5.2) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/.

Esposito, A., Marinaro, M. (2007). What pauses can tell us about speech and gesture partnership. In Esposito, A., Bratanić, M., Keller, E., & Marinaro, M. (Eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*. Amsterdam: IOS Press, NATO Publishing Series, 45-57.

Ferreira, F., Lau, E.F. & Bailey, K.G.D. (2004). Disfluencies, language comprehension, and tree adjoining grammars. In *Cognitive Science*, 28, 721- 749.

Gabrea, M., O'Shaughnessy, D. (2000). Detection of filled pauses in spontaneous conversation speech. In *Proceedings, 6th International Conference on Spoken Language Processing (ICSLP 2000*), Beijing, China, 16-20 October 2000, vol. 2, 678-681.

Giannini, A. (2003). Prolungamenti vocalici e vocalizzazioni. In Cosi, P., Magno Caldognetto, E. & Zamboni, A. (Eds), *Voce, canto, parlato. Studi in onore di Franco Ferrero*. Padova: Unipress, 163-172.

Ginzburg, J., Fernández, R. & Schlangen, D. (2014). Disfluencies as intra-utterance dialogue moves. In *Semantics and Pragmatics*, 7(9), 1-64.

Graziano, M., Gullberg, M. (2018). When speech stops, gesture stops: evidence from developmental and crosslinguistic comparisons. In *Frontiers in psychology*, 9, 879.

Hartsuiker, R.J., Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. In *Experimental psychology*, 57, 169-177.

Hieke, A.E. (1981). A content-processing view of hesitation phenomena. In *Language and Speech*, 24(2), 147-160.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M. & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. In *Cognition*, 133, 692-697.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In Key, M.R (Ed.), *Nonverbal communication and Language*. The Hague: Mouton, 207-227.

Kendon, A. (2004). *Gesture. Visible action as utterance.* Cambridge: Cambridge University Press.

Kita, S. (1990). The temporal relationship between gesture and speech: a study of Japanese-English bilinguals. Master's Thesis, Department of Psychology, University of Chicago.

Kita, S. (2000). How representational gestures help speaking. In McNeill, D. (Ed.), *Language and gesture*. Cambridge: Cambridge University Press, 162-185.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. In *Language and cognitive processes*, 24(2), 145-167.

Kita, S., Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. In *Journal of Memory and Language*, 48(1), 16-32.

Kita, S., Alibali, M.W. & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. In *Psychological Review*, 124(3), 245-266.

Kita, S., Davies, T.S. (2009). Competing conceptual representations trigger co-speech representational gestures. In *Language and Cognitive Processes*, 24(5), 795-804.

Kong, A.P., Law, S., Kwan, C.C., Lai, C. & Lam, V. (2015). A Coding System with Independent Annotations of Gesture Forms and Functions during Verbal Communication: Development of a Database of Speech and GEsture (DoSaGE). In *Journal of Nonverbal Behaviour*, 39(1), 93-111.

Krauss, R.K., Chen, Y. & Gottesman, R.F. (2000). Lexical gestures and lexical access: a process model. In McNeill, D. (Ed.), *Language and gesture*. Cambridge: Cambridge University Press, 261-283.

Krauss, R.M., Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In Messing, L., Campbell, R. (Eds.), *Gesture, sign and speech*. New York: Oxford University Press, 93-116.

Levelt, W.J. (1983). Monitoring and self-repair in speech. In *Cognition*, 14(1), 41-104.

Levelt, W.J. (1989). *Speaking: From intention to articulation* (Vol. 1). Cambridge, MA: MIT Press.

Lickley, R.J. (2015). Fluency and Disfluency. In Redford, M.A. (Ed.), *The handbook of speech production*. John Wiley & Sons, 445-474.

Lickley, R.J. (1998). HCRC Disfluency Coding Manual. Edimburgh: Human Communication Research Centre Technical Report TR-100, University of Edimburgh.

Liszkowski, U. (2008). Before L1: a differentiated perspective on infant gestures. In *Gesture*, 8, 180-196.

Maclay, H., Osgood, C.E. (1959). Hesitation Phenomena in Spontaneous English Speech. In *WORD*, 15:1, 19-44.

Mahl, G.F. (1956). Disturbances and silences in the patient's speech in psychotherapy. In *The Journal of Abnormal and Social Psychology*, 53(1), 1-15.

Mayberry, R.I., Jaques, J. (2000). Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In McNeill, D. (Ed.), *Language and gesture*. Cambridge: Cambridge University Press, 199-214.

McNeill, D. (1992). *Hand and mind*. Chicago: The University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

Mondada, L. (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. In *Discourse Studies*, 9(2), 195-226.

Morsella, E., Krauss, R.M. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture-speech theories. In *Journal of Psycholinguistics Research*, 34(4), 415-427.

Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family? In Müller, C., Posner, R. (Eds.), *The Semantics and Pragmatics of everyday Gestures*. Berlin: Weidler.

Nobe, S. (2000). Where to most spontaneous representational gestures actually occur with respect to speech? In McNeill, D. (Ed.), *Language and gesture*. Cambridge: Cambridge University Press, 186-198.

Nobili, C (2019). *I gesti degli italiani*. Roma: Carocci.

Origlia, A., Savy, R., Cataldo, V., Schettino, L., Ansani, A., Sessa, I., Chiera, A. & Poggi, I. (2019). Human, all too human. Towards a disfluent Virtual Tourist Guide. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization,* Larnaca, Cyprus, 9-12 June 2019, 393-399.

Origlia, A., Savy, R., Poggi, I., Cutugno, F., Alfano, I., D'Errico, F., Vincze, L. & Cataldo, V. (2018). An Audiovisual Corpus of Guided Tours in Cultural Sites: Data Collection Protocols in the CHROME Project. In *Proceedings of the 2018 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage*, Castiglione della Pescaia, Italy, 29 May 2018, 1-4.

O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech. In *Proceedings: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92)*, San Francisco, USA, 23-26 March 1992, 521-524.

Pine, K.J., Bird, H. & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. In *Developmental Science*, 10(6), 747-754.

Poggi, I. (2007). *Mind, hands, face and body. A goal and belief view of multimodal communication.* Berlin: Weidler.

Ragsdale, J.D., Fry Silvia, C. (1982). Distribution of kinesic hesitation phenomena in spontaneous speech. In *Language and Speech*, 25(2), 185-190.

Schnadt, M.J., Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, (Vol. 28, No. 28), Vancouver, Canada, 26-29 July 2006, 750-755.

Senigalliesi, S., Sparano, A., Schettino, L., Savy, R., Dell'Orletta, F., Lubello, S. & Basile, G. (2018). Frequenze lessicali e sintattiche nei testi scritti e orali della descrizione artistico/architettonica: analisi sul corpus delle Certose Campane. Poster presented at *Gruppo di Studio sulla Comunicazione Parlata* (GSCP 2018), Università degli Studi di Napoli "L'Orientale", 12-14 dicembre 2018.

Seyfeddinipur, M. (2006). Disfluency: Interrupting speech and gesture. PhD thesis, Radboud University Nijmegen, Nijmegen.

Shriberg, E. (1994). Preliminaries to a theory of speech disfluencies. PhD dissertation, University of California, Berkeley.

Shriberg, E. (2001). To "err" is human: Ecology and acoustics of speech disfluencies. In *Journal of the International Phonetic Association*, 31, 153-169.

Shriberg, E.E., Lickley, R.J. (1993). Intonation of clause-internal filled pauses. In *Phonetica*, 50.3, 172-179.

Sloetjes, H., Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. In *Proceedings of the 6th international Conference on Language Resources and Evaluation (LREC 2008)*, Marrakesh, Marocco, 28-30 May 2008.

Stam, G., Tellier, M. (2017). The sound of silence: the functions of gestures in pauses. In Breckinridge Church, R. Alibali, M.W. & Kelly, S.D. (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating*. Amsterdam: Benjamins, 353-377.

Strangert, E. (2003). Emphasis by pausing. In *Proceedings of the 15th international congress of phonetic sciences*, Barcelona, Spain, 3-9 August 2003, 2477-2480.

Swerts, M. (1998). Filled pauses as markers of discourse structure. In *Journal of pragmatics*, 30(4), 485-496.

Viola, I., Madureira, S. (2008). The roles of pause in speech expression. In *Proceedings of the 4th Conference on Speech Prosody*, Campinas, Brazil, 6-9 May 2008, 721-724.

Voghera, M. (2017). *Dal parlato alla grammatica: costruzione e forma dei testi spontanei.* Roma: Carocci.

Yasinnik, Y., Shattuck-Hufnagel S. & Veilleux N. (2005). Gesture marking of disfluencies in spontaneous speech. In *The 4th Workshop on Disfluency in Spontaneous Speech*, 173-178.

Zellner, B. (1994). Pauses and the temporal structure of speech. In Keller, E. (Ed.), *Fundamentals of speech synthesis and speech recognition.* Chichester: John Wiley, 41-62.

Acknowledgments

Work funded by the Italian National Project PRIN “Cultural Heritage Resources Orienting Multimodal Experiences (CHROME)” (#B52F15000450001).

DOI: 10.17469/O2106AISV000014

CLAUDIA ROBERTA COMBEI, GIOVANNA MAROTTA

Where do accents come from? Factors affecting the degree of foreign-accented Italian[1]

Following similar research on non-native speech, in this paper we assessed the degree of perceived foreign-accented Italian. This is an extensive study on 6 accents (i.e. English, German, French, Romanian, Spanish, and Russian, in comparison to L1 Italian), supported by a database of 42 speakers whose performance was evaluated by 288 Italian listeners, using a six-points scale. Such an approach allowed us to perform comprehensive descriptive and inferential analyses, the outcomes of which suggested that while having comparable levels for all variables, some accents (i.e. German and Spanish) are perceived as being stronger than others. Additionally, it seems that the degree of perceived foreign accent may be predicted by speaker-dependent variables generally ascribable to the quality and quantity of input received (i.e. presence/absence of pronunciation training) and proficiency. Finally, the samples of read speech were perceived as more accented than those of spontaneous speech, presumably as a result of the fact that speakers might not be fully familiar with written Italian.

Key words: perceived foreign accent, foreign-accented Italian, non-native Italian.

1. *Introduction*

The differences between native and foreign pronunciations have been a topic of constant research interest in the community of Second Language Acquisition (SLA). Over the last few decades, scholars have described, analysed, and explained various phenomena observable in the perception and production of the language being learned (L2[2]).

It has long been established that after their birth, healthy infants can perceive, discriminate, and learn the sounds and prosody of all languages (Eimas, Siqueland, Jusczyk & Vigorito, 1971; Grieser, Kuhl, 1989). Yet, by the time children have reached one year old, they have already become aware of the speech contrasts that are phonologically relevant in their native languages (L1s), and they have started to neglect the rest (Werker, Gilbert, Humphrey & Tees, 1981). Oftentimes, it has been suggested that the older one gets, the harder it becomes to achieve native-like pronunciation skills in an L2.

[1] Authorship note: this paper is the result of a collaboration as it discusses data taken from Combei's (2019) Ph.D. dissertation, supervised by Marotta. For academic purposes, Combei is held responsible for §2.2, § 2.3, §3, §4, and §5, while Marotta for §1, §2.1, and §2.2.

[2] Throughout this paper we will use second language (L2) to refer to any language (i.e. second, third, fourth, etc.) different from the native language (L1).

However, rejecting the assumption that a loss of neural plasticity and the neurofunctional reorganisation result in foreign-accented speech, and instead following Moyer (2013), we believe that this age-related issue occurs fundamentally because L2 learners, unlike L1 learners, start acquiring the L2 with a consolidated L1 phonology, involving an established metalinguistic awareness, but at the same time also a possible source of interference. Lövdén, Wenger, Mårtensson, Lindenberger & Bäckman (2013) and Flege, Wayland (2019) criticise Lenneberg's (1967) assumption that neural maturation could block L2 speech learning, pointing out that, instead, learners cannot prevent the interaction between their L1 and L2 subsystems. Moreover, not all late learners manage or aspire to a native-like proficiency in the L2. Some might be constrained or simply want to remain grounded in the culture of their home country and therefore they will 'continue to use other languages in their daily lives' (Gut, 2007: 75). Consequently, this might unavoidably decelerate the process of achieving native-like pronunciation.

More realistically, recent research has proposed that, in time, the late learners' sensitivity towards subtle phonetic cues of an L2 is not entirely lost, but rather it is 'obscured' (Piske, 2008). As a matter of fact, older learners might actually be able to learn phonological patterns even faster than young learners (Moyer, 2013). On a similar note, Flege (2009) suggests that 'input', rather than age, is responsible for how learners sound when they speak a foreign language.

Anyway, simply put, in case learners of an L2 do not manage to sound like a native speaker they are said to have a 'foreign accent'.

2. *Topic and goals*

2.1 Foreign accent matters

In fact, on the course towards achieving the much desired native-like linguistic proficiency, some hurdles are more difficult than others. It is reasonable to think that when learning an L2, a great difficulty lies in accommodating perception and production mechanisms, so as to take into account new patterns or contrasts, but also similar features. Many theories indicate that especially in the first stages, learners may apply the phonology of their L1 or that of other languages they acquired before the target L2 (Flege, 1995; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelsonm, 2008). Native speakers would perceive these phenomena as non-native or foreign, thus, the name 'foreign accent'. Following Munro (1998) and Jilka (2000), but abandoning the idealistic concept of a 'standard accent'[3], Combei (2019: 65) proposes a more encompassing definition of foreign accent, namely 'a non-pathological manner of speaking in a foreign language that deviates from the pronunciation generally accepted as native by the native speakers of that language'.

[3] Following Lippi-Green (1997), Jenkins (2000), and Derwing (2010) in general and Bertinetto, Loporcaro (2005) for the specific case of the Italian language, we believe that the standard accent is merely a utopian concept reflecting some actors' and some television presenters' pronunciation.

Foreign accent has evolved from initially being treated 'largely as a theoretical puzzle' to a topic of 'great public interest' (Moyer, 2013: i). This may be explained also in light of the fact that, in our globalised world, where many people are active speakers of at least one L2[4], the likelihood of hearing foreign accents has increased. Generally, native speakers understand if their language is spoken with a foreign accent; and in fact, accent-detection and accent-rating studies have demonstrated that people recognize non-native pronunciations in a wide variety of speech samples, including short excerpts (Flege, 1984).

Nonetheless, this does not mean that attitudes towards foreign accents are always neutral. As a matter of fact, native speakers have different reactions to non-native speech, that could range from positive to negative (Bianchi, Calamai, 2012; Levis, Moyer, 2014; Calamai, 2015).

On another note, speaking with a foreign accent may lead to unintelligibility, which affects communication, but it can also have implications at personal, social, communicative, and professional levels. Furthermore, in recent years, accents have posed challenges to Automatic Speech Recognition (ASR) systems; in fact, it has been shown that accent is one of the factors that reduces the performance of ASR (Behravan, Hautamäki & Kinnunen, 2015).

2.2 A brief review on the perception of foreign-accented Italian

By the beginning of 2000, the majority of empirical research on foreign accent had been conducted on English. However, in recent years, various studies focused on L2 Italian, as well. In this subsection we are presenting a brief review of some empirical studies that quantified the degree of perceived foreign accent in Italian[5] (cf. Combei, 2019, for a complete up-to-date review on this topic).

In their pioneer study on accent rating, Marotta, Boula de Mareüil's (2010) study show, among other things, the two German speakers were perceived as having a stronger accent than that of the French and the Spanish informants, both in samples of read and spontaneous speech.

De Meo and her colleagues have recently published various studies on the perceived foreign accent. For instance, Vitale, De Meo & Pettorino (2012) tested the reactions of Italian listeners in a perceptual experiment aimed primarily at evaluating the levels of persuasiveness of non-native speech. While listening to eight radio ads recorded by native speakers of Italian, Chinese, French, and Russian, 164 listeners had to indicate whether the speakers were Italian or foreign, to assess the degree of foreign accent and persuasiveness of each person, and to mention the features that influenced the rating. The authors report that five out of six L2 speakers were rated as accented by almost all listeners. Additionally, their accents were judged as

[4] According to United Nations' International Migration Report (United Nations, 2017), the number of international migrants worldwide reached 258 million in 2017, displaying an ever-growing trend.

[5] Keeping in mind the scope of this paper, we are presenting only studies that examined the Italian listeners' perception of foreign accent. Theoretical research as well as acoustical studies on non-native Italian speech are not discussed.

'strong'. Regarding persuasiveness – correlating to flat and less variable pitch contour – only the French female speaker received positive ratings.

In two similar studies, Vitale, Boula de Mareüil, & De Meo (2014; 2017) assessed the prosody in yes/no questions and statements produced by twelve Chinese learners and four Italians. Instrumental analyses were performed on the utterances examined, by extracting various prosodic features. The results show that prosody may account for differences between elementary, intermediate, and advanced Chinese learners of Italian. Additionally, the perceptual experiment – using prosody transplantation – conducted on 40 Italian listeners revealed that non-native segments with a native Italian prosody are judged as less accented than native Italian segments with a non-native prosody. Finally, the importance of prosody was confirmed also for discriminating between questions and statements.

Similar experiments were performed by Pellegrino and Dellwo (2015); their paper examined the role of rhythmic cues, from the amplitude envelope (ENV), on the perception of non-native Italian. As part of that study, the authors asked ten Italian listeners to assess the native-likeness in two different types of utterances: those produced by a German speaker of L2 Italian and by a native speaker of Italian, and those synthetically manipulated with either the ENV of a German learner of Italian and that of native speaker of Italian. The results show that rhythmic features in the speech ENV affect the listeners' perception of foreignness.

2.3 Goals

This paper, just like those summarised above, assesses the degree of perceived foreign-accented Italian in comparison to native varieties. However, differently from previous studies on spoken L2 Italian, we are focusing on six foreign accents (i.e. English, German, French, Romanian, Spanish, and Russian) and we are investigating whether speaker- and style-related factors affect the degree of accentedness. The database used for this corpus consists in 42 speakers and 288 listeners (see §3.2 and §3.3).

Our hypothesis is that the age of onset is not the only variable hindering the path towards native-like pronunciation. To our mind, it is more reasonable to assume that foreign accent may depend on more than one factor, a composite variable incorporating the quality and quantity of L2 exposure and use, possibly but not necessarily, the length of stay in the L2-speaking country, and L2 (pronunciation) training (Combei, 2019). In fact, Flege (2009) suggested the fundamental role of 'input'[6] on the attainment of L2 pronunciation skills. The approach we propose here may be advantageous in explaining and understanding the production and perception of non-native Italian speech from sociolinguistic and educational perspectives. Therefore, the results of this study may be beneficial in the field of second language acquisition, especially in teaching Italian pronunciation to foreigners.

[6] According to Flege (2009: 175) input may be defined as 'all L2 vocal utterances the learner has heard and comprehended, including his own, regardless of whether these utterances have been produced correctly by L2 native speakers, or incorrectly by other non-native speakers of the L2'.

3. *The study*

3.1 The experiment

This study assesses and interprets the degree of foreign-accented Italian. In order to trigger quantitative judgements expressed by monolingual[7] native Italian listeners, an accent-perception experiment was carried out[8]. We built an *ad-hoc* computer-based task that was conducted over the Internet on a WordPress[9] platform from May 2017 through January 2018. It was distributed to various interest groups via mailing lists. This allowed us to reach a solid, unbiased, and varied pool of listeners from all over Italy (see §3.3).

The first section of the experiment was compulsory and qualifiable for further rounds. The questions were designed to collect background information on the listeners. We sampled for the region of origin, the sex, the L1 (only monolingual Italian speakers were admitted), the age range (i.e. 18-30, 31-45, 46-65), the highest level of education achieved (i.e. high-school, B.A./B.Sc., M.A./M.Sc./M.B.A.), occupation (i.e. student, employee, unemployed), proficiency level in foreign languages (i.e. English, German, French, Romanian, Spanish, and Russian), the degree of familiarity with the foreign accents investigated (i.e. English, German, French, Romanian, Spanish, and Russian), and experience in the field of linguistics (i.e. having attended B.A. and/or M.A. courses in Linguistics).

The participants that qualified for the experiment were reminded to activate the audio on their computers and to use headphones during the experimental session. A detailed description of the task and a short trial were provided as well.

In the actual accent rating experiment, the Italian listeners were asked to assess the degree of foreign accent in the speech samples they were listening to, using a 6-gradients Likert scale (Likert, 1932). The top of the rating scale (i.e. "0") was labelled with "no foreign accent – native speaker of Italian", whereas the bottom (i.e. "5") was labelled with "very strong accent". The in-between values were the following: "1" – "mild accent", "2" – "moderate accent", "3" – "marked accent", "4" – "strong accent". The type of scale and the number of gradients complies with those normally used in similar accent perception experiments (cf. Piske, MacKay & Flege, 2001, for a survey on methods and tools employed in measuring foreign-accented speech).

---

[7] The term "monolingual" is used here to refer to speakers that are neither early nor late "perfect bilinguals"; however, these speakers may be proficient in other languages without being considered "bilingual" (Grosjean, 2010).

[8] This paper presents the results of one of the three tasks of a large accent-perception experiment. The other two tasks were: accent recognition (1) and the identification of features of foreign-accented speech (2). However, it is worth mentioning that for each task original speech samples produced by new speakers were used (i.e. the listener had not heard them in the other tasks). Moreover, in the experiment described in this paper the listeners did not know the L1 of the speakers they were listening to.

[9] WordPress is a content management system based on a Hypertext Pre-processor and a MySQL. A detailed description of this platform is available at this webpage: https://wordpress.com (WordPress, 2019, accessed on the 17th of March 2019).

3.2 Speakers and stimuli

The stimuli were automatically randomised for each experimental session and they could be played only twice, in order to avoid, or at least to reduce, responses and associations that could have emerged due to bias or the authors' intentions.

In order for the stimuli to meet balancedness and representation criteria, when selecting the speakers, we controlled for the following variables: the speaker's mother-tongue, the occupation (all speakers were enrolled as regular or exchange students in degree, Ph.D., or specialization programmes in Bologna), the age (ranging from 20 to 30 years old), the sex, the self-assessed[10] proficiency level in Italian (ranging from B1 to C2 levels, based on *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment-CEFR*[11]), the length of stay in Italy (i.e. less than 12 months, less than 24 months, more than 24 months), the age of onset (i.e. the first exposure to the Italian language: infancy, adolescence, adulthood)[12], the predominant language learning method (i.e. purely naturalistic and guided-naturalistic), and the presence or absence of specific pronunciation training.

The accents considered here are English, German, French, Romanian, Spanish, and Russian, compared to native varieties of Italian elicited by speakers from the Centre and the North of the country. Ten-to-fifteen-seconds-samples of read and spontaneous speech produced by a balanced set of 42 speakers – 6 for each L1 – were included in the perceptual experiment. The read speech samples were elicitated with the support of a newspaper article[13], while in order to obtain spontaneous speech, we asked speakers to talk about their most recent holidays (cf. Combei, 2017 and Combei, 2019, for a complete description of the speech samples).

The stimuli were previously validated by five experts[14] who assessed the degree of foreign accent (inter-rater reliability: ICC (A, 1) = 0.76[15]; p-value 0; 95% – confidence interval).

Considering that one of the hypotheses we wish to test is whether speaker- and style-related factors have an effect on the degree of foreign accent, we rigorously

[10] Considering the number of speakers involved and the scope of this study, we rely on the informants' self-assessed level of Italian, as it would not have possible to actually test the proficiency prior to the experiment.

[11] A description of *CEFR* is available online at this webpage: https://europass.cedefop.europa.eu/it/resources/european-language-levels-cefr (CEDEFOP, 2018, accessed on the 6th of April 2019).

[12] The age of onset (i.e. the first exposure to Italian) coincides with the age of arrival for informants who learned Italian naturalistically; for all the other informants it refers to the age they started to learn Italian at school. In either case, the age labels correspond to the following values: "infancy" between 3 and 12 years, "adolescence" between 13 and 18, and "adulthood" 18 or older (cf. Combei, 2017).

[13] Although the lexical frequency was not controlled, the excerpt is linguistically rich, so as to enable further acoustic and perceptual analyses (cf. Combei, 2017).

[14] The expert informants were 1 full professor in Linguistics and 4 Ph.D. students in Linguistics from the University of Pisa, all having a solid background on the topic of this study. This ensured the choice of a balanced dataset, highly necessary for this kind of experiment.

[15] According to Cicchetti's (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.76 is considered 'excellent'.

controlled for all the variables mentioned at the beginning of this section (i.e. the corpus is well-balanced for all variables and at all levels). Below, we are summarising the speakers' characteristics.

As far as the sex is concerned, 61.90% of the stimuli used for the accentedness rating task were uttered by female voices, while 38.10% by male voices. Besides the native speakers (i.e. 14.29% of the corpus) that were exposed to Italian from birth, for all other informants, the age of onset is distributed as follows: infancy, 21.43%; adolescence, 23.81%; and adulthood, 40.47%. Regarding the method used for learning Italian, 50.00% of the non-native speakers acquired it naturalistically, and 50.00% using both guided and naturalistic approaches (cf. Klein, 2003, for the definition and characteristics of these two acquisition paths). At the dataset level, 38.89% of the learners were given specific pronunciation training, while 61.11% were no. As concerns the length of stay in Italy, 44.44% of the speakers had spent less than a year there by the time they were recorded, 19.44% between one and two years, and 36.12% for more than two years, and generally less than five years. In this experiment we used only samples produced by speakers having at least a B1 level in Italian. The proficiency level among non-native speakers is balanced for each L2 and distributed as follows: B1 (30.56%), B2 (22.22%), C1 (19.44%), and C2 (27.78%).

3.3 Listeners

An adequate number of listeners for each of the levels of the variables mentioned in §3.1 was collected, even if they were not evenly distributed.

For instance, out of 288 participants, 82.99% were women and 17.01% men. As far as the age groups are concerned, 82.02% of the listeners were in the first group (i.e. 18-30 years), 13.20% in the second group (i.e. 31-45 years), and 2.78% in the third group (i.e. 46-65 years). This is not surprising since the experiment was distributed on the internet; thus, it indirectly guaranteed major visibility to younger respondents. Therefore, the majority of the participants (i.e. 61.80%) were bachelor's students and their highest level of education was a high-school diploma; there followed, with similar percentages, the master's students that had a bachelor's degree and the workers that had a master's degree. Another variable that emerged from the sampling is the region where listeners had spent most of their lives. We achieved a fairly well-stratified sample; there were listeners from all Italian regions, except for Molise. Most of the times, the number was directly proportional to the number of inhabitants of that region, all other variables considered. As a matter of fact, most respondents were from Lombardy (i.e. 17.70%), Piedmont (i.e. 11.11%), Sardinia (i.e. 10.07%), Sicily (i.e. 9.72%), and Apulia (i.e. 7.29%). There were relatively fewer informants from Valle D'Aosta, Basilicata, and Marche. We were also interested in the listeners' background in Linguistics. The values for the levels of this variable are well-balanced: those that had attended B.A. and/or M.A. classes in Linguistics represent 59.37% of the entire sample, while those that replied 'no' represent 40.63%. Additionally, we asked participants to indicate their proficiency level in foreign languages and their familiarity with the accents investigated in this study, by using a "0" through "5" scale[16]. As one may expect, everyone (i.e. 99.65%) spoke English, at least to some extent. Similarly, almost 80% of the respondents had some knowledge of French or Spanish. On the contrary, very few respondents spoke or understood Russian (i.e. 20.14%) or Romanian (i.e. 4.86%). Not surprisingly, young people claimed that they had a better knowledge of foreign languages than the older ones (see Table 1), especially regarding English, Spanish, and French.

Table 1 - *The listeners' language proficiency as a function of the age group*

Language	Age group: 18-30 Average level (±S.D.)	Age group: 31-45 Average level (±S.D.)	Age group: 46-65 Average level (±S.D.)
English	3.85 (±0.95)	3.29 (±1.14)	3.25 (±1.17)
German	1.10 (±1.48)	0.78 (±0.32)	1.37 (±1.68)
French	1.90 (±1.44)	1.76 (±1.34)	1.50 (1.07)
Spanish	1.99 (±1.56)	1.97 (±1.62)	0.75 (±1.04)
Romanian	0.08 (±0.39)	0.19 (±0.69)	0.00 (±0.00)
Russian	0.45 (±0.98)	0.11 (±0.51)	0.13 (±0.35)

Finally, we were interested in the participants' familiarity with the six foreign accents under consideration. Listeners rated themselves with the scale described earlier. Overall, 98.96% of them had at least some familiarity with the English accent; there followed the French accent (92.36), the Spanish accent (88.54%), and the German accent (78.12%). At the opposite side, there were the Russian and the Romanian accents, both familiar to less than 50.00% of the listeners. Considering that Romanian- and Russian-speaking people represent some of the largest communities in Italy (ISTAT, 2018), these data may seem unexpected. However, we believe this could be explained in terms of the characteristics of our sampling set; it includes a lot of young people that might have not met native speakers of Romanian and/or Russian. The average levels of familiarity with the accents are detailed in Table 2.

Table 2 - *The listeners' accent familiarity as a function of the age group*

Language	Age group: 18-30 Average level (±S.D.)	Age group: 31-45 Average level (±S.D.)	Age group: 46-65 Average level (±S.D.)
English	3.86 (±1.15)	3.32 (±1.28)	3.62 (±0.92)
German	2.30 (±1.77)	1.87 (±1.66)	3.25 (±1.16)
French	2.76 (±1.53)	2.79 (±1.47)	2.88 (1.25)
Spanish	2.81 (±1.67)	2.97 (±1.60)	2.38 (±1.77)
Romanian	1.08 (±1.36)	1.45 (±1.52)	1.00 (±0.76)
Russian	1.54 (±1.56)	1.24 (±1.38)	1.25 (±1.04)

[16] "0" means that the informant has no knowledge of that language, while, on the contrary, "5" means that she/he is highly proficient.

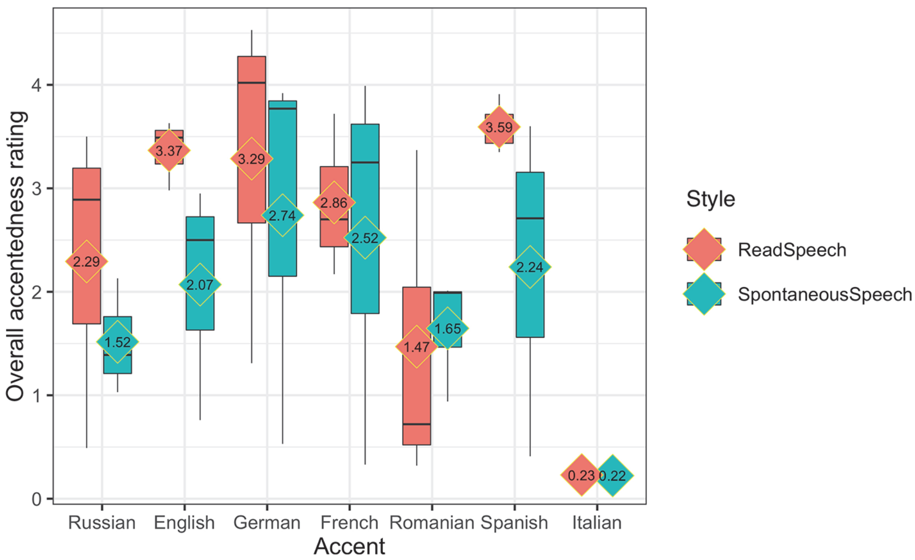
The unbalanced distribution of listeners is inevitable in this type of experiment, since, in fact, one cannot fully control crowdsourcing data collection. Nevertheless, we reached an adequate number of participants for each variable and each variable level.

4. *Predicting the degree of foreign accent*

4.1 Models built around speaker- and style-dependent factors

Before analysing the results, an inter-rater reliability test was performed on the listeners' judgements; this yielded a positive outcome: ICC (A,1) = 0.68[17]; p-value: 0; 95% – confidence interval, suggesting that most of the times there was agreement between the listeners' ratings. A chart aggregating the scores for all accents as a function of the style is shown in Figure 1.

Figure 1 - *Accentedness ratings for each accent as a function of the style*



Following Piske et al. (2001), we were interested in assessing which factors contributed to the perception of accentedness in Italian. For such purpose, after a treatment coding of the categorical variables, we fitted Linear regression models in R (R Core Team, 2018), where we tested the effect of all the variables under control on the accentedness score (i.e. the dependent variable). Since the levels of some speaker-dependent variables regarding native speakers of Italian generated aliased coefficients in the models due to perfect multicollinearity, we excluded this group from the analyses presented below. Not surprisingly, in fact, Italians were perceived

[17] According to Cicchetti's (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.68 is considered 'good'.

as being native speakers and they were assigned ratings around "0" (see Figure 1). Thus, their data would not have added any relevant information to the models.

We started with a model that was built around all independent variables (i.e. the style; the speaker's sex, L1, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian), while the response variable was the accentedness score received by the speakers (Model 1 in Table 3). This model could estimate 77.00% of the variation in the accentedness ratings. However, since the sex variable was the weakest predictor for the degree of foreign accent (p-value: 0.86), we dropped it. In fact, Model 2 (see Table 3) predicts 78.1% of the variation. Finally, in Model 3 (see Table 3) we excluded also the second weakest factor, namely the age of onset (p-value: 0.68); even if there is a difference between the ratings received by the three groups (i.e. infancy, adolescence, and adulthood), it is not statistically significant. This finding actually might suggest that starting to learn Italian at a young age – especially during adolescence – does not necessarily guarantee native-like productions. Therefore, in the end, Model 3 explains 79.3% of the variation in the accentedness ratings.

Table 3 - *Summary of the three multiple linear regression models*

	Model 1	Model 2	Model 3
Dependent variable	accentedness rating	accentedness rating	accentedness rating
Independent variables	style, sex, L1, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian	style, L1, age of onset, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian	style, L1, method of learning Italian, pronunciation training, length of stay in Italy, and proficiency level in Italian
Observations	36	36	36
Residual standard error	0.626 on 19 degrees of freedom (sigma: 0.254)	0.611 on 20 degrees of freedom (sigma: 0.247)	0.594 on 22 degrees of freedom (sigma: 0.240)
Multiple R-squared	0.875	0.875	0.870
Adjusted R-squared	0.770	0.781	0.793
F-statistic	8.33 on 16 and 19 degrees of freedom, p-value: 1.678e-05	9.34 on 15 and 20 degrees of freedom, p-value: 5.346e-06	11.34 on 13 and 22 degrees of freedom, p-value: 6.578e-07

The assessment of the linear model assumptions was performed with a global test on 4 degrees of freedom (level of significance: 0.05). The assumptions for Global Stat, Skewness, Kurtosis, Link Function, and Heteroscedasticity are acceptable. The fit information (i.e. Anova Type II Test) for Model 3 is shown in Table 4.

Table 4 - *Anova Type II Test for Model 3*

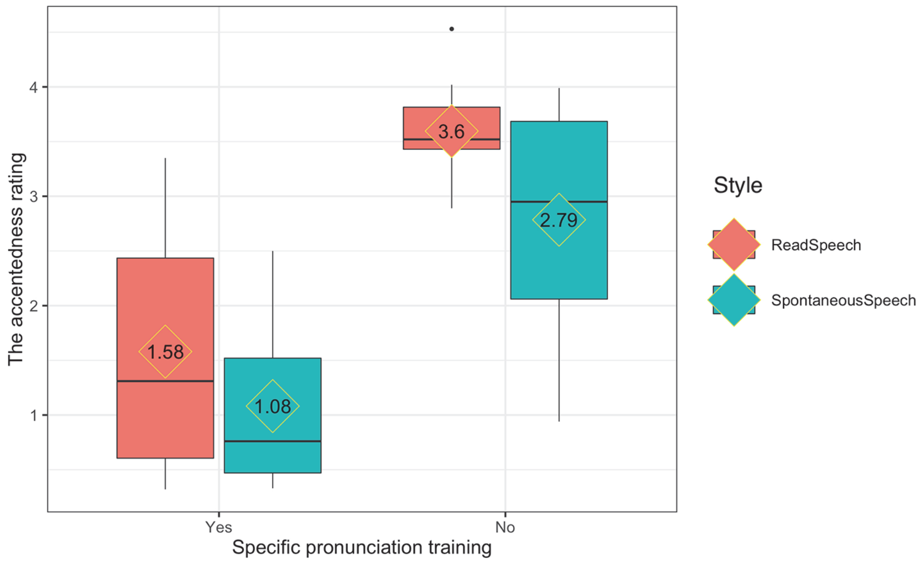
	Sum of squares	Degrees of freedom	F value	Pr(>F)
L1	8.14	5	4.61	0.005**
Style	1.66	1	4.70	0.04*
Italian learning method	1.53	1	4.30	0.06
Specific pronunciation training	4.36	1	12.33	0.001**
Length of stay in Italy	1.29	2	1.82	0.18
Proficiency level in Italian	5.56	3	5.24	0.01*
Residuals	7.78	22	-	-

Generally, the samples of read speech were perceived as more accented than those of spontaneous speech (p-value: 0.04), presumably as a result of the fact that speakers might not be fully familiar with written Italian and because some reading errors might have arisen due to conflicting orthographic norms between the speakers' L1 and Italian (Wottawa, Adda-Decker, 2016). At the same time, in spontaneous speech, the speaker would tend to use familiar and less problematic constructions. This finding is consistent with previous research on foreign accent; in fact, Oyama's (1976), Thompson's (1991), and Kolly, Dellwo (2013) also found that read speech was perceived to be more strongly accented than spontaneous speech.

All other grouping variables and levels being equal and comparable (i.e. corpus balancedness and representativity), most of the times, Romanian speakers were rated with "no accent" or "mild accent" scores. Romanian and Italian share indeed a high number of the segmental and suprasegmental features, but it is still unclear whether the "good" ratings received by Romanian speakers may be ascribable to this vicinity. Next, the samples produced by German speakers were perceived as the most accented. They were followed in this order by Spanish, French, English, and Russian. This is consistent with the SLA literature that has linked pronunciation deviations, and therefore, foreign accent, to blocking mechanisms ascribable to L1 (Flege, 1995). Therefore, as shown in Table 4, the speaker's L1 seems to predict the degree of foreign accent (p-value: 0.005).

Furthermore, the speakers who received specific pronunciation training were perceived as less accented than those who did not (p-value: 0.001). A clearer representation of this evidence is shown in Figure 2. This finding could advocate for shifting the view towards the role of input in SLA (Flege, 2009).

Figure 2 - *Accentedness ratings as a function of the style and pronunciation training*



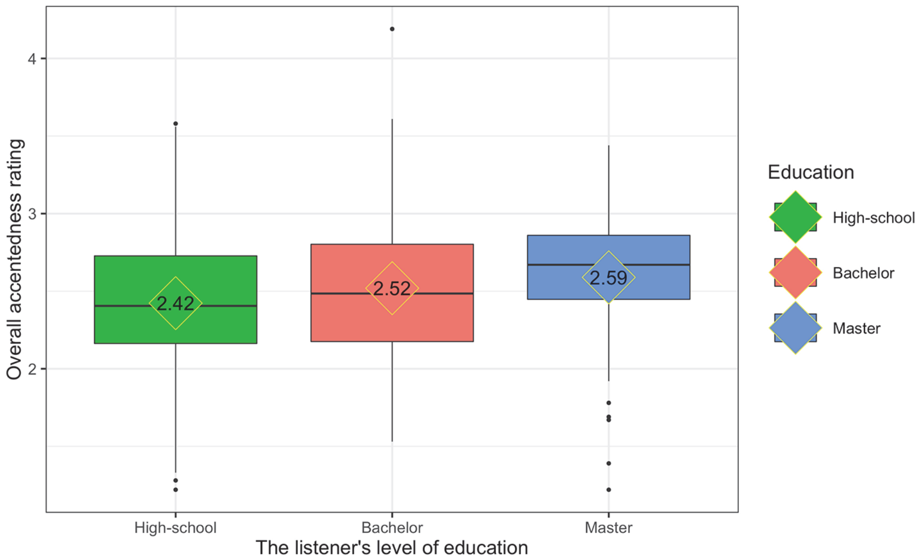
However, as reported in Table 4, having learned Italian in a specific learning setting is not generally associated to a better pronunciation score (p-value: 0.06). Moreover, the length of stay in Italy does not seem to have any role in predicting the accentedness (p-value: 0.18), probably because the difference between the levels is too small (i.e. 12 months) to display any significant effect. Nevertheless, the speakers that stayed in Italy for more than 2 years received better pronunciation ratings than the other two groups. A similar rationale behind the rather feeble but still statistically significant effects of the proficiency level in Italian (p-value: 0.01), together with the fact that it may not be a fully reliable predictor, since speakers self-assessed their performance. Nevertheless, the speakers that claimed they had a C2 level in Italian received indeed good ratings, suggesting that their pronunciation is close to that of a native speaker.

4.2 Models built around listener-dependent factors

After a treatment coding of the categorical variables, we verified the strength of the effect of listener-dependent factors (i.e. sex, age group, education level, region of origin, occupation, background in linguistics, proficiency in foreign languages, and familiarity with the accents investigated) on the score they assigned to speech samples, by attempting to fit a Linear regression model. Its results may be summarised as follows: $F(26,261) = 1.07$, $p = 0.37$; $R^2 = 0.096$; Adj. $R^2 = 0.006$, Residual standard error: 0.46 on 261 degrees of freedom.

Given the model complexity, meaningful results were obtained only for the education variable (see Figure 3), namely, the higher the listener's education level, the more severely they would judge the speakers' degree of accentedness (Sum of squares: 1.53, Degrees of freedom: 2, F value: 3.6087, p-value: 0.03). Even by dropping several factors, such as the listeners' region of origin, age group, sex, and background in Linguistics, the model did not improve its predictive strength. In fact, the results of the final model, having only the listeners' education level and occupation as independent variables, may be summarised as follows: $F(4,283) = 2.35$, $p = 0.05$; $R^2 = 0.03$; Adj. $R^2 = 0.02$, Residual standard error: 0.46 on 283 degrees of freedom.

Figure 3 - *Accentedness rating as a function of the listener's education level*



5. *Conclusions and further research*

Based on our data, the degree of foreign-accented Italian seems to be determined especially by the following factors: the speaker's L1, the presence or absence of specific pronunciation training during the learning path, the proficiency level in Italian, and the style. Therefore, through the ears of Italian listeners, some accents are stronger than others (German accent – strong vs. Romanian accent – mild). Next, receiving pronunciation training seems to improve the learner's accent, making it sound more native-like. Moreover, C2 levels in Italian generally correspond to good pronunciation skills. Read speech was perceived as more accented than spontaneous speech, probably as a result of reading issues arisen due to different orthographic norms. In fact, some speakers might not have had a complete familiarity with written Italian. Listener-dependent factors do not seem to have a significant role in predicting accentedness ratings.

However, in order to confirm all the afore-mentioned associations at a larger scale, in future research, other accents should be investigated as well. Finally, it would be desirable to sample for and test the role of sociopsychological factors, such as attitude, motivation, and personality on the degree of perceived foreign accent.

Bibliography

Behravan, H., Hautamäki, V. & Kinnunen, T. (2015). Factors affecting i-vector based foreign accent recognition: A case study in spoken Finnish. In *Speech Communication*, 66, 118-129.

Bertinetto, P.M., Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. In *Journal of the International Phonetic Association*, 35(2), 131-151.

Bianchi, F., Calamai, S. (2012). Voci italiane e straniere a confronto. Indagine sugli stereotipi associati agli accenti stranieri. In *Cambio*, 4, 149-164.

Calamai, S. (2015). Between linguistics and social psychology of language: the perception of non-native accents. In *Studi e Saggi Linguistici*, 53(2), 289-308.

CEDEFOP (2018.). Livelli europei - Scheda per l'autovalutazione. https://europass.cedefop.europa.eu/it/resources/european-language-levels-cefr/ Accessed 06.04.2019.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. In *Psychological Assessment*, 6(4), 284-290.

Combei, C.R. (2017). CorAIt – A non-native speech database for Italian. In Basili, R., Nissim, M. & Satta, G. (Eds.), *Proceedings of the Fourth Italian Conference on Computational Linguistics,* Rome, 11-13 December 2017. Torino: Accademia University Press, 113-118.

Combei, C.R. (2019). A corpus-based study on foreign-accented Italian. PhD Dissertation, University of Pisa.

Derwing, T.M. (2010). Utopian goals for pronunciation teaching. In Levis, J., LeVelle, K. (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, September 2009. Ames, IA: Iowa State University, 24-37.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971). Speech Perception in Infants. In *Science,* 171(3968), 303-306.

Flege, J.E. (1984). The detection of French accent by American listeners. In *The Journal of the Acoustical Society of America*, 76, 692-707.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in Cross-Language Research*. Baltimore: York Press, 233-277.

Flege, J.E. (2009). Give Input a Chance! In Piske, T., Young- Scholten, M. (Eds.), *Input Matters in SLA.* Bristol: Multilingual Matters, 175-190.

Flege, J.E., Wayland, R. (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. In *Journal of Second Language Studies*, 2(1), 1-44.

Grieser, D., Kuhl, P.K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. In *Developmental Psychology,* 25(4), 577-588.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, Massachusetts / London, England: Harvard University Press.

Gut, U. (2007). Foreign Accent. In Müller, C. (Ed.), *Speaker Classification. I Fundamentals, Features, and Methods*. Berlin: Springer, 75-87.

ISTAT. (2018). Bilancio demografico nazionale. Anno 2017. https://www.istat.it/it/files//2018/06/bilanciodemografico2018.pdf Accessed 20.03.2019.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent: Identifying intonational deviations by means of F0 generation and resynthesis.* Stuttgart: IMS.

Klein, W. (2003). *Second language acquisition*. Cambridge: Cambridge University Press.

Kolly, M.J., Dellwo, V. (2013). (How) can listeners identify the L1 in foreign accented L2 speech? In *Travaux neuchâtelois de linguistique*, 59, 127-148.

Kuhl, P.K., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). In *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979-1000.

Lenneberg, E. (1967). *The biological foundations of language*. New York: Wiley and Sons.

Levis, J.M., Moyer, A. (2014). *Social dynamics in second language accent*. Boston, MA: De Gruyter Mouton.

Likert, R. (1932). A technique for the measurement of attitudes. In *Archives of Psychology*, 140*,* 5-55.

Lippi-Green, R. (1997). *English with an Accent. Language, ideology, and discrimination in the United States.* London: Routledge.

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U. & Bäckman, L. (2013). Structural brain plasticity in adult learning and development. In *Neuroscience & Biobehavioral Reviews*, 37(9), 2296-2310.

Marotta, G., Boula de Mareuil, P. (2010). Persistenza dell'accento straniero. Uno studio percettivo dell'italiano L2. In Schmid, S., Schwarzenbach, M. & Studer-Joho, D. (Eds.), *La dimensione temporale del parlato, Atti del V Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, Zurigo (Svizzera), 4-6 febbraio 2009. Torriana: EDK Editore, 475-494.

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge, UK: Cambridge University Press.

Munro, M. (1998). The effects of noise on the intelligibility of foreign-accented speech. In *Studies in Second Language Acquisition*, 20, 139- 154.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. In *Journal of Psycholinguistic Research*, 5(3), 261-283.

Pellegrino, E., Dellwo, V. (2015). Listeners pay attention to rhythmic cues when judging the nativeness of speech. In *The Journal of the Acoustical Society of America,* 138(3), 1946-1946.

Piske, T. (2008). Phonetic Awareness, Phonetic Sensitivity and the Second Language Learning. In Cenoz, J. (Ed.), *Encyclopaedia of Language and Education*, Vol. 6. Berlin: Springer, 155-166.

Piske, T., MacKay, I.R.A. & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: a review. In *Journal of Phonetics*, 29(2), 191-215.

R CORE TEAM (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ Accessed 20.03.2019.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. In *Language Learning*, 41, 177-204.

United Nations (2017). International Migration Report 2017: Highlights (ST/ESA/SER.A/404) Department of Economic and Social Affairs, Population Division. https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2017_Highlights.pdf Accessed 20.03.2019.

Vitale, M., De Meo, A. & Pettorino, M. (2012). La competenza prosodico-intonativa nell'italiano L2: analisi e sintesi del segnale fonico di cinesi, giapponesi e vietnamiti. In Ferreri, S. (Ed.), *Linguistica educativa*. Roma: Bulzoni Editore, 329-342.

Vitale, M., Boula de Mareüil, P. & De Meo, A. (2014). An acoustic-perceptual approach to the prosody of Chinese and native speakers of Italian based on yes/no questions. In *Proceedings of the 7th International Conference on Speech Prosody*, Dublin, 48-652.

Vitale, M., Boula de Mareüil, P. & De Meo, A. (2017). Chiedere in italiano: le domande polari e lo sviluppo della competenza prosodica in parlanti cinesi di italiano L2. In Gundmundson, A., Álvarez López, L. & Bardel, C. (Eds.), *Romance Languages: Multilingualism and Language Acquisition*. Bern: Peter Lang, 143-159.

Werker, J.F., Gilbert, J.H., Humphrey, K. & Tees, R.C. (1981). Developmental aspects of cross-language speech perception. In *Child Development*, 52(1), 349-355.

WORDPRESS (2019). Create a Free Website or Blog. http://www.wordpress.com/ Accessed 17.03.2019.

Wottawa, J., Adda-Decker, M. (2016). French Learners Audio Corpus of German Speech (FLACGS). In *Proceedings of LREC 2016, Tenth International Conference on Language Resources and Evaluation*, May 2016, Portorož, Slovenia, 3215-3219.

DOI: 10.17469/O2106AISV000015

PIERO COSSU

The metathesis in the dialect of Suelli (CA): A descriptive analysis of the rhotic consonant

The paper aims at providing phonetic grounds to support the idea of Campidanese metathetic aberrant clusters (*mr, sr, tsr, tʃr*) as branching onsets. A phonological interpretation is presented, followed by the phonetic analysis of the rhotic consonant in marked clusters recurring in the dialect of Suelli (Campidanese Sardinian group). In spite of their intrinsic articulatory complexity, such clusters do not show acoustic differences as compared to cross-linguistically unmarked consonant clusters. Our analysis is focused on the rhotic consonant: its phonetic behaviour in such clusters is compared to that of /r/ in unmarked clusters and intervocalic environment. In particular, we looked for patterns analogous to those found in literature concerning marked clusters of the Campidanese typology. The analysis is based on fieldwork data and has showed that no particular acoustic pattern is at work when aberrant clusters are concerned. It will be claimed that the cross-linguistically systematic presence of a vowel portion preceding taps in *Cr* clusters and the high number of rhotic variants do not have phonological relevance in determining the syllabic status of Campidanese aberrant clusters.

Key words: rhotic consonant, Cr-cluster, metathesis, Sardinian, Campidanese.

1. *Introduction*

The metathesis of the rhotic consonant is a phonological phenomenon affecting all macro-varieties of Sardinian[1] (Logudorese, Nuorese, Arborense and Campidanese, according to the tetrapartition proposed by Virdis, 1988). In southern dialects[2] it has been systematic, while in central and northern dialects[3] its diffusion is much more limited (Lai, 2013a). In Campidanese dialects the rhotic consonant cannot occur in coda position: diachronically, every *CVrCV(CV)* lexical structure has been repaired to the form *CrVCV(CV)* as in /ˈpurpu/ > [ˈprupu] “octopus”, /ˈporku/ > [ˈproku] “pig”. In south-western Campidanese area metathesis has produced a series of typologically marked clusters ([sreβiˈðɔri] < SERVITOREM “servant”, [ˈmrekunizi] < MERCURI DIEM “Wednesday”, [ˈʧreβu] < CERVUS “deer”, [ˈʦrufuru] < SULPHUR “sulphur”). This phenomenon has been associated to rustic dialects (Wagner, 1941, Paulis, 2001, Virdis, 2013). Fieldwork data show that such clusters occur also in a

[1] We refer the reader to Wagner (1941) for a comprehensive picture of the kinds of metathesis that diachronically affected Sardinian dialects, and Lai (2013) for a phonological analysis.
[2] In most of the Campidanese group.
[3] Logudorese, Nuorese and Arborense.

more central area of Campidanese (such as the dialect of Suelli, cf. Cossu, 2017). In order to study patterns of Campidanese metathesis and the diffusion of aberrant clusters we have analysed an original dataset including recordings of elicited speech of the dialect of Suelli[4]. Speakers were asked to translate 26 sentences presented in Sardinian Italian to their own dialect. The situation in which recordings[5] took place was partially informal: speakers were informed that the purpose of our study was to document and store the *parlata* (dialect) of the village of Suelli. They were presented with sentences and after some moment during which speakers commented how Suelli's dialect is different than nearby varieties[6], they produced the translation. We let speakers free to choose lexical forms judged most appropriate.

The aim of this paper is twofold. On one hand we want to determine the syllabic status of clusters in Campidanese Sardinian, whether etymological or metathetic (both unmarked and marked) by comparing *Cr* clusters to other consonantal clusters (e.g. *sC* or geminates) through a phonological analysis. On the other hand, a preliminary analysis of the acoustic behaviour of the rhotic in metathetic aberrant clusters is provided. The analysis aims at finding potential different acoustic patterns between the rhotic within aberrant clusters and within other consonant groups. Acoustic patterns of the rhotic in marked clusters are compared to other contexts (intervocalic singleton, intervocalic geminate and especially in unmarked clusters). For what concerns Sardinian rhotics, there is an alternation between taps [ɾ] and trills [r], the formers occurring in intervocalic singleton context (V/r/V) and the latters in intervocalic geminate context (V/rː/V) (Contini, 1987), but no articulatory description is provided when postconsonantal. Such complementary distribution of [ɾ] and [r] in intervocalic position is also found in Spanish, Catalan (Recasens, Pallarès, 1999) and in Sicilian (Celata, Meluzzi & Ricci, 2016).

At this stage, ultimate and general purpose of this paper is thus to describe the behaviour of the rhotic within Campidanese marked metathetic clusters, that we firstly try to prove being tautosyllabic. In fact, in the diachrony of French the occurrence of bilabial nasal /m/ followed by rhotic /r/has determined the raising of an epenthetic plosive between these segments which is homorganic to the first consonant (CAM(E)RA > *chambre*[7]). Analogously, the Italian realisation of 'Israel' may

---

[4] The corpus used here is that created for Cossu's (2017) work.

[5] Recordings were carried out on the software Praat (version 6.0.27). The microphone employed is a Samson Meteor Mic. Sampling frequency: 44100 Hz.

[6] We would like to point that such comments were neither asked nor required by the researcher, but they have raised spontaneously. We suggest that this was due by the willing to highlight how different is a particular dialect compared to other Sardinian ones, those spoken in nearby villages too, even though, linguistically, no particular phonetic or syntactic difference is present. The only difference among dialects of the same area is the output of intervocalic Latin L: in Suelli's dialect the output is a voiced uvular fricative [ʁ] (or its deletion) as opposed to [β] of villages in the area. Virdis (1978) claims that such rhotacism, common to other areas, is not present in Trexenta (one of the historical sub-regions of Sardinia) where Suelli is located.

[7] Ségéral, Scheer (2005: 255) claim that plosive [b] emerged as the realization of /r/ in strong position, thus in onset position, since, due to syncope, a heterorganic [mr] sequence was created. The Authors also claim that new [br] cluster is an affricate, that is a complex segment, monopositional.

include an epenthetic [d] between the sibilant and the rhotic (Busà, 2013: 26). We would like to highlight that these Authors interpret such clusters as heterosyllabic. Our analysis of segmentally analogous Campidanese clusters, as it will be showed, do not report any epenthetic consonant.

2. *Sardinian rhotic metathesis*

Due to metathesis, the rhotic /r/ moves leftwards from coda position to word-initial position as the second segment of a *Cr* cluster, as in [froˈmiɣa] < FORMIC(U)LA "ant", [ˈproku] < PORCUS "pig". Other patterns occur when the phonological environment is different from *CVrCV(CV)*:

a. if the post-coda onset is a sonorant, except the bilabial nasal /m/, metathesis does not occur and a post-lexical vowel having the same quality of the preceding nucleus surfaces (/ˈpɛrla/ > [ˈpɛrɛla] "pearl", /ʤɔrˈnalɛ/ > [ʤɔrɔˈnali] "newspaper", Molinu, 1999);
b. if the preceding onset is a sonorant, with the exclusion of /m/, the rhotic consonant moves from left to right (/ˈnɛrbju/ > [ˈneβrju] "nerve", Molinu, 1999);
c. if the following onset is a voiced alveolar plosive /d/, /r/ moves from left to right (/ˈkarda/ > [ˈkaðra] "blow struck with a stick");
d. if the following onset is a voiceless alveolar plosive /t/, instead of metathesis a total regressive assimilation occurs (/ˈmɔrtu/ > [ˈmottu] "dead");
e. if there is no preceding onset, /r/ moves from left to right (/ˈɛrba/ > [ˈɛβra] "grass", /arˈʤɔla/ > [aˈʒrɔla] "farmyard".

Diachronically, Sardinian dialects has been affected by other kinds of metathesis of the rhotic consonant that will not be discussed in this paper (see Lai, 2013a for a review and analysis), since aberrant clusters do not surface as the output of such more ancient metatheses (*ibidem*). Synchronically, /r/ is able to migrate over the lexical edge in some phonological context[8]. The rhotic /r/ may in fact move from the underlying syllabic structure of a lexeme to the second position of a consonantal cluster in the preceding word (/sːa ˈɛrba/ "the grass", /nun ʧi at ˈɛrba/ "there is no grass" (Virdis, 1978, Molinu, 1999) or /santu ˈjɔrʤi/ "Saint George" surface as [sˈsrɛβa], [nun ʧaˈðrɛβa], [santru ˈjoʒi], respectively[9]).

Campidanese aberrant clusters due to rhotic metathesis are basically four: [ʧr, ʦr, mr, sr][10]:

i. [sr], as in /sirˈbɔnɛ/ > [sriˈβɔni] "wild boar";
ii. [mr], as in /ˈmɛrkunis/ > [ˈmrekunizi] "Wednesday";
iii. [ʦr], as in /tˈtsurfuru/ > [tˈtsrufuru] "sulphur";
iv. [ʧr], as in /ˈʧɛrbu/ > [ˈʧreβu] "deer".

[8] A phenomenon which Lai (2013a) labels as Synchronic South-Western Metathesis (examples in the text).
[9] Surface forms are also characterized by apocopes and lenitions.
[10] Sequences such as [ʣr, ʃr, ʒr] are also a potential output but their diffusion is lexically much more restricted and therefore they are not part of the dataset analysed in this study.

The next section will deal with the phonological status of such consonantal sequences. If they are heterosyllabic, such consonantal groups should be interpreted as coda-onset sequences; if tautosyllabic, they should instead be analysed as branching onsets.

3. *Phonological status of the aberrant clusters*

Kaye, Lowenstsamm & Vergnaud (1990) claim that neutral segments (such as sibilants and sonorants) cannot interact within a constituent, in other words there cannot be a branching onset, that is a tautosyllabic consonantal group, of the kind resulted from Campidanese metathesis (especially concerning [sr] and [mr][11]). Campidanese Sardinian morphonological rules in external sandhi provide useful clues in order to differentiate hetero- from tautosyllabic consonantal sequences. Along with *Cr* clusters, other sequences of two consonants may occur in word-initial position: *sC* clusters or the two parts of a geminate. In Sardinian, the voiceless alveolar affricate /ʦ/ and the voiceless postalveolar fricative /ʃ/ are underlyingly geminate (Bolognesi, 1998). The scenario is outlined as follows:

a. voiceless obstruents are affected by lenition (Virdis, 1978, Bolognesi, 1998) in CV# #CV context, surfacing as voiced fricatives (/sːu ˈpanɛ/ > [sːu ˈβani] "the bread", /sːa ˈtawla/ > [sːa ˈðawla] "the board», /sːu ˈkanɛ/ > [sːu ˈɣani] "the dog", /sːu ˈʧɛlu/ > [sːu ˈʒelu] "the sky", /sːu ˈfɔɣu/ > [sːu ˈvoɣu] "the fire" (Virdis, 1978);
b. within the same context, voiced obstruents and the bilabial nasal /m/ surface as in the underlying form or are optionally lengthened (Virdis, 1978[12], Ladd, Scobbie, 2003);
c. voiceless obstruents in Vs# #CV contexts are maintained (Bolognesi, 1998, Lai, 2013); voiceless palatal affricate /ʧ/ is deaffricated and preceding /s/ is totally assimilated (/is ˈʧiðas/ > [iʃ ˈʃiðaza] "the weeks" vs /sːa ˈʧɛna/ > [sːa ˈʒɛna] "the dinner");
d. voiced obstruents in Vs# #CV context are lenited and preceding /s/ is deleted (/is ˈbraʦɔs/ > [i ˈβraʦuzu] "the arms"; /is ˈdɛntɛs/ > [i ˈðɛntizi] "the teeth"; /is ʤɛˈnuɣɔs/ > [i ʒeˈnuɣuzu] "the knees"; /is ˈgatɔs/ > [i ˈɣatuzu] "the cats".
e. intervocalic *sC* clusters and underlying geminates are not affected by lenition nor other processes;
f. in Vs# #sC and Vs# #C.C contexts an epenthetic vowel is inserted between the first word's last consonant and second word's first consonant; the quality of this vowel is a copy of the preceding nucleus (/is sˈkɔlas/ > [izi sˈkɔlaza] "the schools"; /is ʃːiˈvɛdːas/ > [izi ʃːiˈvedːaza] "the basins"). Bolognesi (1998), Lai (2013b).

Of particular interest here is pattern (f). It concerns postconsonantal heterosyllabic clusters, that is a coda-coda-onset sequence, a syllabic structure illegal in many phonologies, such as in Italian (see Kaye et al., 1990 for details). The presence of a vowel is required between codas: in Italian this is solved by the choice of the determiner /lo/, which occurs in

[11] We refer the reader to the article cited and Government Phonology literature in general.
[12] Virdis (1978) claims that they may have a threeway output pattern: maintenance, deletion or they may lenit in fricatives. Ladd, Scobbie (2003) only report cases of maintenance and optional lengthening, which is also the output in this study's dataset.

complementary distribution with /il/ and /l/[13] (Marotta, 1993b; Kaye et al., 1990; Davis, 1987). An analogous pattern is shared by Campidanese phonology, as illustrated in (f).

The Vowel Intrusion Test (or Vowel-Zero Alternation: Lai, 2013b) is thus a valid diagnostic in order to verify the syllabic status of word-initial consonantal clusters in Campidanese. The presence/absence of an epenthetic vowel determines if a cluster is hetero- or tautosyllabic: Vowel Intrusion test is thus applied to aberrant clusters. External sandhi lenition patterns are checked too.

3.1 Vowel intrusion test

Recalling (f) in Section 3, a front high vowel [i] is expected between the plural determiner /is/ and the following noun beginning with *sC* or a geminate /tʦ, ʃʃ/ (Bolognesi, 1998):

i. /is sˈkalas/ > [izi sˈkalaza] "the stairs";
ii. /is ʃˈʃimprɔs/ > [izi ʃˈʃimpruzu] "the fools";
iii. /is tˈtsuɣɔs/ > [izi tˈtsuɣuzu] "the necks".

The interpretation will be that *sC* and geminates are heterosyllabic, or multiconstituent (coda+onset), sequences. Contrarily, no epenthesis occurs if the determiner /is/ precedes a single *C* or a *Cr* cluster[14]:

iv. /is ˈpurpɔs/ > [is ˈprupuzu] "the octopuses";
v. /is fɔrˈmiɣas/ > [is froˈmiɣaza] "the ants";
vi. /is ˈprɛðis/ > [is ˈpreðizi] "the priests";
vii. /is ˈkanɛs/ > [is ˈkanizi] "the dogs".

The patterns illustrate that examples (i-iii) concerns heterosyllabic sequences, and (iv-vi) are tautosyllabic clusters instead, that is branching onsets.

By the application of the same test to aberrant clusters it will be showed that we are dealing with branching onsets:

viii. /is ˈʧɛrbɔs/ > [iʃ ˈʃrɛβuzu] "the deers";
ix. /is mɛrˈʤanis/ > [i mrɛˈʒanizi] "the foxes";
x. /is sirˈbɔnɛs/ > [is sriˈβɔnizi] "the wild boars".

External sandhi patterns in viii-x are those enucleated in Section 3 as regards simple onsets. In fact, such patterns apply to syllabic constituents, rather than to simple segments[15]. Being the voiceless alveolar affricate underlyingly geminate, the [i]-epenthesis is expected:

xi. /is tˈtsurpɔs/ > [izi tˈtsrupuzu] "the blind people".

---

[13] /lo/ occurs before coda-onset sequences (#sCV or #C.CV), /il/ before (branching) onsets (#CV, #CRV), /l/ before onset-less nuclei (#VC(C)).

[14] Examples (iv) and (v) refers to metathetic unmarked clusters, (vi) is an etymological *Cr* cluster (<PRESBYTER) and (vii) is a simple onset.

[15] See Marotta (2008) for an analogous pattern in Tuscan varieties and Brun-Trigaud, Scheer (2010) for lenition in branching onsets in the diachronic evolution from Latin to Modern French.

Moreover, we would like to point that not all /s/ plus liquid clusters are to be interpreted as tautosyllabic. In fact, as reported by Lai (2015: 289), when in clusters of this kind the sibilant is post-lexical, in our case a negative verbalizer, the consonant group created this way is heterosyllabic. This is again inferable by the presence of an epenthetic [i] after the determiner /is/, as in the example:

xii. /is slumˈbaus/ > [izi zlumˈbauzu] "the cripples".

3.2 Lenition patterns of aberrant clusters

To check lenition patterns of aberrant clusters in external sandhi may sound redundant, but we think it is crucial to corroborate the idea that *sC* sequences and [sr] clusters are different phonological items. The preconsonantal sibilant /s/ in coda position assimilates the voicing feature of the following segment:

xiii. /sbalˈlaj/ > [zbalˈlaj] "to mess up" (Virdis, 1978).

On the contrary, in the metathetic aberrant clusters made up of an alveolar fricative and a rhotic, the former surfaces as voiceless (being its underlying form /s/) in post-pausal position or absolute initial position:

xiv. ##/ˈsɔrga/ > ##[ˈsrɔɣa] "mother-in-law";
xv. ##/sirˈbɔnɛ/ > ##[sriˈβɔni] "wild boar".

Such examples highlight that sibilants in *sC* sequences occupy the coda, while in [sr] clusters they are the first segment of a branching onset: just like in other branching or simple onsets, the voiceless segment surfaces as such, but as voiced [z] in intervocalic position.

A presumably satisfying amount of evidence has been provided concerning the tautosyllabicity of the marked consonantal groups in Sardinian by the application of the vowel intrusion test and the comparison of lenition patterns to those of hetero- and tautosyllabic consonantal sequences[16]. Next section will concern the phonetic characteristics of rhotic consonants of the Sardinian inventory.

4. *Description of rhotics in Sardinian*

Sardinian inventory is characterised by two rhotic consonants[17]. Singleton /r/ is prototypically realised as an alveolar tap [ɾ], while geminate /rː/ is phonetically realised as an alveolar trill [r] (Contini, 1987). The same scenario occurs in Spanish, Catalan (Recasens, Pallarès 1999) and Sicilian (Celata et al., 2016).

[16] See Loporcaro (2005) for a heterosyllabic interpretation of *muta cum liquida* clusters in Proto-Romance and Italian southern dialects and Lai (2013a) for Sardinian metathetic clusters as complex segments.
[17] Intervocalically, in complementary distribution.

4.1 *Taps*

Alveolar taps are produced by a brief (15-25 ms in Busà, 2013; 15-30 ms in Recasens, 2007) contact between the tongue tip and the alveolar ridge. The movement is upwards and muscularly controlled (Ladefoged, Maddieson 1996, Recasens, 2007: 1). In Italian they are in allophonic variation with non-sibilant alveolar fricatives [ɹ̝] and alveolar approximants [ɹ]: the three of them realised with an apical tongue gesture (Spreafico, Celata, Vietti, Bertini & Ricci, 2015). Taps may also be produced with a burst or fricated release, the latter having a mean duration of 20-30 ms (Busà, 2013).

4.2 Trills

Contrary to taps, the apical movement in trills is aerodynamically driven (Ladefoged, Maddieson, 1996): tongue tip vibration is "initiated when there is enough pressure difference across the apicoalveolar constriction" (Recasens, 2007: 1, see also Solé, 2002b). Closeness between articulators and intensity of the airflow must be maintained constant in order for a trill to occur: if such requirements are not met, a non-trilled output will occur (Laver, 1994). Solé (2002b) demonstrated that long fricative rhotics, approximants and voiceless trills may surface when proper tongue configurations (lax tip of the tongue and stiffness of the dorsum) or aerodynamic requirements are not satisfactorily set, for example in fast and spontaneous speech.

4.3 Postconsonantal rhotics

Both trills and taps may occur in this position. The formers are fully voiced and produced generally with two apical contacts (Recasens, 2007, for Catalan dialects). More interestingly, postconsonantal taps are flanked by a svarabhakty[18]. A great number of studies reported the cross-linguistic occurrence of a vowel portion between a consonant and a tap (Navarro, 1963; Rochette, 1973; Harms, 1976; Massone, 1988; Blecua, 2001; Ramírez, 2006; Frigeni, 2009; Baltazani, Nicolaidis, 2011; Busà, 2013; Savu, 2012; Celata et al., 2016; Spreafico, Kaland, Galatà & Vietti, 2016 and others). The phonetic description is consistent along the studies reported for what concerns duration and quality[19], but different scholars propose different interpretations as regards the reasons of its emergence. In sum, there are two main lines of studies: (a) the intrusive vowel is analysed as a post-lexical segment, that is as an epenthesis; (b) it is intended as part of the articulation of the tap, so phonologically irrelevant and phonetically (or physiologically) determined. For example, Busà (2013) claims it occurs in order to avoid /sr/ sequences in Italian (as in /israˈɛlɛ/); Savu (2012), concerning Slavic languages, claims that it is part of the articulation of the tap and it is present also when [ɾ] is flanked by vowels. This study endorses the line that sees such vowel as the product of a phonetic mechanism.

[18] By svarabhatky it is meant the post-lexical vocalic insertion between two consonants.

[19] The intrusive vowel seems to be influenced by the following nucleus.

5. *Description of the dataset and analysis*

Seventeen speakers divided in three age groups (20-40, 41-60 and 61-85 years old) were recorded. Speakers come from the village of Suelli (40 km due north to Cagliari, central-western Campidanese dialectal domain[20] (Virdis, 1988). Description of this area is provided in Virdis (1978, 1988). Stimuli were elicited through a Sardinian-Italian to Sardinian translation task. Twenty-six sentences[21] in Sardinian-Italian were presented to the speakers. Twelve sentences were expected to include an aberrant metathetic cluster (*ʧr* = 4 times; *tsr*, *mr*, *sr* = 3 times each), fourteen sentences included intervocalic rhotics and/or unmarked *Cr* clusters. In order to reduce the cognitive load[22], some sentence included more than one target word, except those with an expected aberrant cluster (one per sentence). Sentences were presented in random order. Initially, unmarked clusters were not meant to be considered, but since six words included in word-initial position such sequences, /pr/, /kr/, /tr/ and /fr/ have also been analysed.

Aberrant clusters were expected in, e.g., <la carne di cinghiale è buona ma costa cara> (corresponding Campidanese translation [sːa ˈβetːsa dɛ zriˈβɔni ɛ ˈβɔna ma ˈɣɔsta ˈmːɛða] "wild boar meat tastes good, but it is expensive"), <verso le undici facciamo merenda a pane e formaggio> ([faˈʧɔri dɛ iz ˈundizi mruˈdːzauzu a ˈpːani e ˈkːazu] "we are going to have bread and cheese for brunch around 11 a.m.").

In this study, the parameters taken into account are: manner of articulation of the rhotic, duration of the rhotic (including the svarabhakti portion, when present), duration of the svarabhakti and closure portion separately. The annotation[23] system follows Celata, Vietti & Spreafico (2019). We have chosen such system since it is able to elegantly illustrate sub-portions of rhotics, particularly useful here in order to describe the vowel portion as part of the tap, that is a mere phonetic product, phonologically irrelevant.

5.1 Intervocalic singleton /r/

As expected, in line with previous studies (Celata et al., 2016), three rhotic variants are found as allophones of singleton /r/: taps [ɾ], fricatives [ɹ̝] and approximants [ɹ]. Rhotics have been labelled as 'approximants' when they show a continuous formant structure and a reduced waveform. Busà (2013: 30), following Willing, Bradley

[20] Actually, given the rhotacised ([ʁ]) output of Latin intervocalic L, Suelli's dialect should be associated to Central Campidanese as described in Virdis (1988). The Author inserts Trexenta, the area where the dialect *sub iudice* is spoken, within central-western Campidanese domain which is characterised by deletion of intervocalic /n/ and nasalisation of preceding vowels. Suelli's dialect is devoid of such phenomena.

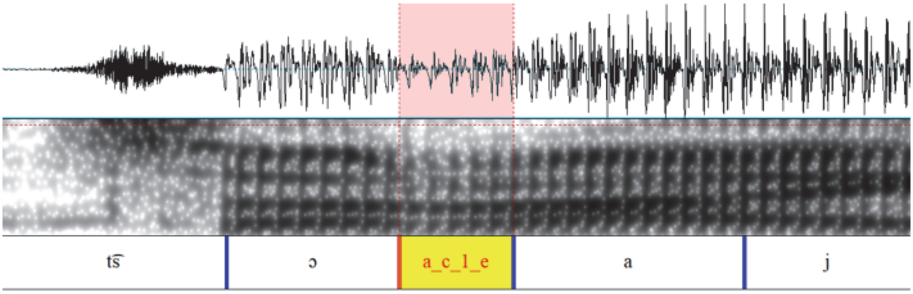
[21] Multiplied by number of speakers, a 440 sentences corpus was created.

[22] That is, to reduce the number of sentences to translate.

[23] The system conveys four pieces of information: (i) 'r' (trill), 't' (tap), 'f' (fricative), 'a' (approximant); (ii) digits (that refer to the first, second, and so on, cycle of apertures/closures which are indicated by (iii) 'a' and 'c', respectively); (iv) 'e' (single-phase segment), 'f' (flanked by other infrasegmental rhotic portions), 'r' and 'l' (the last or first, respectively, portion of a multiple-phase rhotic).

(2008: 93), calls such variants ‘perceptual taps’, “produced like approximants”. Figure 1 shows a token of intervocalic [ɹ].

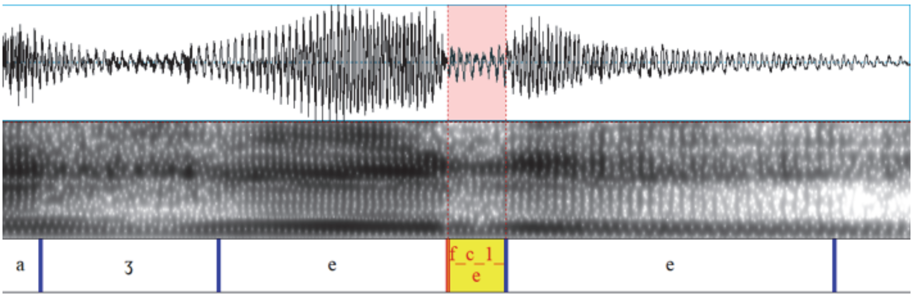
Figure 1 - *Spectrogram of an approximant [ɹ] in the word* Lotzorai, *toponym*



Perceptually, they are indistinguishable from taps and speakers are not aware of the acoustic difference. It is strongly suggested that approximants and fricatives arise when there is no contact between articulators[24] during the realisation of an intended alveolar tap (see Celata et al., 2019: 114).

Rhotics fricatives are labelled as such if the formant structure is strongly reduced and the portion shows aperiodic noise (around 3000 Hz). They are exemplified in Figure 2.

Figure 2 - *Spectrogram of a rhotic fricative in* sa cereʒa *“the cherry”*

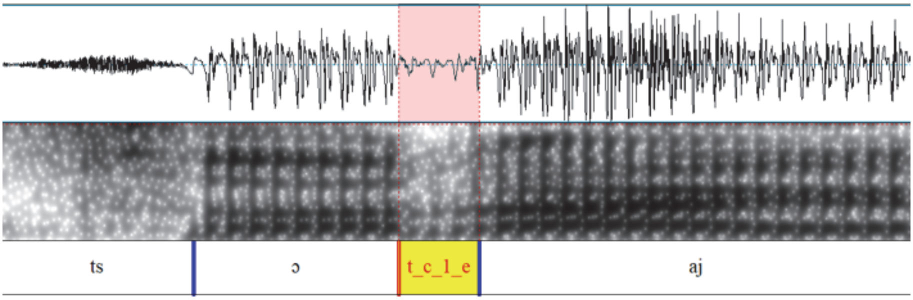


In this case too, no perceptive difference is noticed, neither by the researcher, nor by the speaker.

In the dataset analysed, most of the rhotics labelled as a tap [ɾ] show a certain degree of frication or formant structure, even though very weak. Nevertheless, classification as [ɾ] has been chosen since the waveform is much more reduced than in approximants and fricatives, meaning that the tip of the tongue is extremely close to the alveolar ridge, even though this might sound arbitrary. Low accuracy of speech as that of this corpus might lead to the production of ‘incomplete’ taps as a general pattern. See Figure 3.

[24] In fricatives, the tongue tip sets closer to the alveolar ridge than in approximants.

Figure 3 - *Spectrogram of a tap [ɾ] in* Lotzorai*, toponym*

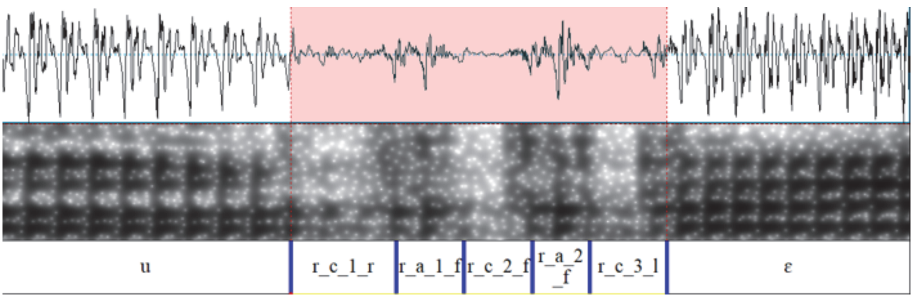


No trilled rhotic occurs in singleton intervocalic environment, as expected (Contini, 1987). Considering all speakers and all occurrences of /r/, approximants are the most frequent tokens (39%), followed by taps (32%) and fricatives (29%). The mean duration of /r/ is 28 ms (st. dev. 10.7). There is anyway a great cross-individual variability as regards the rate of variants produced: some speakers never produced a tap, others produced only fricatives, one produced only taps, a few of them produced all variants. It seems then that speakers have individual patterns concerning the articulation of the rhotic and that this may not be influenced by the phonetic context, rather by speech accuracy. Further work should consider other variables in order to deeper understand the grounds of such scenario in Sardinian dialects.

5.2 Intervocalic geminate /r:/

In this context, trills[25] are possible variants along with fricatives, approximants and taps; compound articulations are found too. By compound articulation it is meant a sequence of two rhotic variants ([ɾɹ], [ɾɹ̝], [ɹ̝ɹ]). Articulations of this kind are reported also for Sicilian (Celata, Meluzzi & Ricci, 2016). Trills with an approximant ending have been labelled as trills [r] since this is the unmarked cross-linguistic pattern (Ladefoged, Maddieson, 1996). Only 5 tokens are single complete constriction rhotics, which have been labelled as taps (mean duration 27 ms). Voiceless trills are also found, as exemplified in Figure 4.

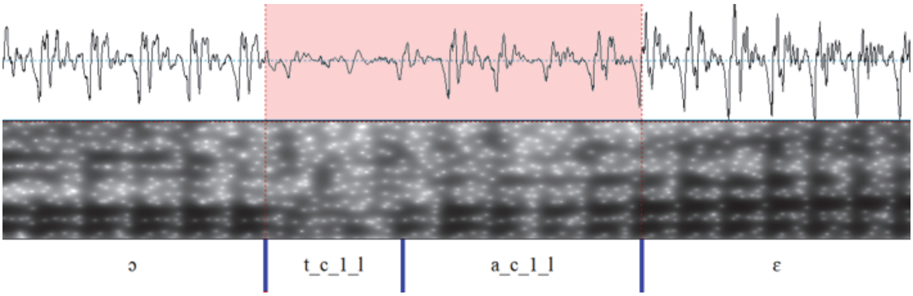
Figure 4 - *Spectrogram of a voiceless trill in* su rei*, "the king"*



[25] The unmarked rhotic allophone (Contini, 1987), for Italian see Bertinetto, Loporcaro (2005: 4).

Figure 5 represents a token of a compound rhotic articulation: a tap followed by an approximant [ɾɹ]. Articulations such as this one have been analysed as a sequence of rhotic consonants. They could be interpreted as trills in which the airflow energy quickly decreases or the tip of the tongue is not maintained close enough to the alveolar ridge for the whole duration of the segment, causing the failing in vibration.

Figure 5 - *Spectrogram of a compound rhotic articulation [ɾɹ] in* torrendi *"coming back"*



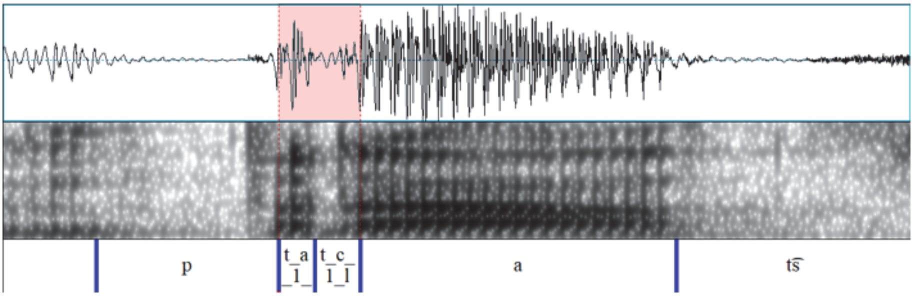
A reduced waveform is indeed visible in the middle of the approximant portion, indicating that the tongue preserves a certain degree of movement. Mean duration of intervocalic geminate rhotics is 77 ms (st. dev. 30.1). Trills and compound articulations are 27% and 26%, respectively, fricatives occur 34% and taps 5%. In this context too, the distribution of variants is cross-individually variable: two speakers produced only fricatives, two speakers realised only trills, others show a tendency towards compound articulations, the most of them produced more than two rhotic variants.

5.3 Postconsonantal rhotics

As regards unmarked *Cr* clusters, the pattern observed aligns with cross-linguistic data. Taps, approximants and rhotic fricatives are found (even though no trill) and two tokens of compound articulations. Concerning the latter variant, it is proposed the same description advanced in Section 5.2. Taps are always preceded by a vowel-like portion, which numerous other studies have defined as svarabhakti or the secondary product of the articulation of a tap (see Section 4.3 for a review). In the dataset, 52% are taps, 40% are approximants, 5% are fricatives, 3% are compound articulations (tap+approximants).

Nevertheless, special attention should be paid to the vowel portion preceding the rhotic closure. See Figure 6, which illustrates a tap.

Figure 6 - *Spectrogram of a tap in* pratza *"courtyard"*

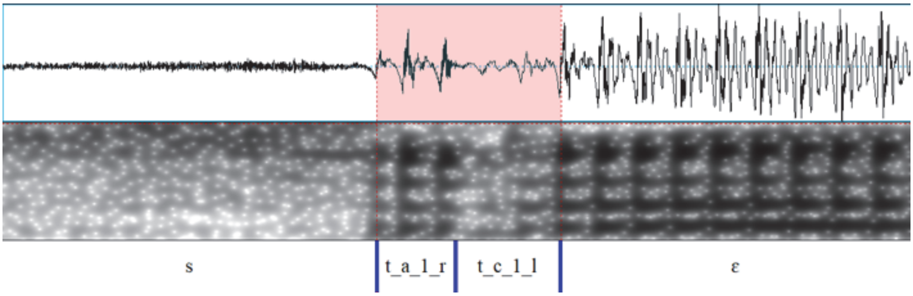


Postconsonant rhotics in unmarked clusters have a mean duration of 44 ms (st. dev. 15.4). Scarcity of tokens for this context do not allow for a deeper analysis of the behaviour of postconsonantal rhotics in Sardinian.

5.3.1 Postconsonantal rhotics in aberrant clusters

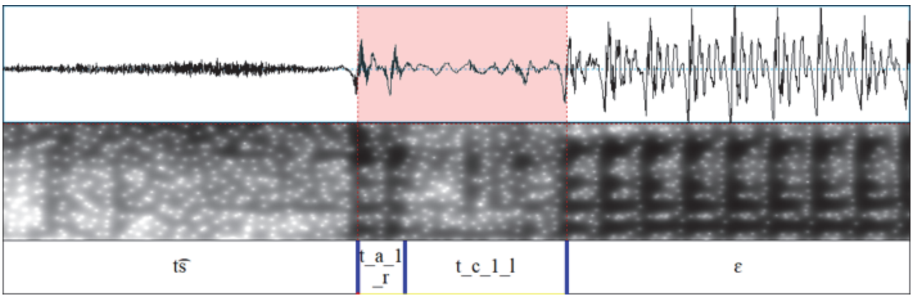
Data are presented analytically. The first cluster considered is *sr*. Sequences of alveolar fricatives and alveolar rhotics are characterised by antagonistic lingual configurations: fricatives require a tension of the tongue's apical-laminal region and advanced dorsum, rhotics need a relaxed tip and retracted postdorsum (Solé, 2002a). In Spanish such complexity is repaired by the regressive assimilation of the fricative to the rhotic (Solé, 2002a); in Italian, an epenthetic voiced alveolar occlusion [d] between the two segments, along with an epenthetic vowel, may occur ([iz^{d}raˈɛle], [autobus^{ə}ˈros:o], Busà, 2013: 26). Based on the dataset analysed in this study, no particular pattern is observed in Sardinian (neither consonantal nor vocalic epenthesis is observed). Taps, fricatives, approximants and, rarely, trills occur as outcomes of the rhotic (46% fricatives, 27% taps, 14% approximants, 10% compound rhotics, 3% trills, total tokens: 42). Only rhotics with clear closure have been labelled as taps: those showing a slightly reduced waveform but having a formant structure indistinguishable from preceding vowel portion have been annotated as approximants (which has sometimes been called "perceptual taps": Busà, 2013). Taps are always preceded by a vowel-like portion, which formant structure is very similar to that of approximants. Its mean duration is 34 ms, slightly longer than the closure phase (23 ms). An example is provided in Figure 7, where closure is longer than the svarabhakti.

Figure 7 - *Spectrogram of a tap preceded by a vowel portion in* cussreba *"that grass"*



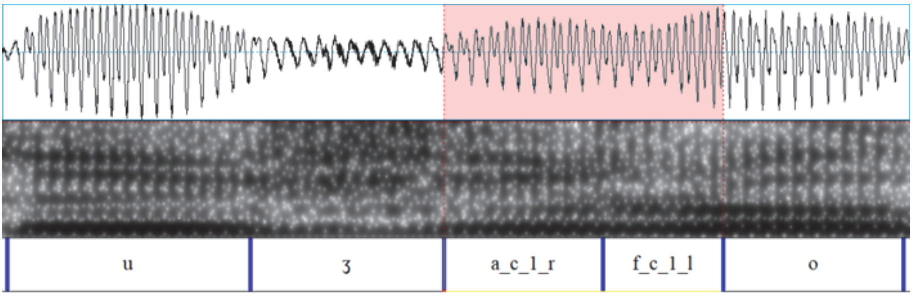
In ʦ*r* clusters the most occurring variant of the rhotic is the approximant (50%, total tokens 44). It is followed by taps (25%), fricatives (11%), trills and compound articulations (3% each). Taps are systematically preceded by a vowel-like portion. Its duration is variable across speakers, averaging from 10 ms to 50 ms. Figure 8 shows the realisation of a tap preceded by an extremely short vocalic portion (10 ms); closure phase is followed by a fricative release (33 ms in total).

Figure 8 - *Spectrogram of a tap in* tzrepi *"sprog"*



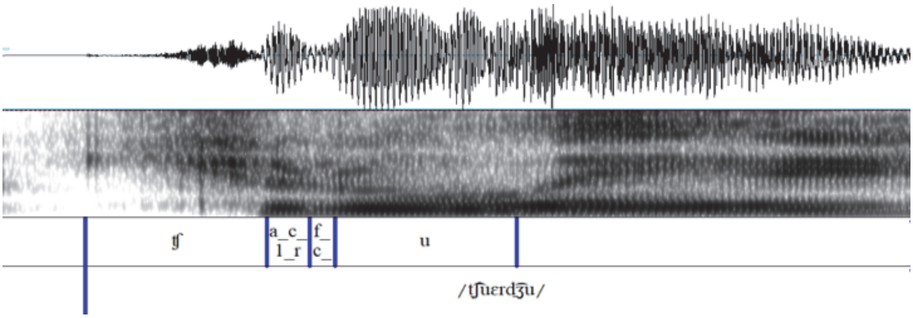
For what concerns ʧ*r* clusters, to our knowledge, no study has ever dealt with the aerodynamics and articulatory characteristics of such consonantal group, maybe due to their high cross-linguistic markedness. The dataset at hand only concerns acoustic data: this implicates difficulties in determining the factors involved during the articulation of an alveolar rhotic preceded by a postalveolar affricate [ʧ]. Nevertheless, the aim of this paper is to investigate the presumably different spectrographically visible patterns of articulation of the rhotic in Campidanese Sardinian aberrant clusters. In this cluster, too, high variability is found: taps (22%), approximants (37%), fricatives (29%), trills (2%) and compound rhotics (10%) are attested. Recall from Section 3 that voiceless postalveolar affricate lenits to [ʒ] in intervocalic environment and to [ʃ] when preceded by /s/. So, tokens of actual [ʧr] clusters are very rare.

Figure 9 - *Spectrogram of intervocalic [ʒɹ̥] in* /sːu ʧɔrˈbɛdːu/ *cluster resulted by the lenition of* /ʧ/ *in such position. The rhotic is analysed as made up of an approximant and a fricative portion*



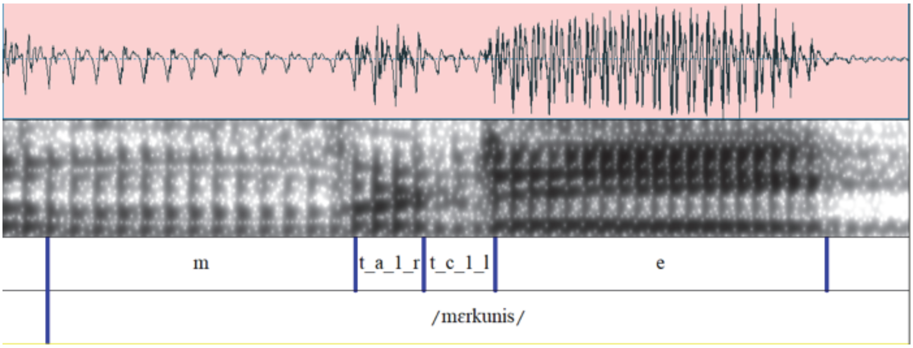
Actual [ʧr] realisations may only occur in absolute position or post-nasally. A token of this kind is showed in Figure 10. The female speaker produced the word [ʧruˈeʒu] "oak" after a pause, so allowing the voiceless post-alveolar affricate to surface as such.

Figure 10 - *Spectrogram of a post-pausal [ʧr] cluster in [ʧruˈeʒu] "oak"*



We finally consider *mr* clusters. Frigeni (2009) claims that after the bilabial nasal /m/, the rhotic consonant goes through nasalisation surfacing as [n]. We have not investigated such phenomenon in this study, since it seems to be only a phonetic sub-product and it does not have any phonological implication. Nevertheless, Frigeni (2009), too, reports that when the rhotic variant is a tap a small vowel between the preceding consonant and [ɾ] occurs. She reports (*ibidem*: 113) that "some tokens [...] do not show either a disruptive vowel fragment or a rhotic portion". What the spectrogram (Frigeni, 2009: 113) shows is an approximant of the same kind of those found in this study's fieldwork data. The same set of variants described in previous sections is indeed observed also in nasal-rhotic clusters: taps (28%), approximants (56%), fricatives (7%), compound articulations (9%).

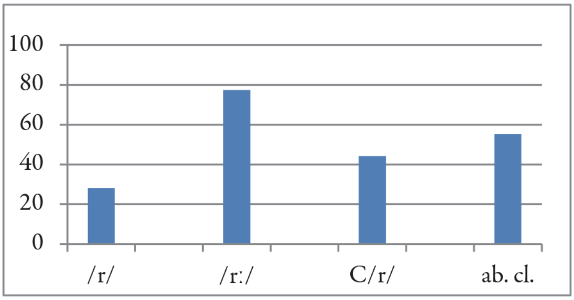
Figure 11 - *Spectrogram of [mr] in /ˈmrɛkunis/. Rhotic is produced as a tap*



5.4 Measurements and general considerations

In Table 1 we present the mean durations of rhotics in aberrant clusters, unmarked clusters and intervocalic context: aberrant clusters (general average is about 55,3 ms): fricative /s/+rhotics (58,8 ms), alveolar affricate /ʦ/+rhotic / (53,3 ms), post-alveolar affricate /ʧ/+rhotic (57,5 ms), bilabial nasal /m/+ rhotic (51,5 ms), unmarked clusters (44,2 ms), intervocalic singleton (28,2 ms), intervocalic geminate (77,3 ms).

Table 1 - *Mean duration (ms) of the rhotic in all contexts considered*



The mean duration of vowel portion and closure phase in taps is as follows, respectively: [sɾ] (34, 6 ms; 23,7 ms), [ʦɾ] (30,1 ms; 21, 5 ms), [ʧɾ] (33,8 ms; 25,7 ms), [mɾ] (27,5 ms; 21, 2 ms), unmarked clusters (25,3 ms; 19 ms).

For what concerns the quality of the vowel portion, no general consideration is possible since the dataset is not balanced: first, on a total of 261 postconsonantal rhotics, only 89 are taps. Vocalic context in which a tap may appear is highly variable, considered a seven vowel set ([i, u, ɛ, ɔ, e, o, a]) and 8 preceding underlying consonants plus their potential lenited form ([p, β, t, k, ɣ, s, z, f, v, m, ʧ, ʒ, ʃ, ʦ]). Such high variability in phonetic environments extremely reduces the possibility to assess an influence by the environment. No statistical analysis is thus possible. However, token by token measurement has been carried out. The data collected reveal a tendency for the vowel portion to be modulated accordingly to the following nucleus

(Busà, 2013). Due to space limitations, single values of first two formants of vowel portion and nucleus are not reported here.

6. *Discussion and conclusions*

This paper focused on the phonetic analysis of the rhotic consonant in Campidanese aberrant clusters compared to its phonetic behaviour in cross-linguistic unmarked clusters and in intervocalic environment (both singleton and geminate). Our aim was to find evidence corroborating the phonological analysis proposed, based on the rhotic behaviour described in other studies and concerning different languages which referred to analogous clusters (such as *mr* and *sr).* Unfortunately, the coda context is absent from the dataset. In fact, as we have claimed and illustrated in Section 2, rhotics in coda position have all been moved to the onset position. As a result, no heavy syllable including a rhotic in the rhyme is present in Suelli's dialect[26].

Based on Suelli's dialect and literature data, we proposed a phonological analysis of typologically marked clusters as branching onsets. We hypothesised that if clusters have different phonological status, this could be mirrored in a different phonetic realisation. That is, if marked metathetic clusters are heterosyllbic (vs. tautosyllabicity of unmaked clusters, both metathetic or etymological), it is plausible to expect some different phonetic strategy, especially concerning what we called aberrant clusters. As it is claimed above, such hypothesis is justified by other studies' data on /sr/ and /mr/ clusters. As reported in Ségéral, Scheer (2005), in French, an epenthetic bilabial plosive [b] surfaced after that bilabial nasal /m/ and rhotic formed a consonantal cluster due to syncope of /e/ in e.g. the diachrony of CAM(E) RA > *chambre*. Authors analyse new [br] clusters as the realisation of /r/ in strong position, that is in post-coda position. According to the Authors, /mr/ clusters in French should have been interpreted as heterosyllabic. Analogously, Busà (2013) reports of an epenthetic alveolar plosive in the realisation of heterosyllabic /sr/ clusters in Italian, for example in *Israele* "Israel". Such patterns, attributed to the realisation of marked heterosyllabic clusters, are not observed in our corpus. We thus believe that absence of plosive epenthesis in [mr] and [sr] clusters (as well as in other sequences) in the dialect of Suelli may corroborate our conclusions concerning tautosyllabicity of metathetic marked clusters we presented in Section 3.

The phonetic analysis carried out shows that the same patterns of articulation of rhotics occur regardless of the markedness of the cluster. Four manners of articulation ([r, ɾ, ɹ, ɹ̝]) plus compound rhotics have been found cross-contextually (post-consonantally and intervocalically). This does not tell us much about phonological status of Campidanese aberrant clusters, but it adds data on literature concerning Campidanese production of rhotics in consonantal clusters. Moreover, data show that trills and compound articulations may only occur as allophones of /rː/ or of cluster-internal /r/. We suggest that this fact might indicate that the gesture asso-

[26] With the exclusion of some very recent Italian loanword, not included in our corpus.

ciated to a postconsonantal rhotic in Campidanese dialect is an alveolar trill [r]. Recall from Section 5 that we have not found trills or compound rhotics in V/r/V context, but [ɾ] may occur as an output of an intervocalic geminate rhotic.

In the postconsonantal environment, taps are always flanked by a short vowel, regardless of the clusters' markedness. This intrusive vowel is interpreted here as a phonetic sub-product arising during the movement of the tip of the tongue from its position in the preceding articulatory configuration to that required to produce a tap (or a trill, too). Data show that in homorganic clusters its duration is slightly longer than in heterorganic ones, increasing even more if *sr* or ʦ*r* clusters are concerned (supposedly, due to the antagonistic configurations of such segments). We suppose that, in clusters of this kind, more time is needed in order to set a sequential and different tongue configuration as compared to heterorganic clusters, resulting in a longer vowel portion. Speakers are not aware of its presence, neither in unmarked nor marked clusters. Further research is certainly needed to deeper investigate the patterns of occurrence of such vowel. A UTI or EPG study is strongly suggested. Nevertheless, the high variability concerning the set of rhotics inter-individually observed in all contexts[27] and the occurrence of a vowel portion in *Cr* clusters are interpreted as phonologically not pertinent. The phonological analysis along with the phonetic description proposed in this paper lead to homogeneously interpret unmarked and aberrant clusters in Campidanese as tautosyllabic. Therefore, Southern Sardinian (Campidanese dialectal group) may be considered a dialect having much looser phonotactic constraints in comparison with others Romance dialects.

Bibliography

Baltazani, M., Nicolaidis, K. (2013). The many faces of /r/. In Vietti, A., Spreafico, L. (Eds.), *Proceedings of the r-atics 3 Conference*, Bolzano, Italy, 125-144.

Bertinetto, P.M., Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. In *Journal of the International Phonetic Association*, 35/2, 131-151.

Blecua, B. (2001). Las vibrantes del español: manifestaciones acústicas y procesos fonéticos. Ph. D. Disseration, Universitat Autònoma de Barcelona.

Bolognesi, R. (1998). The phonology of Campidanian Sardianian. A unitary account of a self-organizing structure. Ph.D. Dissertation, University of Amsterdam.

Brun-Trigaud, G., Scheer, T. (2010). Lenition in branching onsets in French and in ALF dialects. In Karlík, P. (Eds.), *Development of Language through the Lens of Formal Linguistics*. Munich: Lincom, 15-28.

[27] In my corpus, variability concerning articulation of rhotics seems to depend on the speaker, rather than on phonetic context. Regardless of the vocalic context, some speaker tends to produce taps, others prefer approximants, but the most of them produce all variants in random-like fashion. We suggest that extralinguistic factors may have influenced the production of rhotics within our corpus.

Busà, M.G. (2013). Aspects of consonant cluster mutations: the case of /sr/ in Italian. In Sánchez Miret, F., Recasens, D. (Eds.), *Studies in phonetics, phonology and sound change in Romance*, *LSRL 73*. Monaco: Lincom, 23-42.

Celata, C., Meluzzi, C. & Ricci, I. (2016). The sociophonetics of rhotic variation in Sicilian dialects and Sicilian Italian: corpus, methodology and first results. In *Loquens*, 3(1), 1-13.

Celata, C., Vietti, A. & Spreafico, L. (2019). An articulatory account of rhotic variation in Tuscan Italian. Synchronized UTI and EPG data. In Gibson, M., Gil, J. (Eds.), *Romance Phonetics and Phonology*. Oxford: Oxford University Press, 91-117.

Contini, M. (1987). *Étude de géographie phonétique et de phonétique instrumentale du sarde*. Torino: Edizioni dell'Orso.

Cossu, P. (2017). Metatesi a Suelli (CA). Tra fonetica e fonologia. MA Thesis, Università di Pisa.

Davis, J. (1987). A combinatory phonology of Italian. In *Columbia University Working Papers in Linguistics*, 8, 1-99.

Frigeni, C. (2009). Sonorant relationships in two varieties of Sardinian. Ph. D. Dissertation, University of Toronto.

Kaye, J., Lowenstamm, J. & Vergnaud, J.R. (1990). Constituent structure and Government Phonology. In *Phonology*, 7(2), 193-232.

Harms, R.T. (1976). The segmentalization of Finnish 'nonrules'. In *Texas Linguistic Forum*, 5, 73-88.

Ladefoged, P., Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lai, R. (2013a). *Positional effects in Sardinian Muta cum Liquida. Lenition, metathesis and liquid deletion.* Alessandria: Edizioni dell'Orso.

Lai, R. (2013b). Fattori posizionali nella metatesi del sardo sud-occidentale. In *Annali Online di Ferrara – Lettere*, 8(2), 1-24.

Lai, R. (2015a). Lateral relations in Sardinian metathesis: A unified account. In Russo, D. (Eds.), *The Notion of Syllable across History, Theories and Analysis*. Cambridge: Cambridge Scholar Publishing.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Loporcaro, M. (2005). La sillabazione di *muta cum liquida* dal latino al romanzo. In Kiss, A., Mondin, L. & Salvi, G. (Eds.), *Latin et langues romances*. Tübingen: Max Niemeyer Verlag, 419-430.

Marotta, G. (1993b). Selezione dell'articolo e sillaba in italiana: un'interazione totale?. In Accademia della Crusca (Ed.), *Studi di grammatica italiana*, XV, 255-296.

Marotta, G. (2008). Lenition in Tuscan Italian (Gorgia Toscana). In Brandão de Carvalho, J., Scheer, T., & Ségéral, P. (Eds.), *Lenition and Fortition*. Berlin: Mouton de Gruyter, 235-272.

Massone, M.I. (1988). Estudio acústico y perceptivo de las consonantes nasales y liquidas del español. In *Estudios de fonética experimental*. Barcellona: PPU, 14-34.

Molinu, L. (1999). Métathése et variation en sarde. In *Cahiers de Grammaire*, 24, 153-181.

Navarro, T. (1963). *Manual de pronunciation española*. New York: Hafner Publications.

Paulis, G. (2001). Il sardo unificato e la teoria della pianificazione linguistica. In Argiolas, M., Serra, R. (Eds.), *Limba, lingua, language. Lingue locali, standardizzazione e identità in Sardegna nell'era della globalizzazione.* Cagliari: CUEC, 155-171.

Ramírez, C.J. (2006). Acoustic and perceptual characterization of the epenthetic vowel between the clusters formed by consonant + liquid in Spanish. In Diaz-Campo, M. (Ed.), *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. Sommerville: Cascadilla Proceedings Project.

Recasens, D., Pallarèz, M.D. (1999). A study of /ɾ/ and /r/ in the light of the "DAC" coarticulation model. In *Journal of Phonetics*, 27, 143-169.

Rochette, C.-E. (1973). *Les groupes de consonnes en français: étude de l'enchaînement articulatoire à l'aide de la radiocinématographie et de l'oscillographie.* Quebec: Les presses de l'Université Laval.

Savu, C.F. (2012). On the Phonetic Structure of the Rhotic Tap and Its Phonological Implications. MA Thesis, University of Bucharest.

Ségéral, P., Scheer, T. (2005). *What lenition and fortition tells us about Gallo-Romance Muta cum Liquida*. In Geerts, T., Haike, J., van Ginneken, I. (Eds.), *Romance Languages and Linguistic Theory 2003*. Amsterdam: Benjamins, 235-267.

Solé, M.J. (2002a). Assimilatory processes and aerodynamic factors. In Gussenhoven, C., Warner, N. (Eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 351-386.

Solé, M.J. (2002b). Aerodynamic charateristics of trills and phonological patterning. In *Journal of Phonetics*, 30, 655-688.

Spreafico, L., Celata, C., Vietti, A., Bertini, C. & Ricci, I. (2015). An EPG+UTI study of Italian /r/. In *Proceedings of the 18th International Congress of Phonetic Sciences*, 1-5.

Spreafico, L., Kaland, C., Galatà, V. & Vietti, A. (2016). An acoustic analysis of /r/ in Tyrolean. In *Proceedings of Interspeech 2016*, 1002-1006.

Virdis, M. (1978). *Fonetica del dialetto sardo campidanese*. Cagliari: Condaghes.

Virdis, M. (1988). Sardisch: Areallinguistik (aree linguistiche). In Schmidt, C., Holtus, G. & Metzeltin, M. (Eds.), *Lexikon der Romanistischen Linguistik IV*. Tübingen: Mouton de Gruyter, 897-915.

Virdis, M. (2013). La varietà di Cagliari e le varietà meridionali del sardo. In Paulis, G., Pinto, I., Putzu, I. (Eds.), *Repertorio plurilingue e variazione linguistica a Cagliari*. Milano: Franco Angeli, 165-180.

Wagner, M.L. (1941). *Historische Lautlehre des Sardischen*. Halle (Saale): Max Niemeyer.

Willis, E.W., Bradley, T. G. (2008). Contrast Maintenance of Taps and Trills in Dominican Spanish: Data and Analysis. In Colantoni, L., Steele, J. (Eds.), *Selected Proceedings of the 3rd* Conference *on Laboratory Approaches to Spanish Phonology*. Sommerville: Cascadilla Proceedings Project, 87-100.

DOI: 10.17469/O2106AISV000016

LORENZO FILIPPONIO, DAVIDE GARASSINO, DALILA DIPINO

Between phonology and typology. Consonant duration in two Gallo-Italian dialects[1]

Segmental correspondences are sometimes due to completely different factors. We exemplify this claim by considering the allophonic gemination of post-stress consonants in Bologna and Porto Maurizio dialects, two varieties which display contrastive vowel length. After the experimental confirmation of vowel length contrasts and the existence of differences in the duration of the post-stress consonants, we qualitatively analyze intensity contours. Despite similar duration values, in light of different intensity patterns and rhythm accounts, we hypothesize that post-stress gemination derives from close contact in Bolognese, due to a stronger compensative pattern, while it is residual or due to the pressure of standard Italian in Portorino.

Key words: Consonant duration, Northern Italo-Romance, Dialectology, Phonology, Typology.

Introduction

The aim of this paper is to propose a first experiment dealing with the coexistence of vowel length oppositions and a longer duration of post-stress consonants after short stressed vowels in two Northern Italo-Romance dialects. We will do this in § 4 by measuring the quantity of stressed vowels and post-stress consonants (the methods are explained in § 3). A further, particularly relevant aim is trying to find out whether the correspondence of these features must be ascribed to a similar rhythm pattern or not, which will lead us to some typological considerations about rhythm. In order to do that, we will provide a short introduction to intensity contours (§ 5) before concluding with a brief discussion (§ 6). Before starting our analysis, we will explain what is meant here by *rhythm* (§ 1), since we will consider single (phonological) words and not longer speech chains, and we will give a brief description of length patterns in Gallo-Italian Dialects (§ 2).

1. *Some preliminary remarks on rhythm*

We call *rhythm* the distribution of features like stress, quantity and tone (the *suprasegmentals* analyzed by Lehiste, 1970) among the segments inside a given unit.

[1] The paper has been jointly written by the three authors. For academic purposes, LF bears responsibility for §§ 1, 2, 6.2; DG for §§ 5, 6.1; DD for §§ 3, 4. While being solely responsible for any weaknesses or inaccuracies that may be found here, the authors would like to thank Chiara Celata, Franco Cutugno, Christine Mooshammer and an anonymous reviewer for their helpful comments.

Since tone does not play a relevant role in the varieties we are going to deal with, we will refer here only to stress and quantity as rhythm-sensitive features. To classify the different typologies of rhythm, we do not adopt the classical categories of *stress isochrony* and *syllable isochrony* (Abercrombie, 1967 after Pike, 1945, and thereafter a large amount of studies). We prefer instead those of *control* and *compensation* (first in Vékás, Bertinetto, 1991, then in Bertinetto, Bertini, 2008; see Filipponio, 2012a), which are explainable as the tendencies either to keep the distribution of quantity stable or to unbalance it, normally in favour of the prominent elements (viz. the stressed syllables). In this way, segmental changes leading to a bigger imbalance between more and less prominent elements can be ascribed to a compensating pattern, while changes balancing them (or at least no changes) may be due to a controlling pattern. Diachronically, these patterns are cyclical (Filipponio, 2012a), so that a controlling pattern can involve a rhythm structure previously modified by a compensating one; at the end, a total compensation should produce an alignment of prominences building a new control phase pattern.

The question is whether it is possible to deal with rhythm by describing processes involving a single (phonological) word – which, in other words, means asking whether a phonological word can be considered as a rhythm unit. Since phonological words are normally realized inside the speech chain, viz. an environment dominated by inter- and intra-speaker (context-sensitive) variation as well as a huge amount of coarticulation (assimilation, elision and so forth), the attempt to obtain a rhythm picture of a language starting from this situation should be immediately abandoned. In more general terms, every attempt to extract rhythmic properties from the speech chain seems to be doomed to failure: one can surely agree on this with Arvaniti (2012), who claims that rhythm classifications based on metrics such as %V, ΔV, ΔC (Ramus, Nespor & Mehler, 1999), PVI (Grabe, Low, 2002) and Varco (Dellwo, Wagner, 2003) are unsafe: their unsafety depends exactly on the fact that they try to find out regularities by measuring speech chains (phonetic utterances), without regard to prominences (stress, length) and contrasts (stressed/unstressed, long/short) – what speakers/hearers contrarily do.

For all these reasons, the definition of rhythm that we have provided above must be understood as phonological. Therefore, the search for rhythmic features should be exclusively phonology-driven (cf. Dauer, 1983; Bertinetto, 1989; Filipponio, 2012a). In that respect, the phonological word considered as a rhythm unit comes back into play, and the diachrony with it: in the sense that, if some rhythm-driven changes such as lengthening, shortening or loss of segments and/or syllables have been lexicalized, we should be allowed to consider the story of a word structure as a clue to reconstruct the rhythm history of a language. Moreover, allophonic phonological rules, as part of the synchronic phonological derivation (Loporcaro, 2015: 234), must be taken into consideration. In sum, we (try to) consider the co-occurrence of segmental phenomena as epiphenomena related to a superordinate rhythm pattern (see the discussion in § 6).

To provide an example of what we mean, consider for instance the Latin proparoxyton PĔRTĬCĂ 'pole'; doing something 'with a pole' (instrumental) would have been expressed through the ablative PĔRTĬCĀ, with a ˈCVC-CV-CVː structure requiring a stable control rhythm pattern in order to preserve vowel length contrasts in unstressed syllables (to save the contrast ablative ~ nominative). This pattern was no longer available in the late stages of Latin (cf. Loporcaro, 2015: 10-11; 59) and in early Italo-Romance, as long vowels in unstressed syllables were banned and afterwards open syllable lengthening came into force as an allophonic phonological rule: all stressed syllables became bimoraic (heavy), so that the Latin variation ˈCV (light), ˈCVː (heavy, open), ˈCVC (heavy, closed), ˈCVːC (superheavy), was reduced to ˈCVː (open), ˈCVC (closed) via ˈCV > ˈCVː and ˈCVːC > ˈCVC. In fact, having only bimoraic stressed syllables and no long vowels in unstressed syllables is already a compensation-oriented pattern, because the moraic weight will always be higher (or at least the same) in stressed syllables than in unstressed ones.

Consider now some Italo-Romance outcomes like /ˈpɛrtika/ (Tuscan), /ˈpɛːrdga/ (Eastern Apennine Emilian), and /ˈpɛrtɛgɛ/ (in the Western Lombard tiny village of Monteviasco, see Delucchi, 2016: 170): all in all, they can be interpreted as a result of the influence of different rhythmic patterns. Tuscan has kept (controlled) the late Latin/early Romance pattern (/ˈpɛrtika/ = ˈCVC-CV-CV);[2] Emilian has heavily unbalanced the word structure by syncope of the post-stress syllable and the secondary lengthening of stressed vowels (which is regular in the case of a following liquid+plosive cluster: /ˈpɛːrdga/ = ˈCVːCC-CV). In Monteviasco's dialect, on the contrary, stressed vowel lengthening is absent and unstressed vowels are fixed by full vowel harmony: this weight balance between stressed and unstressed syllable indicates a controlling rhythm pattern (Delucchi, 2016: 310) probably arisen after a compensative phase (Delucchi, Filipponio, 2013)[3].

Thus, the observation of three different outcomes of the same Latin word brings us to identify three different rhythm patterns. Obviously, all these phonological forms are prone to every kind of phenomenon as soon as they are put in the speech chain. Nevertheless, they do exist in speakers' knowledge (*langue*): as said, since rhythm should be understood as a phonological feature, a phonetic analysis dealing

[2] With some early compensative oscillation which have left a trace in syncopated proparoxytones (PERSĬCA > *pesca* 'peach', PŎSĬTU(M) > *posto* 'posed.PART.M.SG, place' etc., see Rohlfs, 1966: § 138 for further examples), sometimes restored (a control-driven balancing?) with a non-etymological /a/ (JŬVĔNE(M) > *giovane* 'young.SG', CHRŎNĬCA > *cronaca* 'chronicle' etc., see Rohlfs, 1966: § 139 for further examples).

[3] The Western Lombard onset-epenthesis pattern /ˈforna/ < **forn* < FŬRNU(M), which builds a full disyllabic word, can be considered another evidence of a compensation-to-control path (first apocope, then rebuilding of the unstressed syllable), while the Eastern Emilian rhyme-epenthesis pattern /ˈfawren/ is compatible with the persistent compensation pattern mentioned above – notice also the secondary lengthening of the stressed vowel before liquid+plosive (like in /ˈpɛːrdga/, here /o/ > /aw/), which took place before the apocope (see Repetti, 1995, following Broselow, 1992, for the definition of rhyme-dialects and onset-dialects and the discussion in Filipponio, 2012a: 77).

with rhythm must be, in a certain sense, a kind of bottom-up process, seeking for stable features, contrasts and prominences instead of neutralizing them.

2. *On length in Gallo-Italian*

In Gallo-Italian dialects, the vowel length due to the late Latin/early Romance allophonic phonological rule of syllable isochrony (see above) became contrastive (see for example Loporcaro, 2015, and the references reported).

The consonantal degemination, which is a typical feature of Northern Italian dialects (Rohlfs, 1966: § 229), can be considered the trigger of this process: ˈCV:C(V) ~ ˈCVC(V) (< ˈCVC:V) = Western Lombard (Milan) /ˈpa:s/ 'peace' (< PA$CE(M)) ~ /ˈpas/ 'step' (< PAS$SU(M)), but the explanation needs some further elements.

2.1 Contrastive length from consonants to vowels

There was a position in which consonantal length was independent from stressed syllable isochrony, viz. the pretonic one. Pretonic degemination, which is attested in the whole Northern Italo-Romance area, eliminated this option (Table 1):

Table 1 - *Effects of pretonic degemination in Western Lombard (Milan)*[4]

Latin	Word structure before degemination	Word structure after degemination	Western Lombard
*CAM$MĪ$NU(M)	…VCˈCV:CV	…VˈCV:CV	/kaˈmĩ:/ 'march'
*CA$MĪ$NU(M)	…VˈCV:CV	…VˈCV:CV	/kaˈmĩ:/'fireplace, chimney'
*CAP$PĔL$LU(M)	…VCˈCVC:V	…VˈCVC:V	/kaˈpɛl/ 'hat'
*CA$SA+**ello*	…VˈCVC:V	…VˈCVC:V	/kaˈzɛl/ 'butcher's slaughter room'

Given this situation, it would be possible to reanalyze the quantity of stressed vowels as phonological and let post-stress consonantal degemination take place without any further consequence. This could be the stage of a very conservative Gallo-Italian dialect, like that of Soglio (Alpine Lombard spoken in Canton Grisons), which still has pretonic degemination but has kept post-stress geminates. Yet, in this case we are quite sure (following Loporcaro, Paciaroni & Schmid, 2005) that consonantal length is still phonological, because this variety lacks long stressed vowels in final position, viz. in an(other) position which is independent from stressed syllable isochrony (cf. Martinet, [1975: 205]). In a case like this, lacking any further evidence, we do not have any conclusive proof to declare that vowel length has become phonological.

[4] All Western Lombard examples are taken from Arrighi (1896). The phonological transcription is adapted according to Salvioni (1884).

The next stage is shown by the dialect of Castello di Sambuca (a conservative Apennine Emilian variety spoken in Tuscany on the Adriatic watershed), where next to pretonic degemination and persistent post-stress geminates (like in the dialect of Soglio) some final long stressed vowels arose due to secondary phenomena such as the loss of final *-i* in primary or secondary hiatus (**stivali* > **stivai* > /ʃtiˈvaː/ 'boots'). Although receding in internal position, vowel length is constantly realized in the prepausal one (Filipponio, Nocchi 2010), which is enough to consider it phonological (Loporcaro 2015), but peripheral in the system (Filipponio, Garassino 2019).

A further stage is displayed by the dialect of Lizzano (Eastern Apennine Emilian, about 15km north-east from Castello di Sambuca): the set of words with a long stressed final vowel is bigger, and the phonetic realization of this feature is constant (Loporcaro, Delucchi, Nocchi, Paciaroni & Schmid, 2006), while post-stress degemination takes place more easily in non-prepausal position and in *Allegroform* than for instance in an isolated realization (see the data in Filipponio, 2012b and the comparison between two Ligurian dialects in Garassino, Filipponio, forthcoming).

This picture shows that the co-occurrence of pretonic degemination and long stressed final vowels is a sufficient diagnostic tool to confirm the presence of contrastive vowel length, regardless of the behavior of post-stress consonants. However, it may be unnecessary to consider stressed final vowels, if post-stress degemination occurs anyway, like in the dialect of Monte di Badi (close to Castello di Sambuca, but in Emilian territory): in this case, one could ask whether the systematic shortness of stressed final vowels is either a conservative feature or should be ascribed to a secondary contact effect with the neighboring Tuscan dialects; but the whole absence of long consonants is obviously enough to claim the phonological value of vowel length (Filipponio, 2012b: 244-246).

Finally, a canonical Gallo-Italian dialect will display both complete degemination and long final stressed vowels. The combinations discussed here are summarized in the table below.

Table 2 - *At the origins of Gallo-Italian contrastive vowel length (henceforth: CVL)*

Dialect	**Pretonic degemination**	**Final long stressed vowels**	**Post-stress degemination**	**CVL-Check**
⌜Tuscan⌝	-	-	-	No
Soglio	+	-	-	No
Castello	+	only prepausal	-	yes (periph.)
Lizzano	+	+	residual gemination	yes
Monte di Badi	+	-	+	yes
⌜Gallo-Italian⌝	+	+	+	yes

2.2 Post-stress consonants in Bolognese and Portorino

According to previous studies, the Eastern Emilian dialect of Bologna (henceforth: Bolognese or BO) and the Western Ligurian one of Porto Maurizio (henceforth: Portorino or PM), which have stable CVL, show an incomplete post-stress degemination.

For example, since Coco (1971) up to the more recent Avesani, Vayra & Longo (2016), Bolognese has been pointed out to have long consonants following a short stressed vowel. Similarly, Filipponio, Garassino (2019) found that Portorino shows post-stress long consonants when the stressed vowel is short, at least when the target words occur in internal and non-focalized position. In other words, these two Gallo-Italian varieties are not completely "typical" and are worthy of further investigation.

3. *Materials and Method*

3.1 Stimuli, participants and procedure

The data set of the current study comes from different recording sessions, conducted by the authors at two different times. The former was carried out by the first author in 2006 across the so-called Valle del Reno for his doctoral dissertation (issued in 2007, then expanded in Filipponio, 2012b). The latter is part of an inquiry conducted by the first and the second author during the summer of 2017 in Genoa and Porto Maurizio (IM) (for further details see Filipponio, Garassino, 2019 and Garassino, Filipponio, forthcoming). These data were originally collected for other purposes. For the sake of homogeneity, in this contribution only real words in utterance-internal position will be analyzed.

As reported in Table 3, the first sub-corpus contains data from Bologna and the close dialects of Bazzano, Castello di Serravalle, Marzabotto and Pian di Venola, which can be assimilated to the Bolognese variety for their dialect characteristics (Filipponio, 2012b). All the other recordings come from the Ligurian coastal city of Porto Maurizio, belonging to the dialectal group of Western Ligurian (cf. Forner, 1988).

The Bolognese data were recorded by means of a Sony TCD-D100 DAT recorder and a Sony ECM-717 tie-clip microphone on Sony TDK DAT tapes, while the Portorino ones by means of a Zoom H2n handy recorder.

In both cases the interviews were directly conducted by the authors in the informants' houses or in relatively silent public places. Each informant was recorded alone in a single session.

Table 3 - *Speakers*

Dialect	Abbreviation	Place	Gender	Age
BO	Baz	Bazzano	M	80
	Bol	Bologna	M	68
	Cds1	Castello di Serravalle	M	78
	Cds2	Castello di Serravalle	F	84
	Mar	Marzabotto	M	69
	Pdv	Pian di Venola	M	67
PM	AcTo	Porto Maurizio	M	66
	BoLu		F	77
	LaMa		F	68
	LuTo		M	70
	TeAn		M	38

On the whole, 11 speakers were analyzed; i.e., 6 speakers (1 female and 5 males) for the Bolognese area and 5 (2 females and 3 males) for Porto Maurizio. All speakers were elderly (aged 66-84), except one from Porto Maurizio, aged 38. None of them reported current or past speech or hearing disorders. All the speakers were born and lived in Porto Maurizio or Bologna and surrounding areas and fluently used the local dialect in everyday conversations.

The stimuli, 14 minimal or sub-minimal pairs, are reported in Table 4. The word pairs (8 for BO and 6 for PM) are paroxytones or oxytones which differ in the length of the stressed vowel (long or short). The speakers were asked to orally translate in their native dialect the sentences (each different) that one of the authors read in Italian.

The whole vowel inventory is covered for BO, while it lacks a minimal pair representative of the high anterior vowel /i/ in PM. Although the target items are not the same in the two dialects, they appear in comparable phonetic contexts. The post-tonic consonant is a fricative, a plosive or a sonorant in both dialects.

The experimental corpus thus includes a total of 137 tokens (80 tokens for BO and 57 for PM)[5].

[5] For various reasons (different internal developments, recording problems, etc.) some of the stimuli are missing: /ˈtaːk/~/ˈtak/ for Cds2, Pdv and Mar; /ˈfaːt/~/ˈfata/for Cds1 and Cds2; /ˈbaːla/~/ˈbala/ for Pdv and Mar; /ˈmeːl/~/ˈmel/for Pdv and Cds2; /ˈtroːpa/ for Mar.

Table 4 - *The list of the stimuli*

Vowels	BO	PM
/aː/ ~ /a/	/ˈsaːk/ 'sack' ~ /ˈsak/ 'dry' /ˈtaːk/ 'heel' ~ /ˈtak/ 'toll' /ˈfaːt/ 'done' ~ /ˈfata/ 'slice' /ˈpaːs/ 'pass' ~ /ˈpas/ 'fish' /ˈbaːla/ 'ball' ~ /ˈbala/ 'bubble'	/ˈnaːzu/ 'nose'~ /ˈmazu/ 'May'
/eː/ ~ /e/	/ˈmeːl/ 'honey' ~ /ˈmel/ 'thousand'	/ˈseːne/ 'meals' ~ /ˈsene/ 'ash' /ˈpeːzu/ 'weight' ~ /ˈpezu/ 'worse'
/oː/ ~ /o/	/ˈtroːpa/ 'too much.F.SG' ~ /ˈtropa/ 'troop' /ˈkoːr/ 'heart' ~ /ˈkor/ '(he/she/it) runs'	-
/ɔː/ ~ /ɔ/	-	/reˈpɔːsu/ '(I) rest' ~ /ˈpɔʃu/ '(I) can'
/uː/ ~ /u/	-	/ˈduːse/ 'cake, sweet.SG'~ /ˈduze/ 'twelve'
/yː/ ~ /y/	-	/ˈfryːtu/ 'fruit' ~ /ˈbrytu/ 'ugly.M.SG, dirty.M.SG'

3.2 Data processing and statistical analysis

The recordings were imported into Praat (Boersma, Weenink, 2016), where the stressed vowel and the post-tonic consonant were manually segmented and annotated. Phoneme boundaries were located at the nearest zero crossing according to the full formant structure criterion (Machač, Skarnitzl, 2009), after a visual inspection of the broadband spectrogram and the waveform. In difficult cases, we relied on repeated listening.

The duration of each annotated segment was automatically measured by using a Praat script. For analyzing the intensity, another Praat script extracted the following values: mean intensity of the stressed vowel, the intensity value and the corresponding time at the maximum peak and at the right end of the stressed vowel.

The statistical analysis was carried out by means of the software R (R Development Team, 2018) and the *lme4* package (Bates, Mächler, Bolker & Walker, 2014). Plots were realized with the help of the package *ggplot2* (Wickham, 2016).

In order to take into account the multiple factors involved, including the random variation, linear mixed models (cf. Baayen, 2008; Winter, 2013; Levshina, 2015) were used. The dependent variable was absolute Vowel Duration in ms for the vowel duration analysis and Post-stress Consonant Duration in ms for the consonant duration analysis. The fixed effects were: (a) Dialect (with two levels: BO / PM); (b) Vowel Length (phonologically short/long vowels); (c) Vowel Type or Consonant Type. In each model, Speakers and Target Items were considered random factors. Different models with random intercepts including a two-way interaction term (Dialect * Vowel Length) were incrementally built and tested for significance by comparison with ANOVA based on the Likelihood ratio test. The best model was selected in each case based on the AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion) goodness-of-fit values (i.e., the best model being the one with the smallest AIC and BIC values; for a more detailed discussion, cf. Levshina 2015: 194).

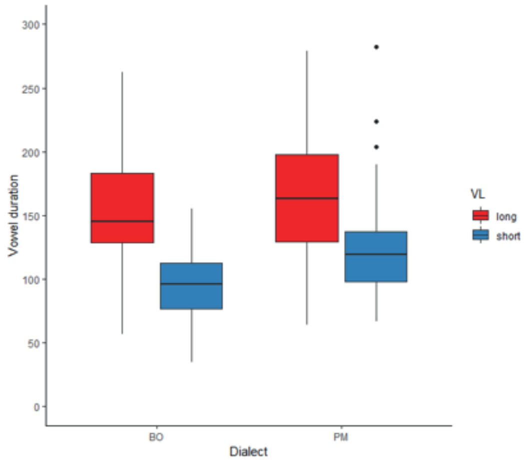
4. *Duration*

4.1. Stressed vowels

In order to describe consonant length and evaluate the nature and the extent of the post-tonic consonant gemination in the varieties at issue, we will consider the whole picture of duration. Firstly, the length distribution of vowels and consonants will be compared between BO and PM. Moreover, being aware of the perceptual salience of the durational ratios to the neighboring sounds in the speech signal, the relationships between vowels and consonants, on the one hand, and long and short segments, on the other hand, will be presented.

The boxplots in Figure 1 show the distribution of vowel duration in BO and PM as a function of phonological length. Long vowels are in red, short vowels in blue.

Figure 1 - *Distribution of vowel duration in the Bolognese area (BO) on the left and Portorino (PM) on the right as a function of the phonological length of the vowel (short vowels in blue and long vowels in red)*



As expected on the basis of the previous literature (cf. for example Uguzzoni, Busà, 1995; Filipponio, 2012b; Loporcaro, 2015; Garassino, Loporcaro & Schmid, 2017; Filipponio, Garassino, 2019), vowel length differences in stressed syllables are clearly realized in both dialects, the median being 96 ms for V vs 145 ms for V: in BO and 119 ms for V vs 163 ms for V: in PM.

Despite the limited and not fully representative composition of the corpus, a series of linear mixed models was performed in order to determine the statistical significance of the results. The best model selected from the analysis (the results of which are shown in Table 5) is as follows: Vowel Duration ~ Vowel Length + Dialect + Vowel Type + Speaker + Target Item.

Table 5 - *Results of the analysis with linear mixed models for stressed vowel duration as the dependent variable*

	β	**Std. Error (SE)**	**Df**	**t value**	**Pr(>|t|)**	
(Intercept)	161.723	12.740	16.665	12.694	5.45e-10	***
VL short	-47.396	7.124	29.144	-6.653	2.63e-07	***
Dialect PM	39.902	19.403	18.207	2.056	0.05436	.
V_type /e/	-24.041	11.114	31.852	-2.163	0.03814	*
V_type /o/	-34.666	10.730	26.275	-3.231	0.00331	**
V_type /ɔ/	-54.043	17.681	33.441	-3.057	0.00438	**
V_type /u/	-39.927	16.929	29.008	-2.358	0.02530	*
V_type /y/	-61.516	17.243	31.017	-3.568	0.00119	**

The effect of Vowel Length on the dependent variable (i.e., Vowel Duration) proved highly significant in both BO and PM (β = -47.40; SE = 7.12; $p < 0.001$). This systematic difference between long and short vowels is consistent with the traditional claim of a phonological opposition of vowel length in these varieties.

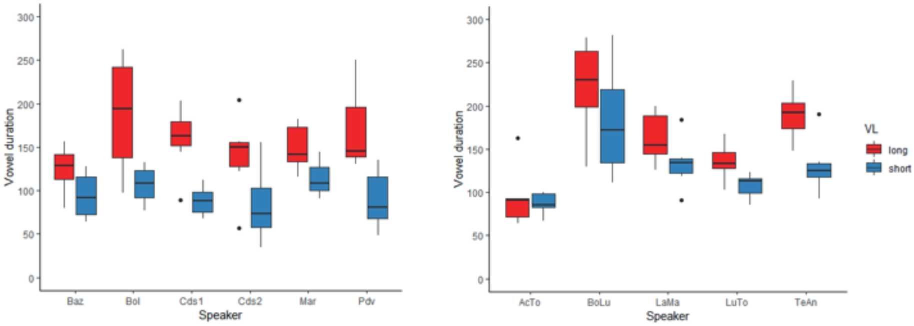
Moreover, the factor Dialect might have an effect on Vowel Duration (β = 39.90; SE = 19.40; $p = 0.05$), according to which BO vowels are generally shorter than PM ones, thus suggesting a tendency for a higher average speech rate in BO. However, this tendency is yet to be verified on the basis of a larger corpus.

Finally, as expected (cf. Lehiste, 1970: 18; Marotta, 1985), the effect of the Vowel Type on duration was confirmed to be significant, each vowel having its own intrinsic duration (compared to the reference value /a/).

If we consider the distribution of durational values separately for each speaker, the picture outlined so far changes slightly.

The graphs in Figure 2 plot the duration measures of the stressed vowel for each BO and PM speaker, respectively. Comparing the phonetic realization of vowel length across speakers, in BO all the speakers clearly realize vowel length oppositions. On the other hand, PM speakers show a higher inter-individual variability, ranging from AcTo, who essentially does not distinguish long and short vowels, to the outstanding contrasts of TeAn, the youngest speaker in the PM group.

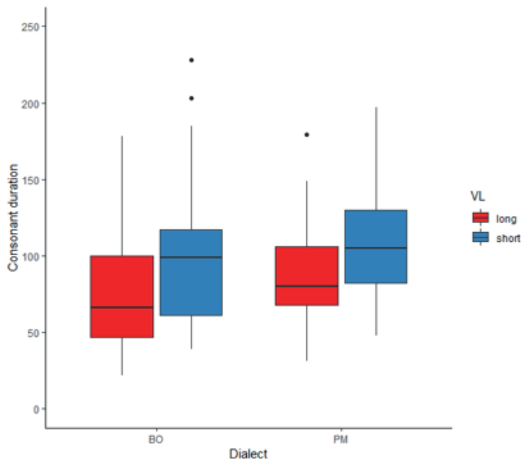
Figure 2 - *Distribution of vowel duration per speaker in BO (on the left) and PM (on the right) as a function of the phonological length of the vowel (short vowels are in blue and long vowels in red)*



4.2 Post-stress consonants

Figure 3 shows the durational values of BO and PM consonants following a long (in red) and a short vowel (in blue).

Figure 3 - *Distribution of consonant duration in BO (on the left) and PM (on the right). Consonants preceded by short stressed vowels are in blue and the ones preceded by long vowels in red*



The medians reveal a temporal difference between the post-stress consonants in both dialects (66 ms for (V:)C vs 99 ms for (V)C in BO and 80 ms for (V:)C vs 105 ms for (V)C in PM), although the boxes are more overlapping than in Figure 1. Even in this case, the results in Table 6 from the best statistical model (Post-tonic Consonant Duration ~ Vowel Length + Dialect + Consonant Type + Speaker + Target Item) show that the duration of the segment is significantly affected by Vowel Length (β = -22.57, SE = 7.62, $p < 0.01$). This finding supports the hypothesis that in these varieties gemination of the post-tonic consonant following a short vowel is a constant feature. In addition, the Type of the consonant proved to have a significant effect just for the liquids ($\beta = -30.57$, SE = 13.56, $p < 0.05$, compared to fricatives chosen as reference value), which were the shortest ones.

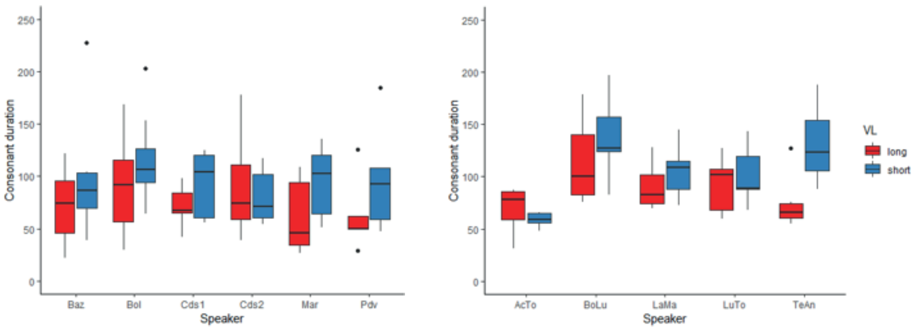
Table 6 - *Results of the analysis with linear mixed models for post-stress consonant duration as the dependent variable*

	β	**Std. Error (SE)**	**Df**	**t value**	**Pr(>|t|)**	
(Intercept)	77.348	12.422	32.225	6.227	5.48e-07	***
VL short	22.566	7.624	25.689	2.960	0.00654	**
Dialect PM	9.161	13.553	26.627	0.676	0.50488	
C_type lqd	-30.570	13.560	27.595	-2.254	0.03230	*
C_type nsl	-15.992	15.648	25.384	-1.022	0.31644	
C_type stp	18.561	10.948	31.941	1.695	0.09973	.

Turning to by-speaker distribution, consonant length appears more variable than the vowel one, as pointed out in Figure 4. All BO speakers realize length differences, except Cds2. In PM, the same speaker who did not discriminate vowels on the basis of length, i.e., AcTo, does not follow the general trend either. A similar distribution concerns also LuTo. Once again, the strongest opposition is realized by TeAn, who is the only young speaker in the PM group.

All in all, the greater variability in consonant length, combined with less clear contrasts, is not surprising in varieties where vowel length is expected to be phonological (cf. Filipponio, 2012b, Loporcaro, 2015; Forner, 1988, Filipponio, Garassino, 2019). It could be considered as a further evidence of the fact that vowel length is still the only relevant feature for distinguishing minimal pairs in these dialects and gemination is just an allophonic, accessory feature.

Figure 4 - *Distribution of post-stress consonant duration per speaker in BO (left) and PM (right) (consonants after short vowels are in blue and consonants after long vowels in red)*



4.3 Paradigmatic and Syntagmatic Ratios

We now consider the paradigmatic relationship between long and short stressed vowels, on the one hand, and post-tonic consonants following short and long vowels, on the other. As shown in Table 7, short vowels correspond to 62% of long vowels in BO or, inversely, the long vowel is 38% longer than the short one. This difference is less marked in PM, where the ratio between the two segments is 0.79, i.e., long vowels are just 21% longer than the short ones.

On the contrary, the ratio between the consonants is similar in the two varieties, thus confirming once again the auditory impression and the preliminary results in Filipponio, Garassino (2019) of a durational difference in post-tonic position for PM.

Table 7 - *Ratio between long and short vowels and between consonants following short and long vowels*

	V/V:	(V:)C/(V)C
BO	0.62	0.78
PM	0.79	0.82

By observing the almost parallel paradigmatic relationships of long/short vowels and the corresponding post-stress consonants in PM (0.79 and 0.82, respectively), we could obtain a further hint about the precarious condition of contrastive vowel length in Western Ligurian, along with the increasingly small number of minimal pairs (cf. Garassino et al., 2017; despite the experimental results in Filipponio, Garassino, 2019).

As far as the syntagmatic relationships are concerned, Table 8 provides the ratio between the stressed vowel and the post-tonic consonant. It summarizes the duration relationships in both dialects and, at the same time, tries to reduce the distortions related to potentially different speech rates.

Table 8 - *Ratio of stressed vowel to post-tonic consonant*

	V:/C	V/C
BO	2.5	1.2
PM	2	1.3

Long stressed vowels are longer than the following consonants, from 2 (for PM) to 2.5 (for BO) times. On the contrary, short vowels tend to be as long as the following consonant. The value of the ratio in favor of the vowel is due to the fact that the absolute duration of the vowels is generally greater than that of the consonants. In other words, there is an initial durational disparity between vowels and consonants, as emerged in the descriptive analysis: on average, consonants are always shorter than vowels and, even when longer, they have the same or a smaller duration than short vowels.

In light of this observation, despite the limited differences between long and short vowels on the paradigmatic level for PM, from a syntagmatic perspective in both varieties vowel quantity contrasts appear still stable and compatible with their interpretation as phonological features.

5. *Intensity*

From the results provided in § 4, one can assume that both varieties are quite similar in terms of allophonic consonant duration, at least in a non-prepausal and prosodic unmarked context. This similarity, however, may appear 'suspicious', especially with regard to allophonic gemination. Bolognese and Portorino have, in fact, quite different features from a prosodic-rhythmic point of view. Bolognese is fully a compensation language (see above § 1), while in Portorino, on the other hand, like in the other Ligurian dialects, compensation phenomena have manifested themselves less dramatically than in Emilian dialects, as is proven by the treatment of unstressed vowels. On the one hand, they have almost completely disappeared in Bolognese, while, on the other hand, they have generally been maintained in Portorino. It is sufficient to consider the developments of Latin PŎRTĬCU(M) 'arcade', a word with a similar structure to PĔRTĬCA (whose developments have been discussed in § 1), to observe the difference between Bolognese /poːrdg/ (with syncope, apocope, lengthening of stressed vowel) and Portorino /ˈpɔrtegu/ (without any of these phenomena).

For this reason, we have decided to include a tentative analysis of the intensity contours of the stressed vowels. In general, absolute data gathered with different instruments in different situations are not fit for this purpose: therefore, the stable difference that was found between the median intensity values of stressed vowels uttered by Bolognese (74.7 (±5) dB) and Portorino speakers (63.8 (±5) dB) needs further investigation and cannot be included in the discussion[6]. In relation to the statements about contrasts and prominences made in § 1, our data offer nonetheless the possibility of verifying if there are any stable intensity patterns whose identification hinges on relative rather than on absolute values.

In light of the already mentioned difficulty of treating intensity, aggravated here by the scarcity of the corpus (n = 137), we have chosen to refrain for now from a statistical analysis, opting instead for some qualitative remarks based on data visualization[7]. Thus, we carried out a preliminary analysis concerning three intensity measures (inspired by previous phonetic research, such as Fischer-Jørgensen, Jørgensen, 1965; Spiekermann, 2000 and 2002 and Mády, Tronka & Reichel, 2005). In what follows, we will then visually inspect and discuss the following parameters:

a. The absolute distance (in ms) from the peak of intensity to the vowel offset, i.e., *Vowel Offset time – Maximum intensity time*;

[6] He, Dellwo (2016: 247) observe, for instance, that "only a turn of the head can lead to a drastic drop of the overall intensity at the receiver's ear (or a microphone)". Such problems are solved by a careful gathering of data and may perhaps become less relevant as the analysis deals with stable and/or contrastive features of intensity contour patterns (viz. with the phonological interpretation of phonetic data, see above § 1).

[7] In fact, more ingredients are needed in order to attempt a careful quantitative study of intensity, such as: truly comparable data for both varieties; normalized measure of intensity (relying, for instance, on z-scores); more fine-grained measures of the intensity curve (i.e., instantaneous velocity of certain specific points vs. average velocity, etc.).

b. The relative distance from the peak of intensity to the vowel offset in relation to the vowel duration, i.e., (*Vowel Offset time – Maximum intensity time*)/*Vowel duration*;
c. The average velocity[8] calculated in relation to the intensity peak and the vowel offset, i.e., (*Vowel Offset intensity – Maximum intensity*)/(*Vowel Offset Time – Maximum Intensity Time*).

Based on the a. and b. measures, we are able to infer the position of the peak of intensity, while c. informs us about the steepness of the curve. As we will see in § 6.1, this information is useful for assessing the relation between the stressed vowel and the following consonant in a phonological perspective.

Figures 5 and 6 show the absolute (ms) and the relative distance from the intensity peak to the vowel offset in both dialects.

Figure 5 - *Absolute distance from intensity peak to vowel offset*

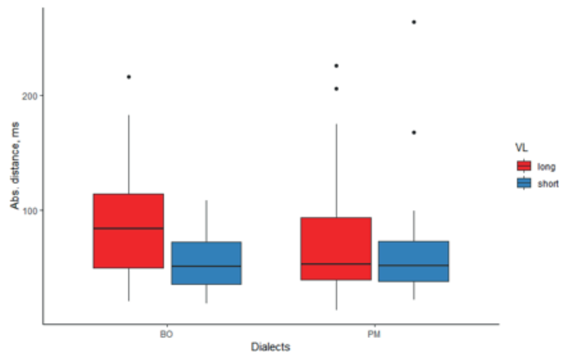
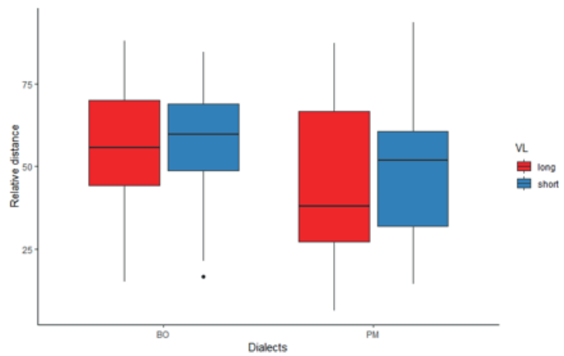


Figure 6 - *Relative distance from intensity peak to vowel offset*



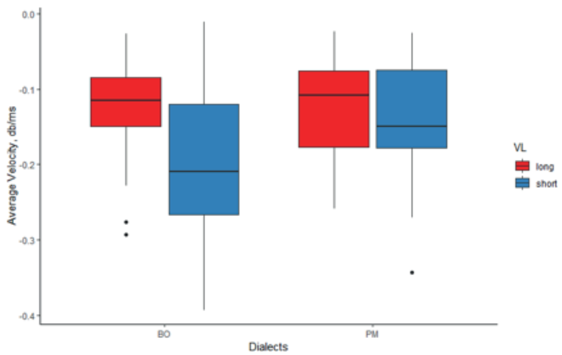
First, it is worth noting that the median values (in particular those of relative distance) hide a high degree of variability. Nevertheless, by crossing the results of ab-

[8] In physics, the average velocity is the distance traveled by an object divided by the time elapsed. In our case, this measure refers to the distance in the intensity curve between the peak of intensity and the vowel offset (divided by the time elapsed between the two).

solute and relative distance, one finds that short stressed vowels behave in the same way, while long vowels show two different patterns: Bolognese speakers place the intensity peak earlier than Portorino speakers, as displayed by the higher absolute and relative distance of the former compared to the latter[9]. In other words, they reach the intensity peak more quickly[10].

Moreover, Figure 7 tells us that Bolognese speakers have bigger fluctuations between peak and offset, which is relevant, because their higher peak values do not correspond in our data to higher offset values, the latter being similar to those of Portorino. For this reason, the difference between the long vowels of the two dialects is 'smoothed' (i.e., the higher distance between peak and offset in Bolognese is balanced by a higher steepness), while Bolognese short vowels fall to the offset more quickly than Portorino ones.

Figure 7 - *Average velocity*

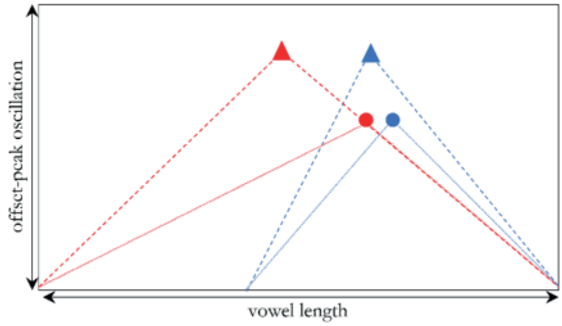


This data, although scarce and extremely variable, suggests that the intensity contour patterns of Bolognese and Portorino are different. In Figure 8 we try to informally visualize this difference. Red and blue lines respectively show long and short vowels; Bolognese is represented by dashed lines and triangle peaks, Portorino by dotted lines and circle peaks.

[9] Following the suggestion of a reviewer, we could also interpret the intensity contours in light of the data on vowel duration (cf. § 4). The absolute intensity-related measure in Figure 5 seems to confirm that vowel length contrasts are more salient in Bolognese than in Portorino: the difference between short and long stressed vowels is, in fact, more striking from an intensity perspective in the former than in the latter. Interestingly, in Figure 6, in which vowel duration is, so to speak, 'neutralized', the relative distance from the peak of intensity to the vowel offset is different between long and short vowels only in Portorino: this suggests that the Ligurian variety, unlike Bolognese, signals vowel length contrasts by means of the intensity contour. Although this data hints at very interesting aspects of the relationship between duration and intensity in the two varieties, due to the absence of a quantitative analysis we prefer to abstain from further speculation.

[10] The mean values of average velocity refer to the difference between vowel offset and the peak of intensity, which accounts for bigger or smaller fluctuations throughout the contour – viz. a relativized value, different than the bare peak of intensity value briefly reported above.

Figure 8 - *Intensity patterns of Bolognese and Portorino*



In the following paragraph, we will explain the meaning of these patterns.

6. *Discussion*

6.1 Eastern Emilian as a syllable-cut language?

In § 4 we have seen that Bolognese and Portorino, viz. both varieties with contrastive vowel length, show a partial post-stress gemination, at least in the elicited contexts. In § 5 we have seen that these dialects seem to have different intensity patterns (although we were not able to conduct a thorough analysis of that). By observing their word structure, we also know that they underwent two different rhythm patterns: today's Bolognese is the result of a strong compensation pattern, while Portorino stabilized its structure after the first (late Latin and early Gallo-Italian) compensative drift. In light of these typological differences, it is plausible to consider the (superficially) similar allophonic gemination as the result of different underlying phenomena.

If we now consider the Bolognese intensity pattern, we observe that the early peak in long vowels as well as the high steepness in the short vowels can be correlates of loose and close contact respectively (as defined by Martinet, 1966) between the stressed vowel and the following consonant. This picture is compatible with the well-known syllable-cut hypothesis, which has been first put forth in the literature concerning Germanic languages (in particular, Standard High German and Dutch): in these varieties, one observes an interaction between stressed vowels and post-stress consonants, in such a way that open syllables can only host long vowels, whereas closed syllable only present short vowels (cf. Fischer-Jørgensen, Jørgensen, 1965; Vennemann, 2000; Spiekermann, 2002; Uguzzoni, Azzaro & Schmid, 2003). Stressed vowels in closed syllables as in the German word *Mitte*, 'middle', cannot fulfil their course since they are 'cut' by the following consonant. As a result, these vowels are short, in terms of vowel quantity, and lax (in this case, [ɪ]), in terms of vowel quality. On the other hands, stressed vowels in open syllables, such as in the German word *Miete*, 'rent', cannot be cut from the following consonant, which belongs to a different syllable, and are long and tense ([i]).

It is worth noticing that Uguzzoni et al. (2003) have suggested on the basis of several in-depth empirical analyses (in particular, see Uguzzoni, Busà, 1995) that some Emilian dialects can be considered syllable-cut varieties. The emergence of this pattern may be explained by "the need to safeguard the short/long distinction in vowels under dynamic stress" (2003: 2718; according to this perspective, syllable-cut serves as a structural 'crutch' for the survival of vowel length contrasts)[11]. Filipponio (2012b: 250-251) has partially revised this view, because a relevant parameter of syllable-cut languages, viz. the absence of a short stressed vowel at the end of the phonological word, is not satisfied by Emilian dialects (cf. Bolognese /prɛː/ 'meadow' but /ba/ 'oxen', /bvo/ 'drunk'), so that we can at least say that Emilian dialects are still quantity languages which are about to become syllable-cut.

However, in spite of its theoretical attractiveness, syllable-cut has proven difficult to study empirically because its phonetic features have not been properly defined yet. One of the main reasons for this probably lies in the complex interaction between syllable-cut, word stress, vowel quantity and quality. In other words, it is very difficult to disentangle its phonetic correlates from the ones of quantity and quality (as is shown very clearly in Mády et al. 2005)[12]. Moreover, the main phonetic correlate of syllable-cut distinctions has been suggested to be intensity, which is a notoriously elusive prosodic feature.

6.2 Between phonology and typology

We can now put together the pieces of the puzzle: If 1. Bolognese shows loose contact with long stressed vowels and close contact with short stressed vowels, and 2. this alternation is compatible with a (at least partially) syllable-cut pattern, whose most important correlate is intensity, then we can suppose that the allophonic post-stress gemination is another correlate of close contact. Let us consider again the different phonological word structures of Gallo-Italian dialects by taking as an example the outcomes of RŪGĬDU(M) 'rough.M.SG'

Table 9 - *Gallo-Italian outcomes of Latin* RŪGĬDU(M) *'rough'*

Dialect	RŪGĬDU(M)	Syncope	Apocope	Compensative force
Ligurian	/ˈrydegu/	-	-	↓
West Lombard	/ˈruvid/	-	+	
Lizzano	/ˈruvdo/	+	-	
Bolognese	/rovd/	+	+	

[11] These Emilian dialects also show spectral differences accompanying vowel length contrasts (cf. Uguzzoni, Busà, 1995).

[12] By means of a fine-grained comparison between Hungarian (a quantity language in which no syllable-cut effects are present) and German, Mády et al. show that the phonetic parameters singled out by Spiekermann (2000 and 2002: number of intensity peaks (*E-Zahl*), their positions (*E-Pos*) and the shape of the intensity contour (*E-Halt*)) do not seem to be specific of syllable-cut languages, since they play some role in Hungarian as well. Ultimately, vowel length may be responsible for all these observed effects.

By observing Table 9, we can hypothesize that the intensity pattern which triggers the close contact is the 'dynamic' implementation of the Bolognese strong compensative pattern.

On the other hand, given the control pattern which nowadays characterizes Portorino, one can argue that the allophonic post-stress gemination is due to other factors: maybe an incomplete degemination (as a conservative feature and not as an epiphenomen of close contact like in Bolognese), or the pressure of standard Italian (which has contrastive consonantal length); this is something we are currently not able to decide.

More generally, in proparoxytones we consider syncope as a stronger correlate of compensation than apocope because the unstressed syllable involved is closer to the stressed one. Indeed, among the consequences of the unbalancing between stressed and unstressed syllables one counts not only the lengthening of the former and the shortening or loss of the latter, but also some attraction effects revealing that the phonological word is treated like a sort of 'macrosyllable'. In this respect, the Apennine Eastern Emilian dialect of Lizzano (see above § 2) behaves exactly like Bolognese, with the only difference of the apocope, and its allophonic post-stress gemination (see above Table 2) can thus be ascribed to the same factors operating in that variety[13].

As a proof, one can observe the exceptions to the reduction of stressed vowel length in proparoxytones in Eastern Emilian and in Ligurian. Eastern Emilian proparoxytones which escape the reduction are characterized by a long tense stressed vowel followed by a sequence of consonantal segments which cannot become a syllable coda, according to Vennemann (1988) (decreasing consonantal strength, unless the first element is a plosive, which systematically undergoes close contact). On the contrary, the presence in Ligurian of proparoxytones with long stressed vowels is not related to these structural constraints, as shown in the following Table 10:

Table 10 - *Vowel length in Eastern Emilian and Ligurian proparoxytones*[14]

Latin	Lizzanese	Ligurian	Consonantal sequence
*JŬVĔNE(M)	/ˈdʒoːvne/	/ˈzuvenu/ (cf. Ghini, 2001: 171)	/v/…/n/
PĔCŎRA	/ˈpegora/	/ˈpɛːguɹa/ > /ˈpɛːgwa/ (Genoese, cf. Parodi 1902-5: 157)	/g/…/r/

[13] The interpretation of Lizzanese allophonic gemination as an epiphenomenon of compensation revises the view of Filipponio (2012b: 251).

[14] Further Emilian examples in Filipponio, 2010; Filipponio, 2012a: 78-79; Filipponio, 2012b: 298-301 (LĔPŎRE(M) > Lizzanese /ˈlevora/ 'hare' undergoes shortening in spite of the consonantal sequence /v/…/r/ because the outcome of stressed ĕ was a diphthong *[eə] – according to Filipponio, 2012b: 273 – or *[je] (see Filipponio, 2017: 274-275) and not a tense vowel). For Genoese, Toso (1997: 16) provides examples of postlexical proparoxytones with a long stressed vowel arising in verb + clitic strings (cf. Loporcaro 2015: 206).

Moreover, Eastern Emilian paroxytones which undergo the reduction display a non-etymological allophonic post-stress length explainable as close contact (Lizzano /ˈruvdo/ (Table 9) > [ˈruv·do]; /ˈpegora/ (Table 10) > [ˈpeg·ora]; cf. Malagoli, 1930), while the Ligurian ones do not (according to Ghini 2001).

Thus, the shortening of stressed vowels in proparoxytones, which is normally explained (for example by Filipponio 2012a) as a general reduction process affecting all Gallo-Italian dialects (except the cases mentioned above), could be ascribed, like post-stress gemination, to phenomena due to different factors, despite their superficial similarity. In Emilian, the shortening would be due to the attractiveness of the stressed syllables in proparoxytones, which draw to themselves the unstressed ones, causing the shortening and hence the close contact. In Ligurian, it would simply be the first step of the loss of contrastive vowel length which in other varieties affects also paroxytones (Western Lombard) and even oxytones (Intemelian Ligurian)[15].

Obviously, our hypotheses need further experimental research to be proved, but what we want to emphasize here is that segmental correspondences are sometimes due to completely different factors and that only the analysis of the rhythm patterns, considered as a phonological feature traceable in the diachrony, can provide solid explanations.

Acknowledgements

The research presented in this paper has been funded by the Swiss National Science Foundation as part of the project entitled *At the intersection between acoustic phonetics and information structure. An empirical investigation on the phonetic realization of vowel length in three Ligurian dialects* (Project Nr. 100015_178932).

Bibliography

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: University Press.

Arrighi, C. (1896). *Dizionario milanese-italiano. Con repertorio italiano-milanese.* Milano: Hoepli.

Avesani, C., Vayra, M. & Longo, V. (2016). Attrito e transfer tra dialetto e italiano regionale. Quantità e lunghezza vocalica nel parlato intramurario di Bologna. In Savy, R., Alfano, I. (Eds.), *La fonetica nell'apprendimento delle lingue*. Milano: Officinaventuno, 391-410.

Baayen, R. Harald (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv*:1406.5823.

Bertinetto, P.M. (1989). Reflections on the Dichotomy 'Stress- vs. Syllable-timing'. In *Revue de Phonétique Appliquée*, 91/93, 99-130.

[15] For a detailed discussion, see Loporcaro (2015) and Filipponio (2012b).

Boersma, P., Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0. 15 (2016). http://www.praat.org.

Broselow, E. (1992). Parametric Variation in Arabic Dialect Phonology. In Broselow, E., Eid, M. & McCarthy, J. (Eds.), *Perspectives on Arabic Linguistics IV. Papers from the 4th Annual Symposium on Arabic Linguistics*. Amsterdam/Philadelphia: John Benjamins, 7-45.

Coco, F. (1971). Effetti della degeminazione consonantica nel dialetto bolognese. In *Atti del VII Convegno del Centro per gli Studi Dialettali Italiani, Torino - Saluzzo, 18-21 maggio 1970*. Torino: Rattero, 162-167.

Dauer, R.M. (1983). Stress-timing and Syllable-timing Reanalyzed. In *Journal of Phonetics*, 11, 51-62.

Dellwo, V., Wagner, P. (2003). Relations between Language Rhythm and Speech Rate. In Solé, M.-J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcellona, 471-474.

Delucchi, R. (2016). *Fonetica e fonologia dell'armonia vocalica. Esiti di -A nei dialetti della Svizzera Italiana in prospettiva romanza.* Tübingen: Francke.

Delucchi, R., Filipponio, L. (2013). Interplay between vowel harmony and rhythm patterns in Swiss Italian dialects. Communication at the *3rd Transalpine Typology Meeting*. University of Zurich, 30.05.2013.

Filipponio, L. (2010). La quantità vocalica nei proparossitoni etimologici al confine tra toscano e gallo-italico. In Iliescu, M., Siller-Runggaldier, H. & Danler, P. (Eds.), *Actes du XXVe Congrès International de Linguistique et Philologie Romanes (Innsbruck 2007)*. Berlin/New York: de Gruyter, II, 67-76.

Filipponio, L. (2012a). Reflexionen über die Diachronie des Galloitalienischen Rhythmus. In Schafroth, E., Selig, M. (Eds.), *Testo e ritmi. Zum Rhythmus in der italienischen Sprache*. Frankfurt a.M.: Peter Lang, 61-87.

Filipponio, L. (2012b). *La struttura di parola dei dialetti della valle del Reno. Profilo storico e analisi sperimentale.* Sala Bolognese: Forni.

Filipponio, L. (2017). Prospettive, retrospettive, ricostruzioni. La tradizione scritta del bolognese (secoli XIII-XIX). In *Lingua e Stile*, 52(2), 241-282.

Filipponio, L., Nocchi, N. (2010). Diagnostica fonetica e diagnosi fonologica. Ossitoni lunghi di sillaba libera a Sambuca Pistoiese (PT). In Schmid, S., Schwarzenbach, M. & Studer, D. (Eds.), *La dimensione temporale del parlato*. Torriana: EDK, 225-248.

Filipponio, L., Garassino, D. (2019). Center and Periphery in Phonology: a "stress-test" for two Ligurian Dialects. In *Italian Journal of Linguistics*, 31(2), 141-167.

Fischer-Jørgensen, E., Jørgensen, H.P. (1969). Close and loose contact ("Anschluß") with special reference to North German. In *Annual Report of the Institute of Phonetics of the University of Copenhagen* (ARIPUC), 4, 43-80.

Forner, W. (1988). Areallinguistik I. Ligurien / Aree linguistiche I. Liguria. In Holtus, G., Metzeltin, M. & Schmitt, Ch. (Eds.), *Lexikon der Romanistischen Linguistik. IV: Italienisch, Korsisch, Sardisch*. Tübingen: Niemeyer, 453-469.

Garassino, D., Filipponio, L. (forthcoming). The impact of information and prosodic structure on the phonetic implementation of vowel length. A contrastive analysis of two Ligurian dialects. In Teixeira Kalkhoff, A., Selig, M. & Mooshammer, C. (Eds.), *Prosody and conceptional variation. Situational conditions of communication, degree of communicational planning, and activity types as parameters for prosodic design*. Bern: Peter Lang.

Garassino, D., Loporcaro, M. & Schmid, S. (2017). La quantità vocalica in due dialetti della Liguria. In Bertini, C., Celata, C., Lenoci, G., Meluzzi, C. & Ricci, I. (Eds.), *Fattori biologici e sociali nella variazione fonetica / Social and Biological factor in Speech Variation*. Milano: Officinaventuno. 127-144.

Ghini, M. (2001). *Asymmetries in the Phonology of Miogliola*. Berlin/New York: Mouton de Gruyter.

Grabe, E., Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In Gussenhoven, C., Warner, N. (Eds.), *Laboratory Phonology 7*. Berlin: de Gruyter, 515-546.

He, L., Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. In *International Journal of Speech Language and the Law*, 23(2), 243-273.

Machač, P., Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Prague: Epocha Publishing House.

Mády, K., Tronka, K.Z. & Reichel, U.D. (2005). Syllable cut and energy contour in vowels: a contrastive study of German and Hungarian. In *ZAS Papers in Linguistics*, 42, 197-217.

Malagoli, G. (1930). Fonologia del dialetto di Lizzano in Belvedere (Appennino bolognese). In *L'Italia Dialettale*, 6, 125-196 [reissued in Filipponio, L., Loporcaro, M. (Eds.) (2011), *Giuseppe Malagoli: il dialetto di Lizzano in Belvedere*. Lizzano in Belvedere: Gruppo di Studi Capotauro, 15-86].

Marotta, G. (1985). *Modelli e misure ritmiche: la durata vocalica in italiano.* Bologna: Zanichelli.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge MA: MIT Press.

Levshina, N. (2015). *How to do Linguistics with R. Data Exploration and Statistical Analysis.* Amsterdam/Philadelphia: John Benjamins.

Loporcaro, M. (2015). *Vowel Length from Latin to Romance*. Oxford: Oxford University Press.

Loporcaro, M., Paciaroni, T. & Schmid, S. (2005). Consonanti geminate in un dialetto lombardo alpino. In Cosi, P. (Ed.), *Misura dei parametri. Aspetti tecnologici ed implicazioni nei modelli linguistici*. Brescia: EDK, 597-618.

Loporcaro, M., Delucchi, R., Nocchi, N., Paciaroni, T. & Schmid, S. (2006). La durata consonantica nel dialetto di Lizzano in Belvedere (Bologna). In Savy, R., Crocco, C. (Eds.), *Analisi prosodica. Teorie, modelli e sistemi di annotazione*. Torriana: EDK. 491-517.

Martinet, A. (1966). Close Contact. In *Word*, 22, 1-6.

Martinet, A. (1970). Remarques sur la phonologie des parlers francoprovençaux. In *Revue des Langues Romanes*, 79, 149-56 [reissued in id. (1975). *Evolution des langues et reconstruction*. Paris: PUF, 195-207].

Parodi, E.G. (1902-5). Studj Liguri. § 3. Il dialetto di Genova dal secolo XVI ai nostri giorni. In *Archivio Glottologico Italiano*, 16, 105-161, 333-365.

Pike, K.L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

R Development Team (2018). *A Language and Environment for Statistical Computing*, v 3.4.2. Wien: R Foundation for Statistical Computing. http://www.R-project.org.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of Linguistic Rhythm in the Speech Signal. In *Cognition*, 73(3), 265-292.

Repetti, L. (1995). Variazione nella sillabificazione: il caso dei dialetti italiani e romagnoli. In *Rivista Italiana di Dialettologia*, 19, 41-56.

Salvioni, C. (1884). *Fonetica del dialetto moderno della città di Milano.* Torino: Loescher.

Spiekermann, H. (2000). *Silbenschnitt in deutschen Dialekten*. Tübingen: Niemeyer.

Spiekermann, H. (2002). Ein akustisches Korrelat des Silbenschnitts: Formen des Intensitätsverlaufs in Silbenschnitt- und Tonakzentsprachen. In Auer, P., Gilles, P. & Spiekermann, H. (Eds.), *Silbenschnitt und Tonakzent*. Tübingen: Niemeyer.

Toso, F. (1997). *Grammatica del genovese. Varietà urbana e di koinè.* Genova: Le Mani.

Uguzzoni, A., Busà, M.G. (1995). Correlati acustici della opposizione di quantità vocalica in area emiliana. In *Rivista Italiana di Dialettologia*, 19, 7-39.

Uguzzoni, A., Azzaro, G. & Schmid, S. (2003). Short vs long and /or abruptly vs smoothly cut vowels. New perspectives on a debated question. In Solé, M.-J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcellona, 2717-2720.

Vékas, D., Bertinetto, P.M. (1991). Controllo vs. compensazione: sui due tipi di isocronia. In Magno Caldognetto, E., Benincà, P. (Eds.), *L'interfaccia tra fonologia e fonetica*. Padova: Unipress, 155-162.

Vennemann, T. (1988). *Preference laws for syllable stucture and the explanation of sound change.* Berlin: Mouton de Gruyter.

Vennemann, T. (2000). From quantity to syllable cuts: on so-called lengthening in the Germanic languages. In *Rivista di Linguistica/Italian Journal of Linguistics*, 12, 251-282.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications.* arXiv:1308.5499, http://arxiv.org/pdf/1308.5499.pdf.

DOI: 10.17469/O2106AISV000017

MANUELA FRONTERA, ANDREA TARASI, ELVIRA GRAZIANO

Le consonanti occlusive sorde aspirate in Calabria: un confronto tra aree dialettali[1]

L'obiettivo di questo studio è oggettivare acusticamente la presenza di post-aspirazione nelle consonanti occlusive sorde /p t k/ realizzate da parlanti calabresi. Verrà preso in esame il contesto fonetico post-nasale, analizzato nei dialetti appartenenti alle aree dialettali tre, quattro e cinque della regione, sulla base della partizione dialettologica proposta da Trumper, Maddalon (1988) e ripresa da Trumper (2016). I dati, strutturati per fasce d'età e sesso dei parlatori, sono analizzati e discussi sia all'interno di ciascuna singola area, che nel confronto fra le aree di riferimento.
Il presente lavoro si inserisce all'interno di una ricerca più ampia, che mira alla descrizione del fenomeno in esame nell'intero contesto calabrese.

Key words: VOT, post-aspirazione, occlusive sorde, dialetti calabresi.

1. *Introduzione*

1.1 VOT: correlati acustici e tendenze

Il fatto acustico per eccellenza che denota l'aspirazione è il *Voice Onset Time* (VOT), cioè il tempo che va dal rilascio della consonante occlusiva all'inizio della sonorità della vocale successiva, e che termina la propria azione nel momento in cui inizia la vibrazione delle corde vocali. Tale periodo, successivo alla fase di chiusura, è altresì definito come *post-aspirazione*, e contrapposto ai casi di aspirazione pre-chiusura ossia di pre-aspirazione (Stevens, Hajek, 2010) caratterizzanti i suoni occlusivi. I valori di post-aspirazione, cioè del succitato ritardo nell'attacco della sonorità, si rivelano un prezioso indice di specificità a livello cross-linguistico, determinante e caratterizzante per ciascuna lingua o varietà (Cho, Ladefoged, 1999). Le tre categorie principali usate per descrivere le variazioni in durata del VOT nelle lingue del mondo sono: *voicing lead VOT*, *short lag VOT* e *long lag VOT* (Lisker, Abramson, 1964), dove le consonanti aspirate figurano come *long lag*, con valori di VOT uguali o superiori ai 60 ms. Cho, Ladefoged (1999) individuano, invece, un limite arbitrario settato sui 50 ms di VOT per distinguere le occlusive sorde non aspirate (*short*

[1] I tre autori hanno sviluppato congiuntamente il lavoro proposto. Ciascuno ha etichettato e analizzato i dati relativi a una delle tre aree dialettologiche indagate, secondo la metodologia esposta al §2. Per fini accademici, i paragrafi 1.1 e 1.2 sono stati redatti dagli autori Frontera e Tarasi; a Frontera si devono la stesura del paragrafo 1.3, della sezione 2 (Metodologia della ricerca), del paragrafo 3.1 e la discussione finale sui risultati (§4); a Tarasi si deve la stesura del paragrafo 3.3; a Graziano si deve la stesura del paragrafo 3.2.

lag) da quelle aspirate (*long lag*). Comparando i vari metodi di categorizzazione, Nodari, Celata & Nagy (2019), infine, giungono a definire come aspirate tutte le occorrenze comprese nel range di durata di 30-100 ms.

Numerosi studi hanno dimostrato che il VOT varia in relazione a fatti specifici: il primo parametro influente, in tal senso, è l'arretramento del punto di articolazione della consonante occlusiva sorda, per cui a consonanti più anteriori corrispondono durate di VOT più brevi rispetto alle posteriori (Nearey, Rochet, 1994; Sorianello, 1996; Cho, Ladefoged 1999). La tendenza a una durata maggiore del VOT si manifesta, inoltre, quando i suoni occlusivi (ancor più se velari) precedono vocali alte e chiuse (Rochet, Yanmei, 1991; Morris, McCrea & Herring, 2008) rispetto a vocali basse o aperte. Un terzo importante parametro riguarda la posizione dell'accento all'interno della parola contenente il suono occlusivo, in base al quale il contesto sillabico postonico favorirebbe un aumento nel livello d'aspirazione della consonante ostruente (Falcone 1976; Sorianello, 1996; Romito, Ciardullo & Tarasi 2015; Nodari 2015; Frontera, 2018). Recenti ricerche hanno sottolineato che la realizzazione acustica del VOT può dipendere, in aggiunta, da variabili di natura sociolinguistica, conferendo informazioni sia sulla presenza di dinamiche di variazione, sia sullo status dei parlanti (Heselwood, McChrystal, 1999; Scobbie, 2006; Nance, Stuart-Smith, 2013; Nodari, 2015, 2017). La variabilità legata a età e sesso (Pichler, Wagner & Hesson, 2018; Wagner, 2012; Yu, Nil & Pang, 2015) è, d'altronde, determinata fisiologicamente. Proprio per questo motivo, nel presente studio, si darà spazio a ulteriori analisi relative a queste ultime variabili, allo scopo di indagare se possano rappresentare un fattore determinante nei valori di aspirazione riscontrati nei soggetti intervistati.

1.2 L'aspirazione in Calabria

L'aspirazione delle occlusive sorde in Calabria è un fenomeno che si manifesta in modo irregolare sul territorio (Falcone, 1977; Piccardi, 2018). Non a caso, negli studi condotti sui dialetti calabresi, essa non è stata mai tenuta in considerazione per stabilire confini o continuum dialettali. Tuttavia, vari approcci, prima uditivo/percettivi poi acustici, hanno cercato di "localizzarne" la presenza sul territorio regionale, corrente nella quale si inserisce lo studio qui proposto.

Il fenomeno oggetto di analisi è già attestato a partire dai primi anni '60: autori come Rohlfs (1966) e Falcone (1976) individuano, attraverso un'analisi uditiva, rispettivamente realizzazioni aspirate della dentale /t/ in posizione intervocalica e dell'intera classe /p t c k/, sia geminate che precedute dai suoni /m n r/, nella provincia di Reggio Calabria e di Cosenza. Canepari (1986), invece, individua presenza di aspirazione nelle zone centrali e meridionali della Calabria, constatando che le occlusive sorde diventano aspirate sia precedute da /N/ o /r/ sia in geminazione, mentre il fenomeno risulta meno frequente, in tutte le posizioni accentuali, se l'occlusiva sorda è preceduta da /l/. Mele, Schmid (2009) e Mele (2009) riscontrano la presenza di aspirazione nella varietà dialettale di San Giovanni in Fiore (Cosenza), dato confermato, in merito al contesto di occorrenza post-nasale, dalle successive analisi acustiche di Romito, Ciardullo & Tarasi (2015). Il caso di San Giovanni in Fiore (CS), in tale specifico contesto, è risultato un esempio lampante di esito inatteso, presentando aspirazione delle occlusive sorde, laddove i dialetti della stessa area dialettale tendono all'utilizzo delle rispettive omorganiche sonore, per effetto di processi fonologici di sonorizzazione (Trumper, Ortale 1978; Sorianello 1996). Limitatamente al contesto di indagine oggetto della presente ricerca (-nC-/-mC), gli ulteriori riscontri finora attestati sul fronte acustico riguardano prevalentemente la zona centrale della regione, ovvero la varietà regionale di adolescenti e anziani dell'area catanzarese, rispettivamente indagata da Nodari (2015) e nello studio pilota di Frontera (2018), nonché il parlato spontaneo (dal corpus HLVC di Toronto) di calabresi emigrati, analizzato da Nagy, Nodari & Celata (2018) e Nodari, Celata & Nagy (2019).

In merito agli ulteriori contesti di aspirazione, Sorianello (1996) aveva già individuato la presenza di aspirazione delle occlusive sorde sia nel dialetto che nell'italiano regionale cosentino. Qui, la durata generale del VOT, sia per l'italiano regionale che per il dialetto, risulta minore nelle occlusive sorde scempie, maggiore nelle post-vibranti e ancora superiore nelle geminate. Fanciullo, Librandi (2002) e Telmon (1993) discutono della presenza di aspirazione nell'italiano regionale calabrese sia per le occlusive sorde geminate che per quelle precedute da nasale o vibrante; pur non basandosi su studi sperimentali, secondo De Blasi (2014), nella stessa varietà, l'occlusiva sorda dentale mostrerebbe la maggiore aspirazione in contesti specifici, ovvero se intensa (-tt-) o immediatamente dopo -r- o -n-. Stevens, Hajek (2010) nello studio condotto sulla durata delle occlusive sorde geminate /p t k/ in più varietà d'italiano (tra cui quelle catanzarese e cosentina), concludono che i maggiori livelli di post-aspirazione in contesto nazionale appartengono ai parlanti di Catanzaro, in modo ancor più evidente nella realizzazione dell'occlusiva alveolare /t/. La presenza di aspirazione, nelle zone regionali limitrofe, è ulteriormente confermata sia in parlanti giovani che anziani, con picchi di durata maggiori nei contesti post-liquidi e in geminazione (Nodari, 2015; Frontera, 2018). Il tratto riscontrato in tali comunità di parlanti sembra persistere anche in contesto migratorio dopo molti anni di esposizione alla L2, sia in parlanti italo-canadesi che italo-argentini, provenienti dalle stesse aree calabresi (Nodari et al., 2019; Frontera, 2019).

Gli studi precursori alle indagini acustiche e le ricerche appena citate denotano una diffusione ampia del fenomeno lungo il territorio calabrese; tuttavia, i livelli più forti di aspirazione sembrano investire soprattutto le aree centro-meridionali della regione. Analizzando la più recente partizione dialettologica calabrese (Trumper, Maddalon 1988; Trumper 2016, si veda l'Immagine 1), l'aspirazione dovrebbe interessare maggiormente le Aree 3 e 4 e, nei contesti specificati, l'area reggina (Area 5), senza considerare il caso del sangiovannese (CS), in Area 2.

Immagine 1 - *Partizione dialettologica calabrese secondo Trumper (2016:227)*

Nonostante questo, esistono ancora molti gap da colmare riguardo alla distribuzione geografica del fenomeno, estendendo l'analisi a più fasce di età e un maggior numero di informanti per le zone già indagate, ed esaminando le produzioni delle stesse occlusive in aree ancora totalmente inesplorate. Risulta interessante, perciò, per questa ricerca, verificare se e quanto differiscono, rispetto ai dati appena enunciati, i livelli di aspirazione presenti in aree in cui tale fenomeno sembrerebbe essere atteso, ma non ancora acusticamente attestato, se non nel parlato di adolescenti (Nodari, 2015) e se esistano differenze correlate alle fasce d'età (Frontera, 2018). Sembra dunque necessario fornire un quadro più completo e vasto di tale fenomeno, includendo un campione più cospicuo di parlanti, di differenti età e provenienze.

1.3 Obiettivi e domande di ricerca

Lo scopo della presente ricerca è quello di fornire nuovi dati acustici sui livelli di aspirazione delle consonanti occlusive nelle varietà calabresi.

Primo obiettivo è quello di verificare se i valori di VOT nelle consonanti occlusive sorde in contesto post-nasale, prodotte da parlanti afferenti ad aree dialettologiche in cui ci si aspetta di riscontrare aspirazione, ottengono valori di VOT superiori rispetto a quelli riscontrati nelle ulteriori aree. Comprovata la presenza di aspirazione, si cercherà di determinare, più dettagliatamente: i) se i valori ottenuti si uniformano ai trend, riscontrati in letteratura, inerenti alle variazioni fra contesti fonetici e accentuali di occorrenza dei suoni occlusivi sordi; ii) se esiste un'area, tra quelle esaminate, in cui i risultati raggiunti siano significativamente divergenti rispetto alle altre aree; iii) se esistono differenze nelle durate di VOT determinate dall'età e dal sesso dei parlanti coinvolti.

2. *Metodologia della ricerca*

2.1 Soggetti e campione

I soggetti coinvolti nello studio proposto, e dettagliati in Tabella 1, sono in totale 36, equamente distribuiti fra le tre aree dialettologiche di interesse, ovvero 3, 4 e 5. All'interno di ciascuna area sono stati individuati due punti geografici di raccolta, costituiti da due distinti paesi/città: Catanzaro (CZ) e Settingiano (CZ) per l'Area 3; Badolato (CZ) e Serra San Bruno (VV) per l'Area 4, Rosarno (RC) e Taurianova (RC) per l'Area 5. Ciascun paese è rappresentato da 6 informanti, un uomo e una donna per ciascuno dei tre range di età considerati (18-35 anni, 36-59 anni, +60 anni), per un totale di 12 partecipanti per area. Ogni soggetto viene identificato tramite un'etichetta riportante le lettere iniziali del proprio nome e cognome, l'età, il sesso (M/F) e l'area di appartenenza (3/4/5).

Tabella 1 - *Distribuzione dei soggetti coinvolti nel corpus d'analisi*

Soggetti (n. 36)	**Range 1 (R1) (y. 18-35)**	**Range 2 (R2) (y. 36-59)**	**Range 3 (R3) (y. +60)**
AREA 3	NL22M3	VG53F3	SM82F3
(n. 12)	RC21F3	CT56F3	GF72M3
6m + 6f	GG26M3	GT59M3	PE63M3
	VM26F3	SS58M3	CB60F3
AREA 4	NT24M4	FP40M2	MRG68M4
(n. 12)	FC25F4	MMG41F4	LC68F4
6m + 6f	FS21F4	PG41F4	DM61M4
	GF26M4	GP40M4	RV68F4
AREA 5	GI22F5	GF36M5	EA60F5
(n. 12)	DM28M5	AS47F5	FM63M5
6m + 6f	NA23F5	CP55F5	FV64M5
	DL24M5	VS41M5	RM65F5

2.2 Stimoli

I valori di *Voice Onset Time* analizzati riguardano le consonanti occlusive sorde /p t k/ in contesto post-nasale di parlato dialettale. Il materiale d'analisi è stato elicitato tramite *task* consistenti nella traduzione, dall'italiano alla varietà dialettale di afferenza, di 30 parole target contenute da 18 frasi, in posizione centrale di enunciato. Le parole sono individuate tenendo conto delle diverse consonanti analizzate, del timbro delle vocali ad esse adiacenti e del contesto accentuale in cui le stesse consonanti compaiono (posizione pretonica/postonica, rispettivamente in parossitono e proparossitono), così come schematizzato in Tabella 2. Il materiale sonoro è stato acquisito utilizzando un microfono dinamico cardioide ad archetto Shure WH20, collegato a un registratore portatile Tascam DR-100 mkIII, con una frequenza di campionamento di 44.100 Hz, 24 bit, mono.

Tabella 2 - *Variabili linguistiche considerate nella presente ricerca*

Variabili Linguistiche	**Punto di articolazione**	/p t k/
	Vocale adiacente	alta > /i u/ non alta > /e a o/
	Sillaba	tonica/atona
	Accento di parola	parossitono proparossitono

2.3 Analisi

I segnali contenuti nel corpus sono stati etichettati e annotati foneticamente secondo la procedura proposta da Abramson, Whalen (2017), ovvero individuando in ciascun segmento occlusivo le porzioni di segnale relative, rispettivamente, alle fasi di chiusura (VLCLO), rilascio (REL) e aspirazione (ASP)[2]. Il rilascio corrisponde all'inizio della fase di esplosione e termina con l'inizio dell'aspirazione, se presente, o con la fine del rumore associato al burst/esplosione; l'aspirazione ha inizio con la fine del rilascio e termina nel punto in cui il rumore cede spazio al segmento vocalico (cfr. Immagine 2). Vengono inclusi nelle misurazioni anche i casi di *overlapping* tra aspirazione e attacco della sonorità, fissando il punto di *crossover* sull'aumento visibile di intensità delle formanti vocaliche (Abramson, Whalen, 2017). Il VOT è misurato a partire dalla prima esplosione della fase di rilascio della consonante occlusiva, fino all'attacco della prima onda periodica della vocale adiacente (Harrington, 2013). In ciascun suono target, le misure vengono estratte automaticamente in Praat (Boersma, Weenink, 2018) tramite la somma delle durate delle fasi di rilascio e aspirazione (REL+ASP), ove quest'ultima sia presente o ben individuabile (Kang, Whalen, 2017)[3].

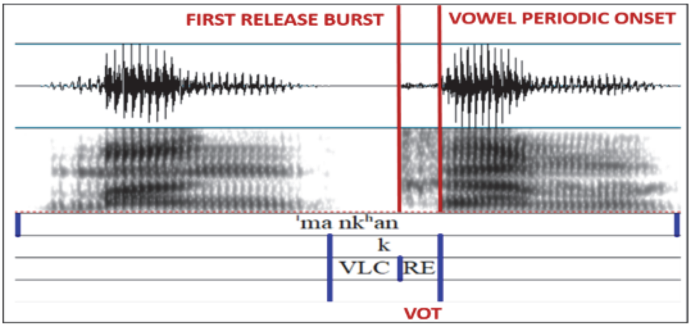
Si considerano "aspirate" tutte le occorrenze di VOT di durata compresa fra 30 e 100 ms (Nodari et al., 2019)[4].

[2] La suddivisione fra fase di rilascio e fase di aspirazione è stata adottata unicamente nei casi in cui tale differenziazione fosse ben netta e visibile.

[3] Ai fini della presente ricerca, ci si concentrerà unicamente sulle misure ottenute dalla somma delle fasi di rilascio e aspirazione. La segmentazione qui operata non trova immediato riscontro nel metodo d'analisi qui adottato, ma vuole rappresentare una base per ulteriori indagini future, focalizzate sull'analisi delle singole fasi di realizzazione dei suoni occlusivi.

[4] In questo primo confronto fra occlusive sorde post-nasali, i dati sono trattati indipendentemente dai valori di *speech-rate* dei parlanti coinvolti. Si utilizzano, dunque, valori di durata non normalizzati. Nei successivi lavori, si procederà alla normalizzazione dei dati ricavati, valutando un possibile confronto di questi ultimi coi precedenti, nonché all'incremento del campione d'analisi per ciascuna area considerata.

Immagine 2 - *Esempio di etichettatura del segnale sonoro analizzato*



2.3.1 Analisi intra-Area

All'interno di ciascuna Area, si procede, in primis, al calcolo delle medie globali relative alle durate di VOT, comprensive di tutti i contesti fonetici presi in esame. I dati descrittivi preliminari forniscono un valore di riferimento orientativo, tale da consentire di rilevare, a primo acchito, la presenza o meno di aspirazione nell'Area analizzata. Successivamente, allo scopo di verificare la conformità dei dati rispetto ai trend sulla variabilità del VOT già presenti in letteratura (cfr. §1.1), le medie globali vengono dettagliate per punto di articolazione dell'occlusiva (/p t k/), altezza della vocale adiacente e contesto accentuale della parola in cui è contenuta la sillaba, a sua volta atona o tonica, di cui la consonante target costituisce l'attacco. In merito alla classificazione delle vocali adiacenti, queste verranno differenziate tra vocali *alte*, tra cui rientrano tutte le occorrenze seguite da /i u/, e vocali *non-alte*, /e a o/. Le combinazioni accentuali prese in esame comprendono l'alternanza di consonanti precedenti una vocale tonica (pretoniche) o atona (postoniche, sia in posizione centrale che finale), rispettivamente in contesto parossitono o proparossitono.

Nella stessa Area di riferimento, si analizzano le medie legate alle variabili: i) *età*, identificata da questo momento in poi attraverso le etichette R1 (Range 1, anni 18-35), R2 (Range 2, anni 36-59) e R3 (Range 3, anni +60); ii) *sesso*. Allo scopo di individuare eventuali tracce di variabilità inter-generazionale, i dati descrittivi vengono poi integrati da analisi statistiche di comparazione fra le medie ottenute (test-t a due code per campioni indipendenti) per i tre range di età. Le stesse comparazioni si eseguono per testare la variabilità tra sessi; ulteriori analisi mirano a verificare un'eventuale correlazione fra le suddette variabili e i valori di VOT ottenuti.

2.3.2 Analisi inter-Area

I dati ricavati da ciascuna Area sono successivamente comparati fra le tre Aree di riferimento: scopo principale è, infatti, quello di comprovare e quantificare acusticamente la presenza di aspirazione nelle zone di interesse, valutando eventuali divergenze significative fra i gruppi d'analisi. I primi confronti riguardano le medie globali delle durate di VOT, in seguito differenziate per range di età.

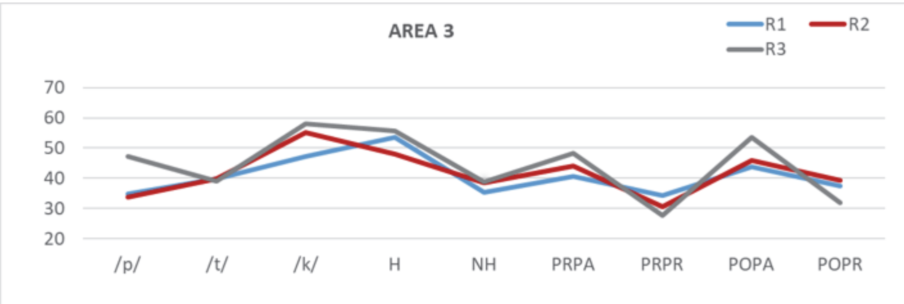
3. *Risultati*

3.1 Area 3

La durata media globale del *Voice Onset Time* attestato in Area 3 è di 43,43 ms (d.s. 23,36).

La variabilità legata al punto di articolazione delle consonanti esaminate riflette i trend letterari, con durate mediamente superiori per le occlusive velari (/k/= 53,11 ms, d.s. 25), seguite dalle consonanti dentali (/t/= 39,74 ms, d.s. 21,73) e dalle bilabiali (/p/= 39,14 ms, d.s. 21,17). Tale tendenza si mantiene stabile in tutte le fasce di età esaminate, con eccezione di R3, i cui parlanti producono aspirazione mediamente più lunga in contesto velare, seguito da quello bilabiale e, in ultimo, da quello dentale. In linea con i dati relativi alle differenze di VOT determinate da vocali adiacenti alte e non alte, i valori ottenuti dai parlanti di Area 3 si conformano, ancora una volta, alle tendenze prestabilite, per cui in tutti i range di età le vocali più alte determinano durate di VOT maggiori (M=52,13 ms, d.s. 25,11) rispetto a quelle meno alte (M=37,05 ms, d.s. 19,73). Come già riscontrato da Sorianello (1996) per i parlanti calabresi di Area 2, ma confermato dai dati ottenuti da Nodari (2015) e Frontera (2018) in Area 3, il contesto accentuale postonico è associato a durate di VOT maggiori, sia all'interno di parossitoni (M=46,72 ms, d.s. 25,04 vs. pretoniche con medie uguali a 43,78 ms, ds. 24,76) che di proparossitoni (M=36,81 ms, d.s. 18,67 vs. pretoniche con medie uguali a 31,27 ms, d.s. 15,25) (cfr. Grafico 1).

Grafico 1 - *Valori medi di VOT in Area 3, differenziati per contesto fonetico*



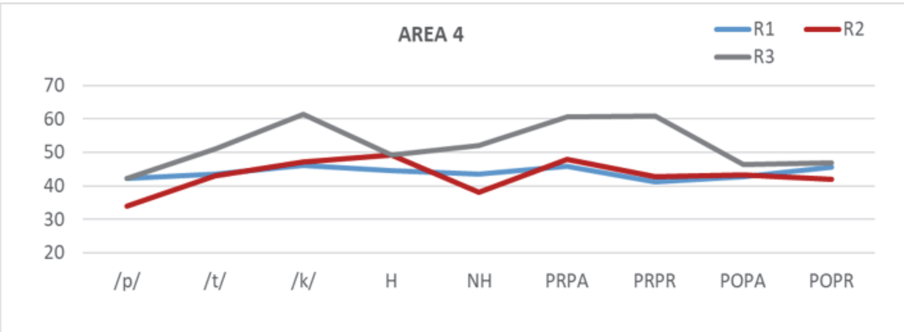
H	Alte	POPA	Postonica in Parossitono
NH	Non Alte	POPR	Postonica in Proparossitono
PRPA	Pretonica in Parossitono	PRPR	Pretonica in Proparossitono

Dal confronto fra medie, eseguito tra i tre range di età degli informanti, non emergono differenze significative cui imputare una presunta variabilità inter-generazionale nei livelli di aspirazione. Il fattore *sesso* sembra avere, al contrario, un peso maggiore e costante: le parlanti femminili producono consonanti più aspirate rispetto agli uomini, in modo statisticamente rilevante all'interno di tutti i range di età considerati (R1: $t(158)=1,97$, $p<.01$; R2: $t(89)= 1,99$, $p<.001$; R3: $t(93)= 1,99$, $p<.001$). Tale dato è rafforzato da quelli inerenti alle correlazioni fra la variabile sesso e le durate di VOT ottenute, più lunghe se prodotte da donne (p<.001).

3.2 Area 4

L'Area 4 presenta una durata media globale del *Voice Onset Time* di 44,10 ms (d.s. 12,70 ms). Come per l'Area 3, anche in Area 4 la variabilità legata al punto di articolazione delle consonanti indagate riflette i trend letterari: in media, valori maggiori di VOT si riscontrano per le occlusive velari (/k/= 53 ms, d.s. 24,15) seguite dalle occlusive dentali (/t/= 45,23 ms, d.s. 18,83) e da quelle bilabiali (/p/= 40,46 ms, d.s. 17,11). Tutte le fasce d'età esaminate mantengono tale tendenza generale. Per quanto riguarda, invece, le differenze di VOT legate alla qualità delle vocali adiacenti, in tutti i range d'età esaminati si riscontrano valori di VOT maggiori, di fronte a vocali alte (M=49,21 ms, d.s. 23,3), minore rispetto alle vocali basse (M=37,98 ms, d.s. 17,05). In merito alla variabilità della durata del *Voice Onset Time* determinata dal contesto accentuale, contrariamente a quanto accade in Area 3, per l'Area 4 si attestano valori di VOT maggiori in contesto accentuale pretonico sia all'interno di parossitoni (M=47,81 ms, d.s. 23,61 vs. postoniche con medie uguali a 43,28 ms, ds. 20,17) sia all'interno di proparossitoni (M=42,61 ms, d.s. 21,1) vs. postoniche con medie uguali a 41,9 ms, d.s. 23,06) (cfr. Grafico 2).

Grafico 2 - *Valori medi di VOT in Area 4, differenziati per contesto fonetico*



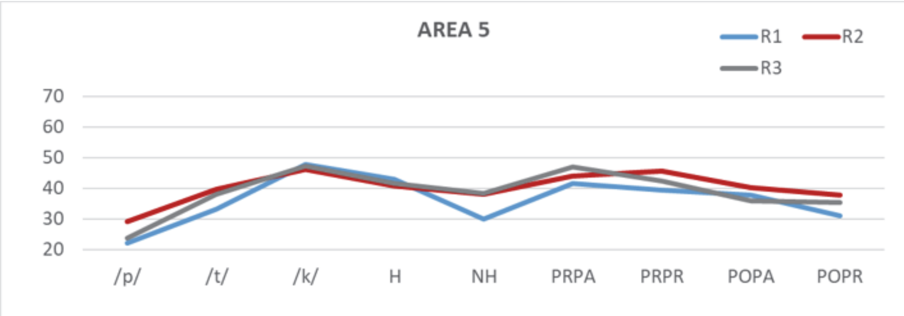
H	Alte	POPA	Postonica in Parossitono
NH	Non Alte	POPR	Postonica in Proparossitono
PRPA	Pretonica in Parossitono	PRPR	Pretonica in Proparossitono

Dal confronto fra le medie globali ottenute per tutti i range d'età, emergono valori significativamente maggiori nei parlanti di R3 rispetto agli informanti più giovani (R3 > R2/R1 per p<.001). Inoltre, per quanto riguarda la variabile *sesso*, sia all'interno di R2 (t(122)=1,98, p<.01) che di R3 (t(124)= 1,98, p<.01) le donne producono consonanti significativamente più aspirate rispetto agli uomini. I risultati vengono confermati dalle correlazioni fra le durate di VOT e le variabili *range d'età* e *sesso*. Il VOT presenta valori maggiori quando le consonanti target sono prodotte dagli informanti più anziani (p<.001) e da donne (p<.001).

3.3 Area 5

La durata media del VOT riscontrato nella quinta area dialettale è pari a 39,76 ms (d.s. 15,97). In merito al luogo di articolazione delle consonanti in esame, l'occlusiva velare ha la durata maggiore (/k/= 48,60 ms, d.s. 16); seguono quella della dentale (/t/= 37,53 ms, d.s. 14,89) e quella della consonante bilabiale (/p/= 30,57 ms, d.s. 12,64). Questa tendenza è stabile in tutte le fasce di età (cfr. Grafico 3). Il rapporto tra durata del VOT e vocali adiacenti alte/non alte rispecchia i trend letterari in quanto, per tutti i range d'età, le vocali alte determinano un *Voice Onset Time* superiore (M=43,63 ms, d.s. 17,25) rispetto a quello delle meno alte (M=36,35 ms, d.s. 13,85). Infine, il VOT ha una durata maggiore nel contesto pretonico sia all'interno dei parossitoni (M=44,68 ms, d.s. 16,64 vs. postoniche con medie uguali a 38,05 ms, d.s. 16,08) che dei proparossitoni (M=41,10 ms, d.s. 12,70 vs. postoniche con medie uguali a 32,90 ms, d.s. 12,21).

Grafico 3 - *Valori medi di VOT in Area 5, differenziati per contesto fonetico*



H	Alte	POPA	Postonica in Parossitono
NH	Non Alte	POPR	Postonica in Proparossitono
PRPA	Pretonica in Parossitono	PRPR	Pretonica in Proparossitono

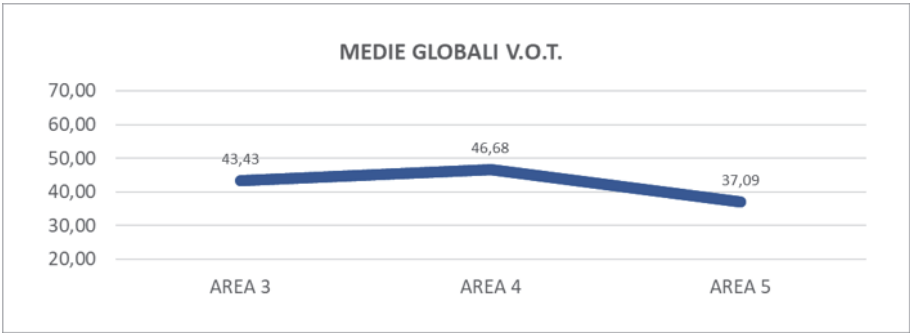
Dal confronto fra medie globali non risultano differenze significative ascrivibili al fattore *età* che possano confermare una variabilità inter-generazionale nei livelli di aspirazione prodotta dagli informanti. Neppure la variabile *sesso* costituisce un fattore discriminante significativo (M=35,42; d.s.=17,48 per gli uomini; M=36,44; d.s.=14,85 per le donne), eccetto nelle produzioni dei parlanti di età media, gruppo in cui gli uomini producono un VOT lievemente superiore rispetto alle donne ($t(94)=1,98$, $p<.05$).

Tali risultati vengono confermati dalle correlazioni tra le durate di VOT e, rispettivamente, le variabili *range d'età* e *sesso*, entrambe non significative.

4. *Discussioni e conclusioni*

Da un'osservazione globale dei dati presentati, le misure acustiche ottenute confermano la presenza di aspirazione delle occlusive sorde /p t k/ in contesto post-nasale, all'interno di tutte le aree dialettologiche investigate (cfr. Grafico 4):

Grafico 4 - *Medie globali di VOT per le Aree 3, 4 e 5*



Prescindendo dall'alta variabilità interna a ciascun gruppo analizzato, le durate medie di VOT rientrano nel range di riferimento di 30-100 ms. L'unica eccezione si riscontra in Area 5 per l'occlusiva bilabiale /p/, la quale mantiene valori medi inferiori ai 30 ms, dunque non presenta aspirazione, in tutti i range di età analizzati (cfr. Tabella 3). Sulla base della partizione considerata, l'Area 4 mostra i maggiori livelli di aspirazione in assoluto, seguita dall'Area 3 e dall'Area 5, quest'ultima con valori significativamente inferiori rispetto alle altre aree. I dati emersi confermano i trend generali sulla variabilità dei valori di VOT, in base ai luoghi di articolazione delle consonanti occlusive e all'altezza delle vocali a queste adiacenti. Tuttavia, in Area 4 e 5 le medie esaminate, in relazione alle posizioni accentuali di occorrenza, risultano essere in controtendenza rispetto a quelle di Area 3 e, di conseguenza, ai dati presenti in letteratura[5].

[5] A tal proposito, si ritiene che le differenze legate al contesto accentuale e al luogo di articolazione siano da imputare alla consistente variabilità intra-parlatore, come dimostra la fluttuazione dei valori di deviazioni standard presentati. In generale, un incremento della numerosità del campione potrà contribuire a rafforzare o smentire i dati fin qui ottenuti.

Tabella 3 - *Medie e deviazioni standard dei valori di VOT relativi alle Aree 3, 4 e 5, differenziati per contesti di occorrenza e range di età*

		/p/	/t/	/k/	A	NA	PRPA	PRPR	POPA	POPR
AREA 3	**R1**	34,8 (18)	40 (19)	47,2 (19)	53,5 (21)	35,3 (14)	40,6 (20)	34,3 (12)	43,9 (20)	37,5 (17)
	R2	33,8 (20)	39,8 (21)	55,1 (25)	48,2 (24)	38,5 (21)	44 (15)	30,6 (14)	46,0 (17)	39,3 (17)
	R3	47,2 (24)	39,0 (30)	58,1 (29)	55,5 (29)	38,8 (28)	48,2 (20)	27,7 (15)	53,6 (20)	32,0 (14)
AREA 4	**R1**	42,2 (11)	43,6 (12)	46,0 (12)	44,4 (12)	43,6 (12)	45,9 (13)	41,1 (9)	42,8 (12)	45,7 (12)
	R2	34,0 (18)	42,9 (21)	47,3 (21)	49,1 (23)	38,0 (17)	47,8 (24)	42,6 (21)	43,3 (20)	41,9 (23)
	R3	42,2 (20)	51 (21)	61,4 (29)	49,2 (23)	52 (24)	60,7 (27)	61 (22)	46,2 (19)	46,8 (21)
AREA 5	**R1**	22,2 (12)	33,2 (16)	47,9 (14)	43,1 (18)	30,4 (12)	41,6 (16)	39,4 (11)	37,9 (18)	31 (11)
	R2	29,2 (12)	39,7 (13)	46,1 (15)	40,7 (14)	38,2 (14)	44 (11)	45,8 (1,5)	40,4 (15)	37,9 (10)
	R3	23,7 (8)	38 (14)	47,2 (15)	41,6 (15)	38,5 (15)	47,1 (19)	42,5 (15)	36,0 (14)	35,5 (14)

H	Alte	POPA	Postonica in Parossitono
NH	Non Alte	POPR	Postonica in Proparossitono
PRPA	Pretonica in Parossitono	PRPR	Pretonica in Proparossitono

Sulla base dei dati ottenuti, il fattore età non sembra avere un effetto costante sulla variabilità di VOT nelle Aree esaminate, dal momento che non sempre le differenze intergenerazionali sono statisticamente significative[6]. Il fattore *sesso* appare, al contrario, un buon indicatore di variabilità (relativamente ai dati qui ottenuti), dal momento che, nella quasi totalità dei casi riscontrati, i valori associati a parlanti femminili superano in durata, in modo rilevante, quelli associati agli informanti maschili. Tale dato, se da un lato sembra contraddire alcuni degli studi legati alla variabilità del VOT fra sessi, per i quali in varie lingue del mondo si attesta una maggiore durata dei valori di aspirazione negli uomini rispetto alle donne (si confronti Piccardi, 2017 per una rassegna in merito), d'altro canto suggerisce che, trattandosi in questo caso di parlato dialettale, è possibile che le donne siano più ferme preservatrici del-

[6] Tuttavia, tali dati vanno considerati come puramente indicativi di una tendenza limitata ai parlanti coinvolti nelle presenti analisi: tenendo conto, inoltre, dei possibili effetti di *age-grading*, legati alle produzioni dei parlanti più anziani (cfr. Pichler, Wagner & Hesson, 2018; Wagner, 2012), si ritiene necessario operare in futuro su dati associati alle differenze di *speech-rate* dei parlanti.

la propria tradizione linguistica; naturalmente, anche questo dato merita di essere specificatamente esplorato.

Infine, i risultati raggiunti in questa ricerca, seppure relativi allo specifico contesto post-nasale, confermano e rafforzano quanto riscontrato percettivamente nelle precedenti ricerche su parlanti delle stesse zone geografiche e varietà dialettali, contribuendo a colmare le lacune esistenti sul piano acustico in merito alle analisi sull'aspirazione nei dialetti calabresi.

Riferimenti bibliografici

Abramson, A.S., Whalen, D. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. In *Journal of Phonetics*, 63, 75-86.

Boersma, P., Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.37. http://www.praat.org/Accessed 11.05.18.

Canepari, L. (1986). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Cho, T., Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. In *Journal of Phonetics*, 27, 207-229.

De Blasi, N. (2014). *Geografia e storia dell'italiano regionale*. Bologna: Il Mulino.

Falcone, G. (1976). Calabria. In Cortelazzo, M. (Ed.), *Profilo dei dialetti italiani*. Pisa: Pacini.

Fanciullo, F., Librandi, R. (2002). La Calabria. In Cortelazzo, M., Marcato, C., De Blasi, N. & Clivio, G. (Eds.), *I dialetti italiani. Storia, struttura, uso*. Torino: Utet, 793-833.

Frontera, M. (2018). Aspirated voiceless stops in elderly speakers from Calabria: a pilot study. In Botinis, A. (Ed.), *ExLing 2018 Proceedings of the 9th Tutorial and Research Workshop on Experimental Linguistics*, Paris, 28-30 August 2018, 33-36.

Frontera, M. (2019). Spanish voiceless stop production by Italian-Argentinian bilinguals from Calabria. Poster presentato al convegno *Bilingualism in the Hispanic and Lusophone world (BHL)*. Leiden University, 9-11 January 2019.

Harrington, J. (2013). Acoustic Phonetics. In Hardcastle, W.J., Laver, J. & Gibbon, F.E. (Eds.), *The Handbook of Phonetic Sciences*, (2nd ed.). Chichester: Blackwell Publishing, 81-129.

Heselwood, B., McChrystal, L. (1999). The effect of age-group and place of L1 acquisition on the realisation of Panjabi stop consonants in Bradford: an acoustic sociophonetic study. In *Leeds Working Papers in Linguistics & Phonetics*, 7, 49-68.

Kang, J., Whalen, D.H. (2017). *Get_vot*. In https://github.com/HaskinsLabs/get_vot. Accessed 12.04.19.

Lisker, L., Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. In *Word*, 20, 3, 384-422.

Mele, B. (2009). *Fonetica e fonologia del dialetto di San Giovanni in Fiore.* Tubingen-Basel: Francke Verlag.

Mele, B., Schmid, S., (2009). Le occlusive palatali del dialetto di San Giovanni in Fiore (CS). In Romito, L., Galatà, V. & Lio, R. (Eds.), *La fonetica sperimentale. Metodo e applicazioni*. Torriana: EDK editore, 349-371.

Morris, R.J., McCrea, C.R. & Herring, K.D. (2008). Voice onset time differences between adult males and females: isolated syllables. In *Journal of Phonetics*, 36, 308-317.

Nagy, N., Nodari, R.& Celata C., (2018). A variationist analysis of Heritage Italian VOT: Phonetic but not phonological fidelity. In *NWAV* 47, NY, 18-21 October 2018.

Nance, C., Stuart-Smith, J. (2013). Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. In *Journal of the International Phonetic Association*, 43(2), 129-152.

Nearey, T., Rochet, B.L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. In *Journal of the International Phonetic Association*, 24(1), 1-18.

Nodari, R. (2015). Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*, Studi AISV 1. Milano: Officinaventuno, 139-153.

Nodari, R., (2017). Indexicality and aspiration in Calabrian Italian: a sociophonetic approach. Poster presentato al *XIII Convegno AISV* (Associazione Italiana Scienze della Voce), Pisa, 25-27 gennaio 2017.

Nodari, R, Celata, C. & Nagy, N. (2019). Socio-indexical phonetic features in the heritage language context: Voiceless stop aspiration in the Calabrian community in Toronto. In *Journal of Phonetics*, 73, 91-112.

Piccardi, D. (2017). Sociophonetic factors of speakers' sex differences in Voice Onset Time: A Florentine case study. In Bertini, C., Celata, C., Lenoci, G., Meluzzi, C. & Ricci, I. (Eds.), *Fattori sociali e biologici nella variazione fonetica*, Studi AISV 3. Milano: Officinaventuno, 83-106.

Piccardi, D. (2018). L'aspirazione nei dialetti italiani meridionali e greci moderni. Tracce di una strategia di preservazione?. In *Lingua e stile*, 53(1), 27-55.

Pichler, H., Wagner, S.E., Hesson, A. (2018). Old-age language variation and change: Confronting variationist ageism. In *Language and Linguistics Compass*, 12(6), e12281.

Rochet, B.L., Yanmei, F. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. In *Canadian Acoustics*, 19(4), 105-106.

Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti.* Torino: Einaudi.

Romito, L., Ciardullo, M.A. & Tarasi A., (2015). Analisi acustica delle occlusive sorde aspirate del dialetto di San Giovanni in Fiore (CS). In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *ll farsi e il disfarsi del linguaggio. L'emergere, il mutamento e la patologia della struttura sonora del linguaggio*, Studi AISV 1. Milano: Officinaventuno, 169-186.

Scobbie, J. (2006). Flexibility in the face of incompatible English VOT systems. In Goldstein, L., Whalen L., H. & Best, C. (Eds.), *Laboratory Phonology VIII: varieties of phonological competence*. Berlin: Mouton de Gruyter, 367-392.

Sorianello, P. (1996). Indici fonetici delle occlusive sorde nel cosentino. In *Rivista Italiana di Dialettologia*, 20, 123-159.

Stevens, M., Hajek, J. (2010). Post-aspiration in standard Italian: some first cross-regional acoustic evidence. In *Proceedings of Interspeech*, Makuhari, Japan, 1557-1560.

Telmon, T. (1993). Varietà regionali. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo, la variazione e gli usi*. Roma-Bari: Laterza, 93-149.

Trumper, J., Ortale, R. (1978). Analisi preliminare del sistema consonantico del dialetto di Cosenza. In *Lingua e contesto*, 5, 3- 82.

Trumper, J., Maddalon, M. (1988). Converging Divergence and Diverging Convergence: The Dialect-Language Conflict and Contrasting Evolutionary Trends in Modern Italy. In Auer, P., di Luzio, A. (Eds.), *Variation and Convergence: Studies in Social Dialectology*. Berlin-NewYork: Mouton-de Gruyter, 217-259.

Trumper, J. (2016). *Geostoria linguistica della Calabria.* Roma: Aracne.

Yu, V., Nil, L.D., Pang, E. (2015). Effects of Age, Sex and Syllable Number on Voice Onset Time: Evidence from Children's Voiceless Aspirated Stops. In *Language and Speech*, 58(2), 152-167.

Wagner, S.E. (2012). Age Grading in Sociolinguistic Theory. In *Language and Linguistics Compass*, 6(6), 371-382.

DOI: 10.17469/O2106AISV000018

RICCARDO ORRICO, RENATA SAVY, MARIAPAOLA D'IMPERIO

Salerno Italian: Intonational phonology and dimensions of variation

Despite the large number of studies on the intonation of Italian varieties, research on Salerno Italian has been rather scarce and a detailed account of the main phonological patterns is still missing. The intonation analysis we report in this paper is framed within the Autosegmental-Metrical theory and the types of utterances we analyzed are taken from a corpus of both read and semi-spontaneous speech. The study adds to the literature on Italian intonation not only in dealing with an under-described variety, but also by going beyond the common consideration of tunes as stable expressions of pragmatic meanings within a linguistic community. Considerable emphasis is here placed on variety-internal variation with data showing that individual variability in the use of intonation cannot be simply dismissed, but needs to be properly addressed.

Key words: Intonation, Salerno Italian, Individual Variability, AM theory.

1. *Introduction*

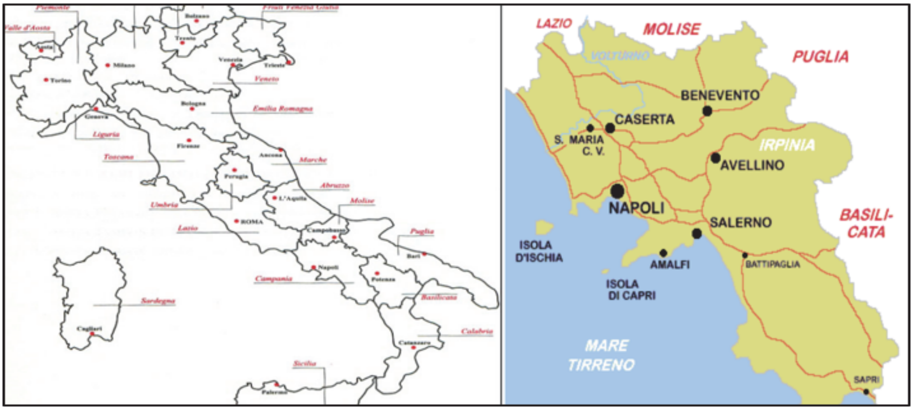
In the last few decades, research on intonation in Italian varieties has received an increasing amount of interest and some of the varieties spoken in the peninsula have been extensively investigated, especially within the Autosegmental-Metrical (AM) theory of intonation (Pierrehumbert, 1980; Ladd, 2008).

Nevertheless, the existing studies have been limited to a small number of varieties, while very little work is, to date, available for others. Much research has been conducted, for example, on Neapolitan Italian (among others, D'Imperio, 2000; 2002), which deeply investigates the interface between the phonology and phonetics of intonation and the role of intonational events in the production and perception of speech acts. The variety spoken in Bari has also been extensively investigated (Grice, Savino, 1997; Savino, Grice, 2011), focusing on the intonation realization of different types of yes-no questions (information-seeking, confirmation-seeking, counter-expectational). In addition, Palermo Italian has been under close examination, first by Grice (1995) and Grice, D'Imperio, Savino & Avesani (2005), who provide a description of the basic tunes used for statements and questions and, more recently, by Gili Fivela, Iraci (2017), in which a detailed description of the realization of a wider range of pragmatic conditions is reported.

Furthermore, a large number of studies have been published also for the varieties spoken in Tuscany (among others, Marotta, Sorianello, 2001; Avesani, Vayra, 2003; Gili Fivela, 2008). Among these, of particular interest is the investigation of Pisa Italian reported in Gili Fivela (2008), in which the author initiates the discussion around variety-internal variation and provides some important insights about the relationship between phonology and phonetics of intonation and the expression of pragmatic meanings.

The issue of variation across and within regional varieties of a language is of paramount importance when dealing with Italian and, in no way, the intonational systems of the few varieties investigated in the past should be taken as representative samples of the variation within this language. Italy is characterized by a complex linguistic system, in which vernaculars and regional-accented Italians coexist. The analysis of segmental phonology has allowed a dialectological classification that puts Italian vernaculars into a spatial continuum, characterized by a chain of inter-comprehension of adjacent vernaculars, moving from the extreme south to the north of the peninsula. On the other hand, it has allowed the definition of three macro dialectological areas, northern, central, and southern, with the boundaries being, on one side, the La Spezia-Rimini line, separating the northern from the central vernaculars, and on the other, the Rome-Ancona line, between central and southern vernaculars (see Loporcaro, 2009 for a detailed account).

Figure 1 - *Map of Italy (left) and a detail of Campania, the region where Salerno is located (right)*



However, the case of Italian intonation is far more complex than the segmental phonology facts. As pointed out by intonational studies adopting an inter-variety perspective (Grice et al., 2005; Savino, 2012; Crocco, 2013; Gili Fivela, Avesani, Barone, Bocci, Crocco, D'Imperio, Giordano, Marotta, Savino, & Sorianello, 2015), the picture of Italian varieties is much more fragmented when it comes to intonation. In addition, comparative studies are scarce, therefore it is not possible to draw a clear and comprehensive map of the intonational systems of Italian.

1.1 Comparative studies in Italian intonation

Within the AM approach to Italian intonation, the first attempt to specifically deal with variation across different varieties is Grice et al. (2005). The study analyzes a wide range of intonational phenomena, such as distribution and combination of pitch accents and boundary tones, the way they are used to express different speech acts, levels of phrasing, downstep, and tonal truncation for four different Italian varieties, *i.e.* Naples, Florence, Bari, and Palermo. The issue of variation is addressed with the goal of trying to find commonalities across the varieties and to propose a common labelling system.

The study does in fact show the existence of commonalities in the varieties analyzed. Statement intonation, for example, appears to be extremely stable across varieties, especially for broad focus utterances, in which a falling tune, analyzed as a H+L* L-L%, is generally employed. Additionally, the way those varieties signal narrow focus in statements is also similar, despite some differences in the way different labelers analyze the phenomenon. Specifically, the study reports the use of a L+H* pitch accent in Naples, a H*+L in Bari and Palermo, and a H* in Florence, despite a very similar rise-fall shape. In fact, looking at the phonetic characteristics of the three accents, it is possible to notice that they are very similar and that the different analyses are often the results of variety-specific choices. For example, the rising analysis for Neapolitan comes from D'Imperio (1999), who provides evidence that it is the rising portion of the configuration the one that characterizes the accent, after looking at its realization in multi-words focused constituents in which the sequence of accents is realized with an f0 medial fall. The first evidence derives from the fact that the f0 target scaling of the medial low tone stays the same when increasing the number of words within the constituent, which proves that it is an actual L target. Moreover, both slope and timing of such L target, with respect to the following peak, stay the same across conditions (while the slope of its interpolation with the preceding peak becomes shallower when number of words increases within the focused constituent), proving that the L target is actually the leading tone of a rising L+H* accent. Grice et al. (2005) argue that there are no reasons to believe that the same would not apply to the other varieties and that the same kind of inspection would probably yield similar arguments in favor of labeling the accent as a L+H*. However, experimental evidence for Bari Italian, in which L target timing is stable relative to the following H peak regardless of the number of syllables following the pitch accent, points towards a justified H*+L falling analysis (Gili Fivela, Savino, 2003).

Another common feature reported in Grice et al. (2015) is the way southern varieties signal questionhood. Specifically, while Florence Italian (central variety) makes use of both pitch accent and boundary tone as a cue for questionhood (H* L-H% for questions vs. H+L* L-L% for statements), southern varieties only use nuclear pitch accent type[1]. The strength of this claim, however, has been challenged in Savino (2012), arguing that a rising accent and a terminal fall is a typical way to signal a yes-no question throughout Italy.

Additionally, we can notice from Grice et al. (2005) that in the three southern varieties analyzed, the accent used for questions is a later peak variant, as opposed to

[1] Though, as then shown by Petrone, D'Imperio (2008), the shape of the interpolation line between the prenuclear and the nuclear accent in late narrow focus utterances also has a role in discriminating the two speech modalities in Neapolitan Italian.

the accent used in statements in the same variety. Again, the differences in the way the accents are labelled across varieties, *i.e.* L*+H in Naples and Palermo, and L+H* in Bari, derive from variety-specific choices, often necessary to signal phonological opposition with early peak accents in the same variety and hence not motivated by (phonetic) target alignment alone.

A second thorough comparative analysis of the intonation in Italian varieties is Gili Fivela et al. (2015). This study, conducted as a part of a much bigger research project including the investigation of intonation in several Romance languages, collects data from thirteen varieties spoken in Italy with a fairly wide range of pragmatic functions. Moreover, it is based on new spoken data, collected through the same methodology in all the varieties considered, hence allowing for a much simpler and coherent inter-variety comparison.

Some of the claims put forward by Grice et al. (2005) find confirmation when analyzing data from a larger number of varieties and pragmatic situations. For example, the realization of broad focus statements is confirmed to be stable across varieties, with a falling pitch accent and a low boundary (H+L* L-L%). Similarly, the realization of the pitch accent signaling contrastive focus mirrors previous analyses, in that it appears to be realized in a very similar way throughout Italy, with a rising-falling f0 movement, whose peak is aligned with the first half of the accented vowel. This is true despite the fact that differences in the interpolation of the accentual targets, together with variety-specific phonological oppositions, might call for different labels (L+H*, H*+L, or, in some cases, H*).

The picture, however, becomes much more complex for yes-no questions. The analyses presented in Gili Fivela et al. (2015) reveal a strong variation across the varieties and, crucially, no evidence points to the existence of clusters of varieties on the basis of geographical or dialectological areas. This lack of homogeneity appears to be also true, though not as strong, for wh-questions. For example, counter-expectational wh-questions appear to be very consistent across varieties.

In addition, continuing the work on Pisa Italian by Gili Fivela (2008), Gili Fivela et al. (2015) clearly acknowledges the role of intra-variety variability. Variation is not only detectable across varieties, but it is argued to be visible also within the same variety. Quite a wide range of possible combinations of pitch accents and boundary tones are available to speakers, without any clearly predictable pattern. Together with the observation that geographically distant varieties can have the same tonal realization for the same function, such a high level of variability is speculated to be dependent on inter-variety interference, triggered sometimes by a number of political and social factors, such as schooling, urbanization, internal migration, and perceived prestige of some varieties relative to others.

1.2 Objectives of the study

The objectives of this contribution are to **i)** provide a description of an under-described variety, namely Salerno Italian and **ii)** address the issue of variability in questions within the variety reported in previous studies, by providing data helping to unveil the possible sources of such variation.

2. *Intonation in Salerno Italian*

2.1 Previous research on Salerno Italian

Salerno is a southern variety of Italian which has not received much attention in previous studies on intonation. To the best of our knowledge, some preliminary considerations about the intonation in SI have only been included in few, recent studies, which adopt a contrastive perspective either on several Italian varieties, such as Gili Fivela et al. (2015), or a Second Language Learning perspective (Savy, Luque Moya, 2015; Orrico, Cataldo, Savy & Barone, 2016). Hence, an investigation centered around the intonational system of this variety is still lacking. As for Gili Fivela et al. (2015), in which several varieties and a wide range of pragmatic situations are considered, details about SI are not provided.

An interesting fact on SI, springing from information collected on previous studies, is the difference of some of its intonational phonology traits from the systems of adjacent varieties, specifically from Neapolitan Italian. Salerno and Naples, which are geographically close (about 60 km) are traditionally considered also very close from a linguistic point of view. If we consider segmental phonology traits, the vernaculars spoken in these two cities are both included within two bundles of isoglosses defining a geographical area (central Campania), whose language is hinged on Neapolitan, and are separated from other southern vernaculars spoken in Campania, which, on the other hand, share traits with the vernaculars spoken in Basilicata (Loporcaro, 2009: 142-144).

Nevertheless, the same cannot be argued for intonation. The difference in the two systems appears to be particularly detectable in yes-no questions: Neapolitan follows the "southern trend" described in Grice et al. (2005) in which questionhood is characterized as a rise-fall, with a low terminal, and discrimination relative to statements is mainly conveyed by pitch accent choice. On the contrary, SI, despite also allowing for rise-falls in questions, appear to allow a final rise, which appears in the prototypical question tune in SI (Savy, Luque Moya, 2015). This trend is observed in different varieties throughout Italy. However, while for some Italian varieties this double choice of boundary for question is often argued to be due to speech style (see Savino, 2012), this does not appear to be the case for SI.

2.2 Methods

The following sections present a description of the main intonational patterns used for statements and questions in Salerno Italian (SI).

We used the same corpus employed in Gili Fivela et al. (2015). This is a modified version of the corpus used for the *Interactive Atlas of Romance Languages* (IARI, Prieto, Borràs-Comes & Roseano, 2010), consisting of speech material collected via a Discourse Completion Task (DCT) and a Reading Task (RT). In the DCT, speakers had to spontaneously react to a written text describing a situation, while in the RT they had to read out loud the target sentence after silently reading a description of the contextual situation. The same contexts were used in both tasks. Gili Fivela et al. (2015) adopted specific criteria of age (20-35 years old) and education (high school or university degree) for speaker selection. For Salerno, speech material was collected from 4 speakers (2 males and 2 females) and for 57 contexts[2]. Two repetitions for both DCT and RT were collected from each speaker.

In the present study we report results from a subset of the 57 situations collected for Salerno. Specifically, we will focus on statements and questions: **i)** broad (48 items) and narrow (32 items) focus condition for statements; **ii)** information-seeking (79 items), confirmation-seeking (45 items), counter-expectational (13 items), and echo (30 items) conditions for yes-no questions.

The intonational analysis was conducted within the Autosegmental-Metrical approach and using a ToBI-like annotation system, as described in previous studies about intonation in Italian varieties (see section 1.1 above). The analysis is primarily concerned with the occurrence of tonal events (pitch accents and edge tones) and, in particular, with the realization of nuclear configurations.

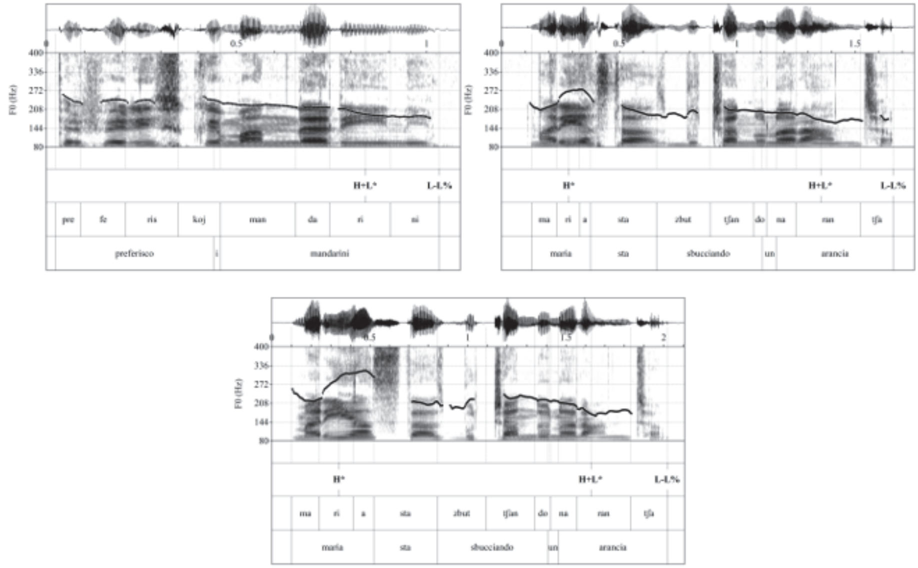
2.3 Statements

Consistent with previous investigations, both broad and narrow focus statements are tonally realized in SI similar to the other varieties of Italian and, additionally, without individual variation.

The tune used for broad focus statements is analyzed as a H+L* L-L%. It is realized as a falling tune, with a falling nuclear pitch accent and a low boundary tone, as shown in the three examples in Fig. 2.

[2] The pragmatic situations included different conditions for statements (broad focus, narrow focus with different positions of the focalized constituent, left/right dislocations, lists), different conditions for questions (information-seeking, confirmation-seeking, counter-expectational, and echo for yes-no questions, information-seeking, counter-expectational, and echo for wh-questions, alternative questions), imperatives, offers, exclamations, and vocatives. The complete list of contexts and target sentences used to collect the original corpus used in the IARI project can be found at http://prosodia.upf.edu/iari, and a more precise explanation of the Italian adaptation of the corpus can be found in Gili Fivela et al. (2015).

Figure 2 - *F0 curve for H+L* L-L% in the statements* Preferisco i mandaRIni[3] *'(I) prefer tangerines' (top left) and two renditions of* Maria sta sbucciando un'aRANcia *'Maria is peeling an orange' (top right and bottom) uttered as broad focus statements by the same female speaker*



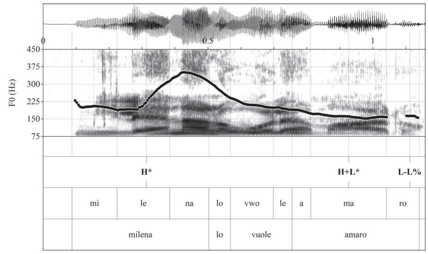
The H+L* pitch accent is phonetically realized as a relatively moderate fall on the syllable bearing the accent. As also shown for other varieties (Gili Fivela et al., 2015), despite always showing an L* target or a falling movement aligned with the accented vowel, the H leading tone can have different scaling or, sometimes, be included in a gradual fall.

Additionally, in case a prenuclear accent is present in a broad focus statement, it is analyzed as a H* pitch accent (top left and bottom panel in Fig. 2), whose phonetic realization involves a rising movement from a low pre-tonic region and a peak that can be realized either on the right edge of the accented syllable or even later, in the post-tonic region. Compare, for example, the prenuclear accents in Fig. 2 with the broad focus statement in Fig. 3, uttered by the same speaker, which more clearly shows the rising movement of prenuclear H* pitch accents. The reason behind its phonological analysis as a monotonal high accent instead of a bitonal sequence of LH tones is twofold. On the one hand, there is a variety internal motivation, stemming from the phonological opposition this accent has with the other rising accents found in the system of tones in SI (see following sections). On the other hand, we have not found a clear phonological opposition between these prenuclear accents. The same has been found for other Romance languages and other varieties of Italian by Prieto, D'Imperio & Gili Fivela (2005), where it is argued that the association between a tone and a segment should not be analyzed only by observing their timing, but also considering

[3] Throughout the article, underlined upper case letters indicate the placement of nuclear pitch accents, while underlined lower case letters indicate pre-nuclear or post-nuclear pitch accents.

how the native speaker perceives it (see also Arvaniti, Ladd & Mennen, 2000) and explain the differences in alignment in allophonic accents by referring to the notion of secondary association with the prosodic word rather than the syllable.

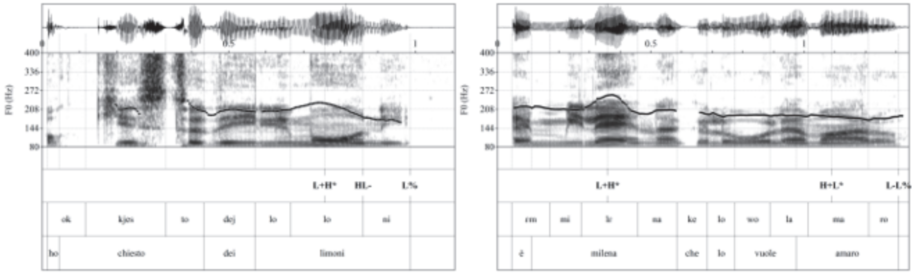
Figure 3 - *F0 curve in the statement* Mil<u>e</u>na lo vuole a<u>MA</u>ro *'Milena takes it (the coffee) black' uttered as a broad focus statement by a female speaker*



Nevertheless, the rendition in Fig. 3 might, impressionistically, be related to a non-exhaustive answer and the topic (Milena) realized as a 'Partial Topic' (Buring, 1997), as if the statement represented the answer to a question like 'How do *your friends* take their coffee?', rather than 'How does *Milena* take her coffee?'. The notion of Partial Topic and its tonal realization has been investigated in D'Imperio, Cangemi (2011) and Brunetti, D'Imperio & Cangemi (2010) for Neapolitan Italian. They find that the non-exhaustive answer presented also a mandatory phrase accent at the right edge of the topic (typically, !H-) and a compressed f0 trend following it, therefore the non-exhaustive reading might stem from those cues and not from a different pitch accent. The same could also be speculated for SI. However, such investigation goes beyond the scope of this paper and future work should be dedicated to answering research questions such as the allophonic status of the two prenuclear accents and the primary cues to the non-exhaustive reading.

As for narrow focus statements, they are realized with a rising nuclear pitch accent, which, all else being equal, is the only tonal difference between this and the broad focus condition. This tune is phonologically analyzed as a L+H* HL-L% sequence, with the pitch accent being phonetically realized as a rising movement starting in the pre-tonic region and a peak aligned with the first half of the syllable, as shown in Fig. 4. The difference in the realization of a L+H* pitch accent and the H* prenuclear pitch accent is that, in narrow focus, three targets are clearly identifiable in the rise-fall movement within the accent region: a L leading tone, typically aligned with the right edge of the pre-tonic syllable, a H* tone, aligned within the first half of the accented vowel, and a second L tone, which is typically reached around the offset of the accented syllable or the onset of the post-tonic syllable. The motivations for the rising analysis mirror the acoustic measurements provided for Neapolitan Italian (see section 1.1) for two reasons. The first one has impressionistically perceptual grounds, since the Neapolitan L+H* sounds the same as the narrow focus accent in SI. In addition, the same opposition with a L*+H rising accent has been found in SI (see the following section).

Figure 4 - *F0 curve for the utterances* Ho chiesto dei li<u>MO</u>ni *'(I) asked for lemons' (left) and* È Mi<u>LE</u>na che lo vuole a<u>ma</u>ro *'(it) is Milena that takes it (the coffee) black' (right) uttered as narrow focus statements by the same female speaker*



In case the focalized constituent is located early in the sentence, a post-nuclear accent is realized, which is analyzed as a H+L*, though it appears much more compressed than when it is realized as a nuclear accent (Fig. 4, right panel).

2.4 Yes-no questions

Results from Gili Fivela et al. (2015) show the general absence of an exact mapping between tunes and yes-no question sub-functions in all the Italian varieties investigated so far. Here, four different sub-functions were considered: information-seeking, confirmation-seeking, counter-expectational, and echo questions. The present section reports intonational data analysis regardless of the pragmatic function. See section 3 for a discussion.

Globally, four different patterns were found for yes-no questions, *i.e.*, two rise-fall and two rise-fall-rise tunes (see Fig. 5). Two different accents can be used as nuclear pitch accents in yes-no questions in SI. Note that both are rising pitch accents, and their phonological opposition is mainly dependent on the temporal alignment of the tonal targets within the accent with respect to the accented syllable. Specifically, we found an early peak pitch accent, labelled as L+H* (left panels in Fig. 5), which is the same accent used to signal narrow focus in statements. The second accent, labelled as L*+H (right panels in Fig. 5), has a rise-fall movement too, though the alignment of its targets is later than for L+H*. Specifically, the first low target is typically realized at the onset of the accented vowel, the peak is reached at the offset of the same vowel, and the second low target is realized in the post-tonic syllable. Both pitch accents can combine with either a low or a high boundary, though distributional data reveal a preference for choosing rising boundaries with a nuclear L+H* and falling ones when a L*+H is chosen.

Figure 5 - *F0 curves for the yes-no question* Sono le <u>NO</u>ve? *'Is it nine o' clock?' uttered with a L+H* HL-H% tune (top left), a L*+H HL-L% tune (top right), and a L+H* HL-L% tune (bottom left), and for the question* Vuoi le <u>MAN</u>dorle? *'Do (you) want the almonds?' uttered with a L*+H HL-H% tune (bottom right)*

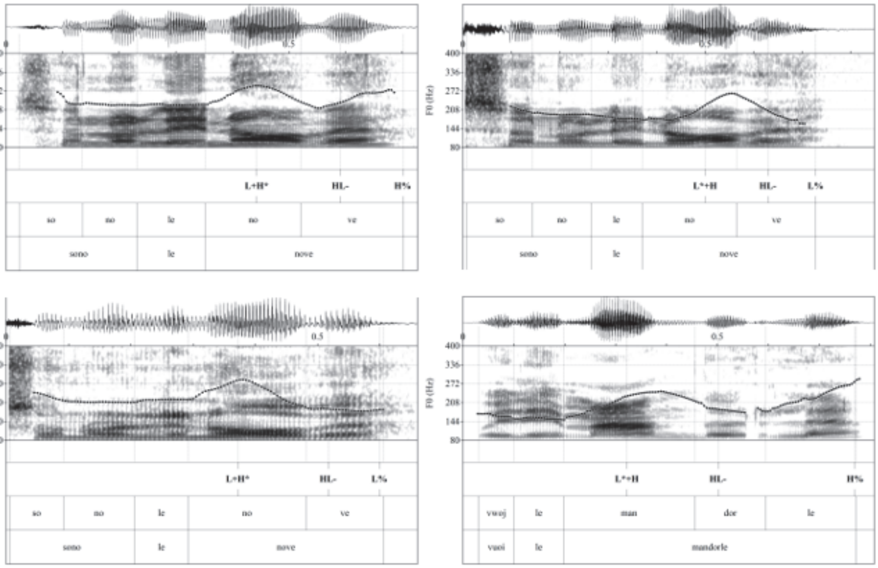
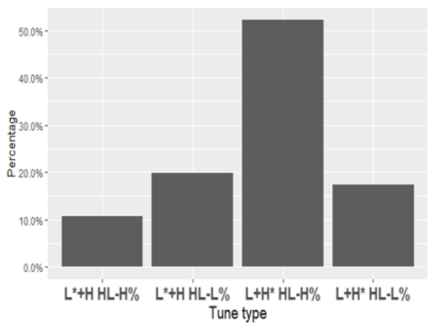


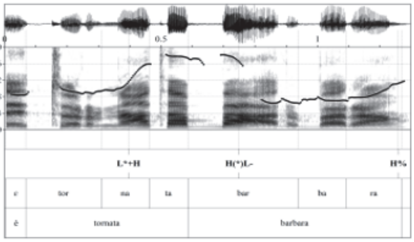
Fig. 6 shows percentages of occurrence of each of the four tunes in our corpus. L+H* HL-H% is the most frequent tune in our data set, which alone reaches more than 50% of occurrences. The same accent, conversely, combines with a low boundary in less than 20% of the cases. As for L*+H, it more often combines with a HL-L% boundary and only in 10% (lowest percentage of occurrences) of the cases with a HL-H%.

Figure 6 - *Percentages of occurrence of the four yes-no questions tunes within our corpus*



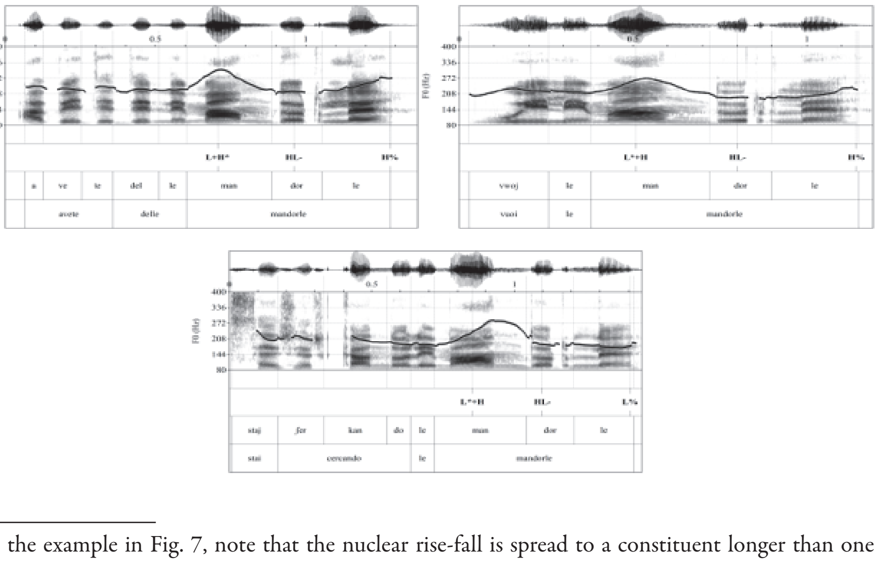
The low frequency of occurrence of L*+H HL-H% might depend on tonal crowding, which implies a relative difficulty to produce the tonal gestures associated with the accent and the tonal edge consecutively, over a short time span. Indeed, L*+H HL-H% is more likely to occur in association with proparoxytone words, where an additional syllable is available between the accent and the boundary, or in cases in which the nuclear pitch accent is not realized on the rightmost accentable syllable in the intonational phrase (see Fig. 7)[4].

Figure 7 - *Example of L*+H L-H% with early nuclear pitch accent in the yes-no question* È tor<u>NA</u>ta <u>Bar</u>bara? *'Is Barbara back?'*



Compare, for example, the three tunes in Fig. 8. The difference between the alignment of the two peaks in the two renditions of L*+H (bottom vs top-right) is pretty clear, with an earlier timing when in combination with an L-H% boundary. In addition, as a consequence of tonal repulsion, the alignment of the peak in L*+H overlaps with the alignment of L+H* (top-right vs top-left). In these cases, the phonological opposition between the two events might be compensated by cues other than alignment, such as the shape of the rise-fall movement, which is a common strategy for category disambiguation (see for example Niebuhr, D'Imperio, Gili Fivela & Cangemi, 2011). Additionally, when the two rising gestures associated with L*+H HL-H% occur very close in time, strategies of temporal readjustment of targets might be used.

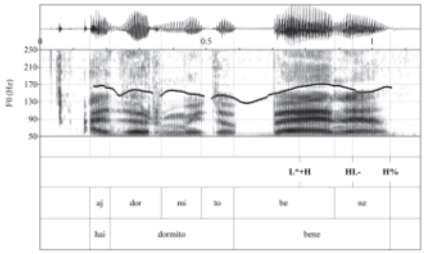
Figure 8 - *Alignment of tonal targets in L+H* (top left) and L*+H in combination with either a L-H% (top right) or L-L% (bottom) uttered by the same female speaker*



[4] In the example in Fig. 7, note that the nuclear rise-fall is spread to a constituent longer than one word and the HL- phrase accent is anchored to the last stressed vowel of the constituent. Following D'Imperio (2000), we label the phenomenon as H(*)L-.

In other cases, strategies of gestural overlap might occur, in which the fall leading to the L- phrase accent, overlapping with the H% rise, is not fully realized, as it can be seen in in Fig. 9.

Figure 9 - *Gestural overlap in L*+H L-H% in the question* Hai dormito <u>BE</u>ne? *'Did (you) sleep well?' uttered by a female speaker*



Here, in order to fully produce the two rising movements, the speaker does not re-adjust the timing of the H accentual target, but instead the curve stays at a relatively high level even during the L- phrase accent due to lack of time to produce both a fall and a rise.

2.5 Summary of statement and question tunes in SI

Table 1 shows the three pitch accents that can be used in nuclear position and how they combine with the two boundary tones to form nuclear tunes.

Table 1 - *Schematic representation of the nuclear tunes used in SI for statements and questions*

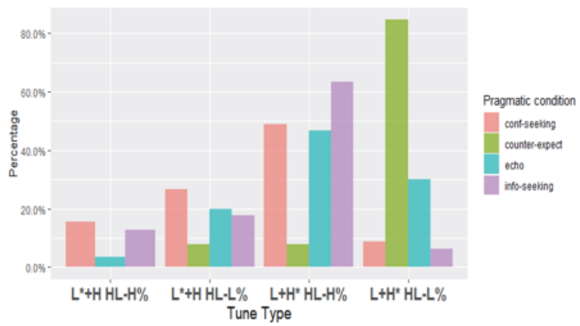
Tune	f0 realization		Utterance type
	Pitch Accent	Boundary tone	
H+L* L-L%			**broad focus statements**
	Pitch Accent	Boundary tone	
L+H* L-L%			**narrow focus statements/ yes-no questions**
	Pitch Accent	Boundary tone	
L+H* L-H%			**yes-no questions**
	Pitch Accent	Boundary tone	
L*+H L-L%			**yes-no questions**
	Pitch Accent	Boundary tone	
L*+H L-H%			**yes-no questions**

3. *Dimensions of variation in SI yes-no question tunes*

Section 2.4 above reports a description of the tunes used for yes-no questions in SI. However, so far no consideration has been made about: **i)** the different pragmatic sub-functions, *i.e.*, information-seeking, confirmation-seeking, counter-expectational, and echo questions; **ii)** the four different speakers; and **iii)** the two task conditions used to elicit the speech material, *i.e.* Discourse Completion and Reading Task. These three dimensions are potential sources of variation in the use of intonation that are worth investigating. This section provides an attempt to characterize the distribution of tunes by taking into account the role that such dimensions of variation might play (see Fig. 10-13).

Recent work on Italian intonation has shown that pragmatic conditions for yes-no questions as those considered here are not a good criterion for tune classification. As reported above, the issue of intra-variety variability has been raised by Gili Fivela (2008) for Pisa Italian and then confirmed for several Italian varieties by Gili Fivela et al. (2015), where they notice that while many varieties share similar tunes for yes-no questions, the same tune can be used for more than one sub-function within the varieties. Fig. 10 shows the distribution of the tunes in SI according to the pragmatic sub-function they were used for. The absence of a one-to-one mapping between tune and function seems to be confirmed also for SI. No specific tune expresses a specific function and, at the same time, no specific function is expressed by a specific tune. The only exception is represented by counter-expectational questions, which are expressed with a L+H* HL-L% tune in almost the 90% of the cases.

Figure 10 - *Distribution of tunes according to the pragmatic sub-function in SI*

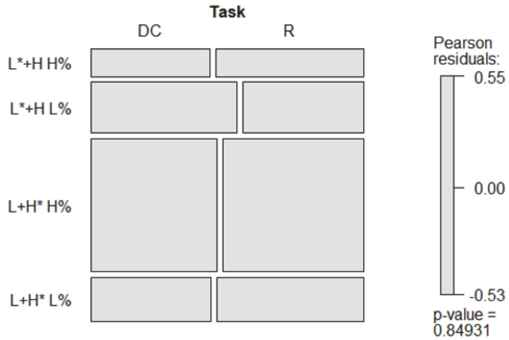


Note that, as for the other pragmatic conditions, different tunes can be used. This situation might be explained by unveiling the other potential dimensions of variability listed above *i.e.* speaker variability and task used for elicitation.

In past studies, it has been argued that speaking style plays an important role in Italian yes-no question intonation. Specifically, the use of H% is argued to be a consequence of a reading effect, since it is not present in spontaneously collected data (Grice, Savino, 1997; Savino, 2012). We tested if this situation applied to our data.

In Fig. 11 we show the distribution of tunes according to the two tasks (Discourse Completion and Reading Task) in a mosaic plot.

Figure 11 - *Mosaic plot showing the distribution of tunes according to task (DC = Discourse Completion; R = Reading)*



The mosaic plot above was created in R using the package *vcd*, which allows for testing the hypothesis of independence between variables using the residual-based method by Pearson (Zeileis, Meyer & Hornik, 2007). The size of the boxes within the plot indicates the frequency of occurrence of each combination of variables, while the residuals give information about how far is the actual occurrence from the one expected according to the null hypothesis (*i.e.* that the variable are independent). As a rule, a residual ranging between 2 and -2 leads to the confirmation of the null hypothesis. The fact that residuals in Fig. 11 are kept within this interval shows that tune and task do not covariate, therefore we are led to believe that the presence of H% is not an effect of 'reading style', but rather it reflects a specific pragmatic choice.

A completely different picture is yielded when we plot the occurrence of tunes for each of the four individual speakers.

Figure 12 - *Mosaic plot showing the distribution of tunes in the four individual speakers (AS, AV, SD, and VM)*

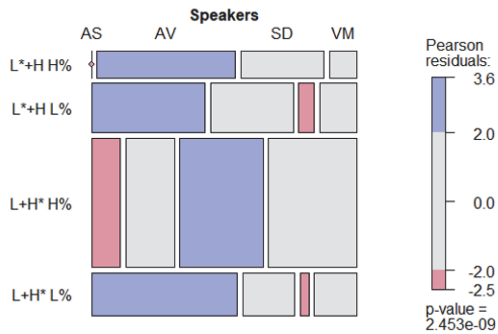
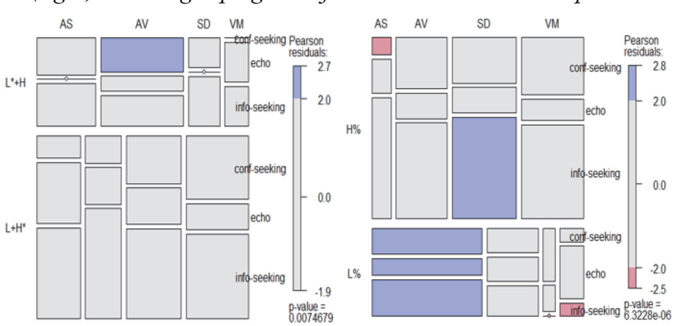


Fig. 12 shows that tune choice is rather dependent on speaker variability. The blue and pink boxes indicate that the occurrence of that specific combination of levels is, respectively, significantly higher and lower than predicted by the null hypothesis. Regardless of the pragmatic condition, speaker-specific differences are quite evident. Note that, for example, there are speakers that prefer to use rising- rather than falling-terminal tunes (SD) or vice versa (AS). AV, on the contrary, makes an overall greater use of L*+H tunes.

Additionally, it would be interesting to observe how this individual variability interacts with the different pragmatic conditions. Italian intonation research that directly addresses individual variability in the mapping between tunes and pragmatic functions is, to the best of our knowledge, absent. Variation at the individual level for Italian varieties has been investigated concerning the interface between phonology and phonetics (Niebuhr et al., 2011), but not concerning the intonation-meaning relationship, while evidence has been shown for other languages. For example, in German, Grice, Ritter, Niemann, & Roettger (2017) show that different speakers use different strategies to encode focus types in intonation. These studies support the claim that individual variability needs to be taken into account when dealing with intonation and intonational meaning. In Fig. 13 we plotted the relationship between individual variability and pragmatic conditions in questions. Note that we excluded counter-expectational questions from the plot. This choice was made for two reasons. Firstly, we have evidence from Fig. 10 that they are realized quite consistently across speakers (they are produced with a L+H* HL-L% in almost 90% of the cases). Secondly, there were very few counter-expectational questions available in the corpus, therefore their inclusion would have created a loss of power to the calculations. Additionally, we plotted boundaries and pitch accents separately. Fig. 13 shows that some of the inter-individual differences reported in Fig. 12 appear to be unrelated to pragmatic condition. For instance, AS makes a greater use of L% than H% in all the conditions. However, we can note that other speaker-specific strategies are also dependent on the pragmatic condition. For example, the frequent use of L*+H by the speaker AV appears to be restricted to confirmation-seeking questions. Additionally, the greater use of H% by SD is only significant in information-seeking questions.

Figure 13 - *Mosaic plot showing the distribution of pitch accents (left) and boundary tones (right) according to pragmatic functions and individual speakers*



In general, abstracting away from inter-individual variation, it can be argued that information-seeking questions are typically expressed with either a L+H* accent or a H% boundary. Counter-expectational questions appear to be expressed with the same pitch accent, though always in combination with a low boundary. As for confirmation-seeking and echo questions, no specific tune seems to be used to express either one of them.

4. *Discussion*

In line with recent research, this study puts SI within the frame of Italian intonational phonology. Despite commonalities within the intonation systems of Italian varieties, most of which are relative to the intonation realization of statements, a great deal of variability is registered, both between SI and other varieties as well as across different speakers.

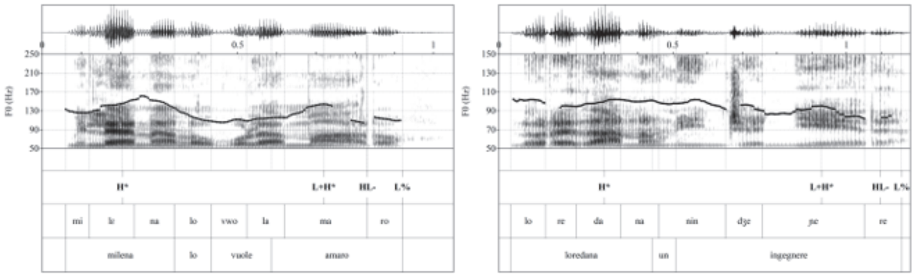
Gili Fivela et al. (2015) found inter-variety variation for Italian and the absence of specific groupings of intonational traits overlapping with dialectological areas defined on the basis of segmental phonology. In addition, in section 2.1 above we also noted that information about SI intonation available in past studies point towards a systemic difference between the varieties of Italian spoken in Salerno and Naples, despite literature on segmental phonology has shown that they share several features. Nevertheless, looking at the picture drawn here for the intonational system of SI, a number of similarities with Neapolitan Italian can be found. Specifically, the two varieties appear to share the pitch accent inventory: they have, for example, the same opposition between two rising accents, L+H* and L*+H, both of which are found to occur only in nuclear position in both varieties. Also, the same pitch accent types (H* and H+L*) seem to occur in pre-nuclear and post-nuclear positions in the two varieties. It therefore appears to be the case that the two varieties have a common basic phonological inventory of intonation primitives. Nevertheless, differences at the level of tune well-formedness and tune-function mapping can be found.

A particularly striking result is the presence of high levels of variability within the data analyzed. We showed that such variability does not depend on the type of task (as shown in Fig. 11) and although a general strategy for expressing information-seeking and counter-expectational questions was detected beyond variability, a uniform pattern is hard to find (Fig. 10; 13).

The absence of a one-to-one correspondence between tunes and functions that has been suggested for different languages. Several studies have shown that a number of factors beyond phonological intonation categories play a role in defining the meaning of a tune. Among these, evidence has been provided for the effect of pragmatic context (among others, Armstrong, Prieto, 2015) and phonetic details (Gili Fivela, 2012; Borras-Comes, Vanrell & Prieto, 2014; Mucke, Grice, 2014).

As for the role of phonetic detail, compare, for instance, the realizations of confirmation-seeking and counter-expectational questions produced by the same SI speaker in the two examples reported in Fig. 14.

Figure 14 - *F0 curves for the confirmation-seeking question* MiLEna lo vuole aMAro? *'Milena takes (the coffee) black' (left) and the counter-expectational question* LoreDAna un ingeGNEre? *'Loredana an engineer?' (right) uttered by the same male speaker*



The two questions are realized with the same pitch accents and boundary tones, however the two pitch accents are compressed in the counter-expectational question, with pitch accent peaks about 50 Hz lower than the confirmation-seeking question. The same can be said for counter-expectational wh-questions, which, as reported in Gili Fivela et al. (2015) are very often, and also very consistently across varieties, produced with the same phonological category as information-seeking wh-questions, but with a much greater excursion within the pitch accent.

Nevertheless, high levels of speaker-specific variation in the use of intonation has been detected (Fig. 12). Such variation appears to be, in some cases, independent from the use of intonation in different pragmatic condition (*e.g.* the use of L% by AS). This strong individual effect is an issue that cannot be readily addressed. One of the main difficulties that arise is to figure out the sources of such speaker-specific variability. Recent work by Kleinschmidt (2019) has tried, for example, to model segmental phonetic variability by looking at how it correlates with socio-indexical features of the speaker, such as age, gender, and dialect spoken, obtaining results confirming his hypothesis. The model also provides evidence of the fact that listeners exploit the covariance of phonetic and social traits to cope with ambiguities in the speech signal, arguing that listeners adapt to variability. However, a lot of unexplained variability at the individual level was still present in Kleinschmidt's (2019) data, suggesting that other factors also come into play.

In a related study (Orrico, Savy & D'Imperio, 2019), for example, we report that socio-indexical features of the speakers can have an effect at the perception level. Specifically, we found that continued exposure (for at least 12 months) to either other varieties of Italian or other languages results in a retuning of intonation categories and their link to specific meanings, which might also have an effect in production. These results are in line with what argued by Gili Fivela et al. (2015), claiming that the main reason for such fragmented picture of Italian intonation might be found in high level of contact across different varieties.

5. *Conclusion*

The aim of this contribution was to provide a better understanding of intonational phonology of Salerno Italian, one of the varieties that has not received much attention by the linguistic community in past research. In line with recent research on Italian intonation, the study was carried out with the assumption that variation, both across and within varieties, exists and needs to be dealt with by unveiling its possible sources. The analysis reported here shows that variability at the individual speaker level represents the strongest, unavoidable level and should be better accounted for in models in which socio-phonetic and cognitive features of speakers are considered. Only by taking into account these aspects of speech we can get a better understanding of the role of intonation in communication.

Bibliography

Arvaniti, A., Ladd, D.R. & Mennen, I. (2000). What is a starred tone? Evidence from Greek. In *Papers in laboratory phonology V: Acquisition and the lexicon*, 119-131.

Avesani, C., Vayra, M. (2003). Broad, narrow and contrastive focus in Florentine Italian. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Vol. 2, 1803-1806.

Borràs-Comes, J., Del Mar Vanrell, M. & Prieto, P. (2014). The role of pitch range in establishing intonational contrasts. In *Journal of the International Phonetic Association,* 44(1), 1-20.

Brunetti, L., D'Imperio, M. & Cangemi, F. (2010). On the prosodic marking of contrast in Romance sentence topic: evidence from Neapolitan Italian. In *International Conference on Speech Prosody*, 1-4.

Büring, D. (1997). *The Meaning of Topic and Focus – the 59th Street Bridge Accent.* London: Routledge.

Crocco, C. (2013). Is Italian Clitic Right Dislocation grammaticalised? A prosodic analysis of yes/no questions and statements. In *Lingua*, 133, 30-52.

D'Imperio, M. (1999). Tonal structure and pitch accent in Italian focus constituents. In *Proceedings of the 14th International Congress of Phonetic Sciences. University of San Francisco, Berkeley*, 101-106.

D'Imperio, M. (2000). The role of perception in defining tonal targets and their alignment. PhD dissertation, Ohio State University.

D'Imperio, M. (2002). Italian intonation: an overview and some questions. In *Probus*, 14(1), 37-69.

D'Imperio, M., Cangemi, F. (2011). Phrasing, register level downstep and partial topic constructions in Neapolitan Italian. In *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*, 10, 75-94.

Gili Fivela, B. (2008). *Intonation in production and perception: The case of Pisa Italian.* Alessandria: Edizioni dell'Orso.

Gili Fivela, B. (2012). Meanings, shades of meanings and prototypes of intonational categories. In Elordieta, G., Prieto, P. (Eds.), *Prosody and meaning* (Vol. 25). Berlin–New York: Walter de Gruyter, 197-237.

Gili Fivela, B., Iraci, M.M. (2017). Variation in intonation across Italy: The case of Palermo Italian. In Bertini, C., Celata, C., Lenoci, G., Meluzzi, C. & Ricci, I. (Eds.), *Fattori sociali e biologici nella variazione fonetica*. *Studi AISV 3*. Milano: Officinaventuno, 167-188.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. & Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In Frota, S., Prieto, P. (Eds.), *Intonation in Romance*. New York: Oxford University Press, 140-197.

Gili Fivela, B., Savino, M. (2003). Segments, syllables and tonal alignment: A study on two varieties of Italian. In *Proceedings of the XVth international congress of phonetic sciences*, 2933-2936.

Grice, M. (1995). *The intonation of interrogation in Palermo Italian.* Tubingen: Niemeyer.

Grice, M., D'Imperio, M., Savino, M. & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. In Jun, S.A. (Eds.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, 362-389.

Grice, M., Ritter, S., Niemann, H. & Roettger, T.B. (2017). Integrating the discreteness and continuity of intonational categories. In *Journal of Phonetics*, 64, 90-107.

Grice, M., Savino, M. (1997). Can pitch accent type convey information status in yes-no questions?. In *Proceedings of the ACL Workshop 'Concept-to-Speech Generation Systems'.* Madrid, 14 July 1997, 29-38.

KleinSchmidt, D.F. (2019). Structure in talker variability: How much is there and how much can it help? In *Language, cognition and neuroscience*, 34(1), 43-68.

Ladd, D.R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Loporcaro, M. (2009). *Profilo linguistico dei dialetti italiani*. Roma-Bari: Laterza.

Marotta, G., Sorianello, P. (2001). La teoria autosegmentale nell'analisi dell'intonazione interrogativa in due varietà di italiano toscano (Lucca e Siena). In *Atti del XXXIII Congresso della Società di Linguistica Italiana*. Roma: Bulzoni Editore, 177-204.

Mücke, D., Grice, M. (2014). The effect of focus marking on supralaryngeal articulation Is it mediated by accentuation? In *Journal of Phonetics*, 44, 47-61.

Niebuhr, O., D'Imperio, M., Gili Fivela, B. & Cangemi, F. (2011). Are There "Shapers" and "Aligners"? Individual Differences in Signalling Pitch Accent Category. In *ICPhS*, 120-123.

Orrico, R., Cataldo, V., Savy, R. & Barone, L. (2016). Transfer, fossilization and prosodic drift in Foreign Language Learning. In Savy, R., Alfano, I. (Eds.), *La fonetica nell'apprendimento delle lingue*. *Studi AISV 2*. Milano: Officinaventuno, 117-132.

Orrico, R., Savy, R., D'Imperio, M. (2019). The perception of speaker certainty in Salerno Italian intonation. In Calhoun, S., Escudero, P., Tabain, M. & Warren,

P. (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 2946-2950.

Petrone, C., D'Imperio, M. (2008). Tonal structure and constituency in Neapolitan Italian: evidence for the accentual phrase in statements and questions. In *SP-2008*, 301-304.

Pierrehumbert, J.B. (1980). The phonology and phonetics of English intonation. PhD dissertation, Massachusetts Institute of Technology.

Prieto, P., Borràs-Comes, J. & Roseano, P. (2010). *Interactive atlas of Romance intonation*. Web page: http://prosodia.upf.eduiari.

Prieto, P., D'Imperio, M. & Gili Fivela, B. (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. In *Language and speech*, 48(4), 359-396.

Savino, M. (2012). The intonation of polar questions in Italian: Where is the rise?. In *Journal of the International Phonetic Association*, 42(1), 23-48.

Savino, M., Grice, M. (2011). The perception of negative bias in Bari Italian questions. In Frota, S., Prieto, P. & Elordieta, G. (Eds.), *Prosodic categories: Production, perception and comprehension*. Dordrecht: Springer, 187-206.

Savy, R., Luque Moya, J.A. (2015). Aspectos prosódicos de las interrogativas en aprendientes italianos de ELE. In *NORMAS*, 7, 283-296.

Zeileis, A., Meyer, D. & Hornik, K. (2007). Residual-based shadings for visualizing (conditional) independence. In *Journal of Computational and Graphical Statistics*, 16(3), 507-525.

DOI: 10.17469/O2106AISV000019

DUCCIO PICCARDI

Bilingual Frequency in a Favorable Context (BFFC) in the Italian dialectal area. Theoretical preliminaries to the analysis of geminate lateral retroflexion and voiceless plosives aspiration in Antona (MS)

The paper builds a theoretical framework for the application of the Bilingual Frequency in Favorable Context (BFFC) formula to the peculiar Italian linguistic setting. BFFC was first devised as a usage-based tool to weight the frequency effect of non-varying cognate words against the probability of variation phenomena in bilingual settings. Since Italian dialects are sister languages of the standard variety, speakers can be considered bilingual. However, no dialectal frequency corpora for the extraction of essential BFFC components are available. The paper suggests overcoming this hurdle using subjective frequency estimates and testing BFFC effectiveness through a picture-naming task and acceptability ratings. A critical overview of the phonetic features of interest is also presented, advancing proposals for future analyses.

Key words: Bilingual Frequency in a Favorable Context, retroflexion, aspiration, usage-based linguistics, dialectology.

1. *Two focal points of usage-based theory*

Langacker (1987: 370) was the first to advocate «the necessity for a nonreductive, usage-based model» of linguistic analysis. In recent years, a converging scientific trend (Bybee, 2006) brought together specialists from different areas, such as grammaticalization, variationist sociolinguistics, phonetics and first language acquisition, to examine corpora of natural language use. Here, communicative experience is deemed responsible for speakers' adjustments of gradient boundaries of linguistic categorizations, conditioning production and perception. Usage-based approaches consider language as a part of domain-nonspecific cognitive processes affecting human behaviors. In particular, great attention is dedicated to cooccurrence detection as a function of frequency of exposure, a «domain-general process of the human mind» (Bybee, Beckner, 2009: 830).

Phonetics adopted a usage-based mindset through exemplar models of representation, thanks to the seminal works of theoretical systematization by Johnson (1997), Pierrehumbert (2001, 2002) and Bybee (2001). In this framework, exemplars are stored as eventive traces in the mind of the individual, including fine-grained phonetic details and extralinguistic information (context, participants etc.). These episodic units are structured in metaphorical clouds, organized by the similarity of their components.

In the first section of this paper, two aspects of usage-based approaches will be discussed, namely which phonological phenomena they can account for and which frequency they are about. Albeit central, these topics may still need some degree of clarification. After introducing a challenging setting of research, i.e. the Italian dialectal area, where frequency corpora are not available, I will discuss the potential for this line of research of recent expedients of usage-based bilingual phonetics. Finally, suggestions and solutions will be sketched in a theoretical preview of a research in progress on the production of geminate lateral retroflexion and voiceless plosives aspiration in the dialect of Antona (MS).

1.1 Variation and reduction in frequency effects

The two conditioning forces that token frequency induces on the role of single words in potential sound change are labeled in Bybee's framework Conserving Effect and Reducing Effect (e.g. Bybee, 2006), accounting for analogical and phonetically based shifts, respectively. As I will propose below, this bipartition may not represent the whole spectrum of inherent phenomena; in this vein, Barras, Honeybone & Trousdale (2007) advanced an opposition between Frequency and Anti-Frequency Effects, putting stress on the directionality of frequency tendencies on language change.

As experience of similar linguistic events grows, speakers' categorization of the so formed exemplar cluster strengthens. Central, denser exemplar groups may attract nearby clouds, promoting analogical change in their direction. In the target clouds, marginal (and less frequent) events are more prone to drift towards nearby stronger categories. High-frequency words tend to *conserve* their categorial affiliation, while low-frequency words lead in analogical change. De Schryver, Neijt, Ghesquière & Ernestus (2008) present a rather straightforward instance of this pattern. In a judgement of standardness test, Dutch participants selected as standard the words containing voiceless allophones of phonological voiced fricatives if the stimuli were in the low frequency range, as assessed through subjective estimations (see below, 1.2). This result, modulated by diatopic factors, substantiates the hypothesis of a sound merger in progress between the voiced and voiceless fricative categories in this language.

In contrast to the trans-categorial, relational assumptions required by usage-based interpretations of analogical change, denser clouds may trigger *per se* neuromotor automation, independently of the structure and the qualities of nearby clusters. Bybee (2002: 268-269) explains this phenomenon as a general process of gestural optimization through constant practice, resulting in anticipatory overlaps (assimilation), omissions of unnecessary components (deletion) and reduction in the magnitude of the action (lenition). These phenomena affect first high frequency words, whose exemplar representations rapidly become dominated by reduced variants. Hay, Foulkes (2016) recently wrote a landmark study observing for the first time a reducing effect leading to actual change in a real time corpus. In New Zealand English, intervocalic medial /t/ sonorization spread following a positive frequency pattern and interacting with the year of birth of the recorded speaker. This result is fundamental for usage-based phonetics, since it anchors to real diachronic observations what was until then limited to predictions of change or speculations on synchronic data.

In the last years, a sociophonetic line of research, pioneered by Hay, Jannedy & Mendoza-Denton (1999), followed Pierrehumbert (2002: 118)'s assertion that word frequency positively affects the rate of change in the composition of any exemplar cloud, favoring newer variants in overcoming the representational balance: in other words, any type of change, and not only reduction, is led by frequent words. This can be interpreted as a major departure from Bybee's point of view, since frequency effects are not considered as unambiguous flags of cognitively determined change in production[1]. In Pierrehumbert's framework, the explanation of variation and change phenomena remains outside the basic assumptions of the usage-based framework *per se*, i.e. the existence of general cross-domain cognitive mechanisms[2]. Even though the source of variation is ultimately physiological, an explanatory proposal should account for the systemic strategies of acceptance and accommodation of the new reduced variants (Todd, Pierrehumbert & Hay, 2019).

Despite these assumptions, a subtle sense of uneasiness still stands in phonetic research in leaving aside Bybee's approach, usage-based inquiries on strengthening phenomena being quite rare. Moreover, under certain circumstances, variation has been reconducted to reduction through phonological arguments. For example, Habib (2011) framed the lexically-driven, frequency affected change from Standard Arabic /θ ð/ to Hims dialectal /t s/ in a reduction process; analogous choices are made by Coetzee and Kawahara (2013) in their explanation of positive frequency patterns in variable geminate plosive devoicing of English loanwords in Japanese. These authors equate loss of phonological distinctiveness with reduction; however, Bybee's reducing effect should not be extended to phonological domains, these being pertinent to relational, trans-categorial properties. Otherwise, (positive) frequency effects risk becoming an ambiguous and potentially confusing explanatory *passe-partout*. In this sense, note that a robust tradition of usage-based studies (e.g. Brown, 2018 and references therein) applies to plosive spirantization the same principles evoked in Habib (2011) for a change in the exact opposite direction.

Last in this paragraph, it is noteworthy that usage-based theories are conceived to provide an explanation to language change, but not as much to stable variation. For example, it has been pointed out that other models, like Lindblom's (1990) H&H Theory, can easily explain stable reductions principally targeting high-frequency words (Forrest, 2017: 142) without recurring to exemplarist arguments. While, denying the pertinence of usage-based linguistics to diachrony, Dinkin

[1] An early discussion on frequency effects on non-reducing change is presented in Phillips' works, e.g. (2006: 41-56).

[2] From a sociophonetic viewpoint, Clopper, Mitsch and Tamati (2017: 39) asserted that «just as easy contexts for the listener allow for gestural economy leading to reduction, easy contexts for the listener allow for more extreme social indexing without sacrificing communication».

(2008: 105) defined synchronic reduction as «the real effect of word frequency», Abramowicz (2007) failed in finding frequency patterns in the realizations of reduced (ING) variants. Forrest (2017) clarified the issue, suggesting that researchers should leave aside raw frequencies and look for proofs of sensitivity to word frequencies in contexts promoting variation: if a word is frequently encountered in such environments, we can expect that traces of variation will also be present in the absence of them. Indeed, Forrest found that frequency in specific grammatical and phonological environments did condition (ING) stable reduction[3], strongly promoting a complexification of the frequency variable, which should include the negative effects of non-varying exemplars in order to boost the accuracy of the models.

1.2 Conceptualization of frequency: from objectivity to subjectivity

Usage-based linguistics is constitutionally bound to the idea that individual lives exert a great pressure on the observed data. The formation of linguistic categories and patterns of use are the ultimate result of «a personal odyssey» (Foulkes, 2010: 19), modulating experimental acquisitions. Drager (2011) adopted an objective, direct and individual concept of frequency analyzing the interviews recorded during one year of ethnographic fieldwork in a New Zealand high school. Here, the more a specific function of *like* occurred in the speech of individual students, the more this specific *like* became target of phonetic reductions. Through her extensive ethnographic effort, Drager described an ideal method for usage-based research, which is able to account for individual behaviors without resorting to external sources. Even so, the author herself (ibid.: 697) admits that «much larger corpora would be required to calculate representative speaker token frequencies».

In most of the other cases, practicality drove usage-based research towards the assumption of the so-called *from-corpus-to-cognition-principle*, i.e. the idea that «frequency in text instantiates entrenchment in the cognitive system» (see Schmid, 2000: 38-40). Word representativeness in a corpus is seen as a window on systemic and, ultimately, cognitive structures. As we saw at the end of the previous paragraph, this objective, indirect and communal take on frequency is undergoing a process of refinement, leaving behind raw corpus values. In Bybee (2002), the scholar assumed that, since the exemplar clusters are activated by phonetic strings, cognitive representations do not involve words taken from a phonetically abstract co-text, but instead occur in acoustic sequences which are not equally prone to variation. Her point was substantially left unrefined until the works by Esther Brown[4]. In her Doctoral dissertation, Brown (2004) investigated the variable phenomenon of Spanish initial /s-/ reduction using a spoken corpus from New Mexico and Southern Colorado. She devised a formula, the Frequency in Favorable Context (FFC), which calculates a score dividing the frequency of a word occurring in a context of potential

[3] Analogous results are reported in Raymond, Brown & Healy (2016) for final /t d/ deletion.
[4] For an alternative operationalization, see Eddington, Channer (2010).

variation[5] by its overall occurrences. This score proved itself to be predictive of the effective reduction probability of /s-/ words. Raymond, Brown (2012) further promoted the validity of the method using logistic regression to discern the significance of FFC values from other cooccurring factors, including overall frequencies. Brown, Raymond (2012) attempted to apply FFC to a diachronic analysis. They observed that weighted values, extracted from a large Medieval Spanish text, outscored raw frequencies with regard to the prediction of the [h] outcomes of Latin initial F- (e.g. *hablar* < FABULARI). Note that in this study FFC was the only parameter reaching significance, differently from absolute measures and word transmission history coding (i.e. oral vs. non-oral). Thus, this research decisively defined proofs by which the use of raw textual frequencies should be avoided. Brown (2013) replicated this analysis, along with that on /s/ reduction, while adding a part-of-speech tagging to the statistical model. Verbs held the highest FFC values, and consequently became the most reduced category. The study tries to support the idea that different involvement in variation and change of distinct grammatical categories is just an epiphenomenon of peculiar patterns of use. Lastly[6], Brown (2018) used word FFC ratings to predict the type and the extent of Spanish initial /d/ reduction elicited in a list of words reading task administered to 18 speakers from disparate Spanish-speaking countries. This format was chosen with the aim of reducing the potential conditioning factor of style. FFC ratings significantly correlated not only with the type of elicited variant (low FFC: [t], [d], [nd] vs. high FFC: [ð], [∅]), but also with the intra-categorial degree of articulatory strength, calculated as the intensity difference between the consonant and the following vowel.

Alongside these developments in corpus-driven methods, a subjective, indirect and individual route is tentatively followed by few researchers. To the best of my knowledge, De Schryver *et al.* (2008) was the first usage-based phonetic study to rely on subjective frequency estimates instead of objective, corpus-extracted ones[7]. Word subjective frequency estimation (or word familiarity[8]) is a major field of interest in cognitive research, pursued since shortly after the compilation of the first textual word frequency corpus (Thorndike, Lorge, 1944; see Howes, 1954) through a variety of experimental designs. Among those, a few influential studies between the end of the Seventies and the beginning of the Eighties popularized rating scales (e.g. Toglia, Battig, 1978; Nusbaum, Pisoni, & Davis, 1984). Overall, this line of research

[5] In this specific case, the context favoring variation is the presence of contiguous non-high vowels.

[6] For Forrest (2017) and Raymond, Brown & Healy (2016) see above, 1.1.

[7] After De Schryver *et al.* (2008), subjective measures have been used in phonetic research by Myers, Li (2009) on syllable contraction and tonal merger in Southern Min; Lee, Kapatsinski (2014) and Song, Dalola (2019) on loss of productivity of Korean /n/ insertion; Nadeu (2016) on the reduction of full vowels in unstressed position in Central Catalan compounds; Zhang, Meng (2016) and Yan (2018) on the variable realization of Shanghai Wu tone sandhi and, lastly, by Yan, Zhang (2017) and Yan (2018) on the same phenomenon in Wuxi Wu.

[8] Here I cluster together these two definitions for practicality; however, it should be kept in mind that some differences between the two concepts have been underpinned, especially from the procedural point of view (Balota, Pilotti & Cortese, 2001).

finds significant correspondences between subjects' perceived word frequencies and corpus data, not without some discrepancies (see e.g. Carroll, 1971, Thompson, Desrochers, 2009). For this reason, Ringeling (1984) argued that subjective protocols may be reliable substitutes for textual corpora when the latter are not available at all. Turning back to linguistic inquiries, Balota *et al.* (2001) succeeded in proposing a standardization of the research protocol which has been thereafter adopted by the majority of usage-based phonetic studies dealing with subjective measures. Balota and colleagues collected frequency estimations for 2938 English monosyllabic words from a total of 2254 respondents, using a 7-point Likert scale with labels assigned at each point. This study was probably accepted by the linguistic community because of its unprecedented quantity of observations; moreover, some of its methodological features, such as the insertion of age and diamesic variables, are in line with variationist concerns.

Subjective frequency estimation represents an attractive middle-ground between the time-consuming attention to the individual of Drager (2011) and the rapid, objective corpus observations. In the next chapter I will introduce an ideal ground for the application of this methodology: usage-based dialectology.

2. *Usage-based dialectology*

Anderwald and Szmrecsanyi (2009: 1126) ironically stated that, «in contrast to sociolinguistics, dialectology and corpus linguistics have been rather uneasy bedfellows until relatively recently». This is due to the general scarcity of publicly available corpora of local varieties. Even though crowdsourcing technology is gradually trying to change this picture (e.g. Cenceschi, Trivilini, Sbattella & Tedesco, 2019), for the time being, usage-based dialectology has found very limited expressions. For example, works by Lynn Clark are inserted in a usage-based framework thanks to the author's corpus of field materials (e.g. Clark, 2008); in other occasions, arguments of English dialectology are inferred from a geo-referenced subset of a corpus of the national language (Clark, Watson, 2011). This procedure cannot be exported to the Italian linguistic landscape, where dialects are not varieties of the standard, but standalone languages. In this chapter, the bilingual dynamics of Italian dialect-standard interactions will be discussed; then, methods and resources of usage-based bilingual phonetics will be explored and tentatively adjusted to fit the Italian situation.

2.1 Italian dialects: advergence in bilingualism

In Italy, dialects are «'sisters' of Italian» (Maiden, Parry, 1997: 7), and not local varieties of the standard language. This is no mere semantic subtlety (Regis, 2013), as the Italian acceptation encodes the peculiar history behind its linguistic landscape. In fact, Italian primary dialects[9] (Coseriu, 1980) are direct developments of the Latin

[9] From this point on, I will refer to this category of idioms as "dialect". On the other hand, here "standard" does not stand for the Italian reference linguistic construct, but for its local instantiations (unless

spoken in the regions of the peninsula. After the linguistic debates of the XVI century these varieties came in an asymmetric sociolinguistic contact with a reference prestige variety based on a Florentine koine (e.g. Regis, 2017). After the achievement of political unity and the diffusion of a uniform educational system (Sobrero, 1996), this reference lect was codified as the Italian standard. In the meantime, the onset of intermediate varieties favored the formation of a sociolinguistic continuum, explaining the acceptation of "dialectology" as, *de facto*, the study of systems historically independent from the national one. In addition to that, the structural distance between the national (and regional) standard and the dialects «justif[ies] treating the Italian situation as bilingual rather than merely bidialectal» (Berruto, 1997a: 394)[10]. Note that both recent neurolinguistic (Schmitt, Auer & Ferst, 2019) and psycholinguistic (e.g. Mello, Miozzo & Peressotti, 2017) studies confirmed that this theoretical standpoint reflects an observable cognitive reality, since experimental data from standard-dialect bilectal subjects are comparable with those from bilingual ones.

Recent surveys promoted by the National Institute of Statistics (e.g. ISTAT, 2017), based on people's self-evaluations, highlighted that, while monodialectal speakers are decreasing, bilingualism is rather stable: on this basis, Dal Negro, Vietti (2011: 89) asserted that «on a national level the bilingual mode appears as the unmarked or default choice». This widespread situation of multiple competences led to the loss of functional compartmentalization of the concurring codes, which are often coexisting at the same diaphasic level ("dilalìa" in the terminology of Berruto, 1987). As a consequence, dialects are gradually losing peculiar lexical and phonological features (e.g. Cerruti, 2016). Berruto (e.g. 1989, 1997b) repeatedly manifested dissatisfaction with the use of the term convergence (Sanga, 1985) describing this phenomenon. In fact, the two main prerequisites of convergence (bilateral participation in a centripetal narrowing and sociolinguistically balanced status) are not met in the Italian milieu. For this reason, the scholar (Berruto, 2005a) strongly promoted the adoption of the term advergence (Mattheier, 1996), defining a unilateral movement of the dialect towards an (idealized) standard.

Given dilalìa, it has been argued that one of the most representative loci of advergence is code-switching[11] (Cerruti, Regis, 2005: 198). Studies on this phenomenon noted an extreme variability and compositional freedom in its Italian manifestations. In this sense, Berruto (2011) asserted that, during code-mixing, speakers do not activate one of two competing grammars, nor a permeation of the two, but both grammars at the same time, in a cumulative, unified fashion. Overall, code-switching can be considered a surface phenomenon, which is better described by variationist

otherwise stated). In the words of Berruto (2018: 510), «nobody in Italy can be considered to be a true native speaker of standard Italian. There are, instead, several regional standardized varieties of Italian».

[10] This assumption does not stand for the speakers of Tuscan varieties, whose competence can be better described with the tools of variationist sociolinguistics (or sociophonetics: see the overview in Calamai, 2017). The same can be said for the Roman speakers, since their dialect incurred in a profound tuscanization during the XV, and mostly XVI, centuries (Trifone, 1992: 28-50).

[11] Here I refer to code-switching as a hypernym of all the phenomena of code alternation and mixture, excluding explicit borrowings.

sociolinguistic accounts[12]. Indeed, recent attempts at explaining specific instances of dialect-standard switching (Cerruti, 2018) implemented a probabilistic view on their emergence, akin to comparative variationist methodologies (Tagliamonte, 2013). This theoretical shift somehow reflects the birth of a usage-based approach to code-switching at the international level (Backus, 2015). In this framework, insertional code-switching and borrowings are seen as residing on a single continuum of cognitive entrenchment. Therefore, researchers should compare the frequencies of the alternating structures in order to predict their degree of fixated production as elements of discourse variation promoting eventual innovation. Considering linguistic units larger than the word this approach has not yet been applied to phonetic features. On a general note, studies on the phonetics of code-switching are still very scarce (see e.g. Muldner, Hoiting, Sanger, Blumenfeld & Toivonen, 2019); however, the point here is, even interpreting code-switching as a variation phenomenon occurring in a single, cumulative grammar, we cannot assess the predictive power of individual lexical items on phonetic realizations if we cannot be sure that the switched slot is going to be replaced with a semantically and phonetically similar element. In other words, if the two switching elements do not share the same etymological base, variation will necessarily occur at the lexical, and not phonetic level.

Since Clyne (1967), it has been noted that cognate[13] words may favor the emergence of code-switching. This triggering effect has been repeatedly reported in the Italian context (Alfonzetti, 1992: 240; Berruto, 2004: 65-67; Cerruti, Regis, 2005: 197). More recently, psycholinguistic research on cognates (Broersma, De Bot, 2006; Broersma, Carter, Donnelly & Konopka, 2019) classified them as cognitive nodes, linking separately grouped, but nevertheless coexisting language representations. During bilingual discourse, both grammars are activated at some level; the occurrence of a cognate bolsters the arousal of the recessive code, promoting a switch in its proximity. Interestingly enough, phonetic discrepancies between the two cognate words do not hinder their role as cognitive bridges in the mind of bilinguals (Broersma, 2009). This has also been confirmed for the Italian situation, as real-time studies (Ghimenton, 2013a, 2013b) noted that standard-dialect cognates hold a pivotal role in children's dialect acquisition through minimal inputs.

Wrapping up this paragraph, I would argue that the assessment of the probabilistic emergence of these cognate nodes as phonetically attributable to one of the two coexistent systems may be a viable route to quantify phonetic advergence from a usage-based dialectological perspective.

2.2 The usage-based contribution to cognate analysis

Today, tentative quantifications of phonetic advergence are still rarities, all concerning cognate analysis. Among those, Auer, Schwarz (2014) is explicitly interested

[12] Such approach preludes to a transformation of code-switching in a variation phenomenon of the sociolinguistically dominant language *tout-court* (see Mioni, 1992: 18 in Cerruti, Regis, 2005: 200).
[13] In this work, I will use this term in its psycholinguistic acceptation, implying a degree of formal and semantic similarity between the words sharing the cognate status.

in frequency effects. The authors searched for lexical gradience in Alemannic dialect-standard German vocalic advergence through a corpus subset of spoken interviews held in Baden-Württemberg. Production data are examined in the light of values extracted from a frequency dictionary based on southwest German materials; however, all the conclusions drawn for this analogical drift towards standard vocalic categories are lacking an appropriate quantification of the attracting force, i.e. the clusters of standard German-like productions.

Recently, usage-based phonetics has developed a valuable tool in order to investigate frequency effects in bilingual settings. Brown, Harper (2009) studied the conditioning factors of Spanish final sibilant reduction in two corpora of spontaneous speech, one from New Mexico, the other from Chihuahua. They devised two binary variables, i.e. the English Exemplar Connection (EEC) and the cognate factor, indicating different levels of similarity between the inquired Spanish words and their English equivalents. In the New Mexican data, gathered from bilingual speakers, the stronger are the Spanish word exemplar connections with their English equivalents without the inquired variation phenomenon, the less probable is their reduction. This is explainable by positing the storage of exemplars from the two different languages in single clouds, variably populated by non-reduced English exemplars. On the contrary, EEC and the cognate factor were not predictors of resistance to reduction in the data from Chihuahua (Spanish monolinguals). Brown (2015), building on these results, refined the above-discussed FFC formula, implementing the weighting of cognate exemplars. She examined Spanish /d-/ reduction in a Spanish-English code-switching corpus from New Mexico. First, the methodology of Brown, Harper (2009) was replicated, coding a binary cognate variable (e.g. Sp. *Diferencia*, Eng. *Difference* vs. Sp. *Domingo*, Eng. *Sunday*) while adding FFC[14] to the analysis. Both factors proved to be significant; however, focusing the analysis on the cognate subgroup, FFC lost its predictive power. With the aim of providing a more fine-grained formalization of the effects of the bilingual exemplar clouds, Brown repeated this test substituting FFC with the Bilingual Frequency in Favorable Context (BFFC). BFFC adds the overall (external) corpus frequency of the English cognate words to the divisor, contributing to counter the contextual, variation-favoring effects. Indeed, BFFC succeeded in modeling the differences within the cognate category; moreover, because of its own structure, the new formula accounted for the effects of the binary cognate variable, evidently providing a more compact procedure for this kind of research.

To date, no other attempts have been made at validating BFFC in other bilingual environments[15]. Nonetheless, I would advance that BFFC could be an invaluable

[14] Both values composing the FFC, i.e. the overall word frequency and the number of word occurrence in reducing contexts, were extracted from an external reference oral corpus. Spanish /d-/ reduction is triggered in non-post-pause, post-lateral and post-nasal contexts.
[15] Brown, Amengual (2015) investigated the same phenomenon in a corpus of Spanish-English Puerto Rican bilinguals as well as Spanish plosive Voice Onset Time (VOT) lengthening in the speech of Texan heritage speakers, while adopting the methodology of Brown, Harper (2009).

parameter for the studies on standard-dialect advergence. In fact, note that BFFC structure can account for both frequency effects and anti-frequency effects, interpreted as analogical variations in the direction of dominant standard representations. If BFFC positively correlates with the probability of occurrence of the inquired dialectal variation phenomenon, this should be classified as endogenous, i.e. promoted by dialectal entrenchment and disfavored by competing standard cognates. On the contrary, if BFFC and variation are negatively proportional, we may be observing an analogical shift towards standard norms, i.e. an *advergence*. Moreover, in both cases, comparing results for BFFC and FFC may contribute to shedding some light on the degree of functional overlapping of the two coexisting codes: a dense, dilalic code-switching milieu should lead to robust inter-language exemplar clustering, thus favoring BFFC. In the next section, a protocol for a usage-based phonetic analysis of peculiar dialectal features will be outlined, summing up the potential benefits of subjective estimates and bilingual frequency expedients.

3. *Naming pictures in the Apuan Alps. An Italian BFFC protocol*

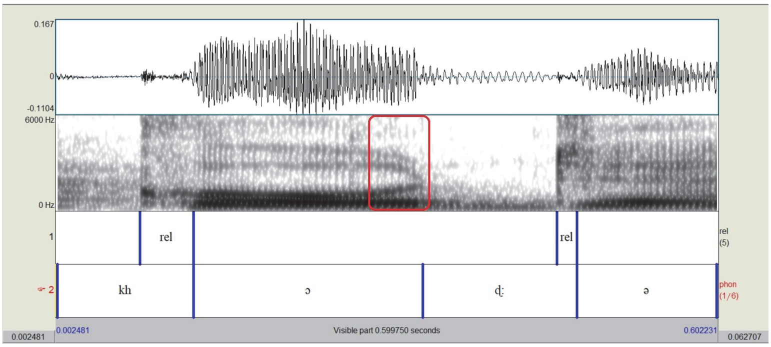
Antona (402 MAMSL; ca. 300 inhab.) is a small village in the upper Frigido valley on the Apuan Alps. Administratively speaking, it is part of the municipality of Massa, a regional chief town in northern Tuscany. The classification of the Antonese dialect is no easy task, as it belongs to the transitional Apuan group (Carpitelli, 1995: 44-47). While Pellegrini (1980^2: 30) in his Map of the Italian Dialects placed the Apuan varieties in the Tuscan cluster, Giannarelli (1912: 12) and Ambrosi (1966/1967: 126) considered them a subgroup of the Lunigianese area. Emblematically, both the volumes on Tuscany (Giannelli, 2000: 129-130) and Lunigiana (Maffei Bellucci, 1977: 18-20) which form part of a seminal series on Italian dialects directed by Manlio Cortelazzo hold the other party responsible for discussing the Apuan varieties, ultimately leaving the Italian tradition without a systematic overview on them. In actual fact, as explicated in Giannarelli's (1912) proposal, these dialects are made of an intricate superimposition of Tuscan and Ligurian isoglosses, not without Emilian echoes. Despite the relative isolatedness of the Apuan communities, their advergence to standard features have been noticed long ago (Ambrosi, 1956: 33); conversely, local identity focuses on specific dialect traits, promoting occasional, irregular overextensions (i.e. hyperdialectism: Carpitelli, 2017: 219). On the basis of a general structural contiguity with the Tuscan varieties, I would stress that usage-based accounts could clarify the effects of standard pressure on specific phonetic characteristics with different degrees of local salience. In the remaining space, I will discuss two Antonese features of potential interest, namely geminate lateral retroflexion and voiceless plosives aspiration; then, I will outline a usage-based research in progress on them.

3.1 Retroflexion and aspiration in Antona

Geminate lateral retroflexion (e.g. [ˈvaɖːə] vs. It. [ˈvalːe], "valley") was first reported for the area in Bottiglioni (1911) and its geographic boundaries, revolving around mount Pisanino, were later defined in Bonin (1952: 146) and Ambrosi (1956). This feature was the object of an intense linguistic debate on its origins, variably hypothesized as Corsican (Bottiglioni, 1926: 6), Ligurian (Bottiglioni, 1930: 18) and Mediterranean (Merlo, 1956), until Savoia (1980: 280 and ff.) put an end to the substratist speculations by advancing solid chronological arguments in favor of the relative recency of the trait. Be that as it may, it is undeniable that the singularity of retroflexion is strongly perceived by local speakers: in a description of the Antonese variety written by a non-specialist (Della Sala, 1986), retroflexion is emblematically put first and classified as typical of the most authentic stratum of the population (i.e. elderly people with limited contacts outside the village).

To date, no acoustic data has been published regarding Antonese retroflex plosives. The only spectrographic (and palatographic) information on the phenomenon in this area is reported in Rossi (1976: 415-416) for the variety of Bosco di Rossano in Lunigiana, where, however, the local realization is the approximant [ɭ] and not the plosive [ɖ]. While studies on analogous retroflexions in other Italian dialects found hints towards their classification as strengthening (i.e. long release phases, gradient loss of formant information; for Sicily and Corsica: Celata, 2006), the directionality of variation is irrelevant from a usage-based point of view: in fact, the application of BFFC to the Antonese dialect would encode the gradient nature of an analogical shift towards the Italian concurrent realization, i.e. [lː].

Figure 1 - *[ˈkʰɔɖːə] (It. [ˈkɔlːa], "glue") uttered in continuous speech by a 75 years old female Antonese speaker. Velar release is 49 ms. long, well above the arbitrary 30 ms. threshold set by Cho, Ladefoged (1999: 223) for aspirated plosives. The red box highlights the descending F3/F4 transitions to the retroflex, leading from vowel midpoint 2845, 3848 Hz to endpoint 2087, 2777 Hz, respectively*



On the other hand, a strictly strengthening, dialect-internal variation in Antonese is voiceless plosives aspiration, a phenomenon brought to light in Giannelli, Savoia (1979-1980: 84). Aspiration is reported as mainly targeting velar and dental segments, both short and long; however, in singleton, intervocalic plosives, aspiration competes with sporadic sonorization, implying, for example, the coexistence of the forms [ˈfoːkʰə] and [ˈfoːk̬ə] for "fire". Again, no acoustic data are available for Antonese aspiration. In Savoia (1980: 242-243) other examples are discussed, and aspiration is synthesized in a variable rule conditioned by social factors, without further specifications. Compared to retroflexion, aspiration is far less salient for local speakers, and is not mentioned in Della Sala (1986). Since the voiceless plosives of standard Italian are realized as unaspirated (e.g. Sorianello, 1996), Antonese has at its disposal both standard-like, unaspirated variants and divergent, aspirated ones. Assuming that positive frequency effects can account for variations other than reduction (cf. Pierrehumbert, 2002), aspiration of Antonese long voiceless plosives should follow cognitive patterns similar to those encoded by the BFFC formula in Brown (2015). Italian exemplars should enhance the density of the unaspirated dialectal ones, countering the emergence of aspirated allophones. Figure 1 illustrates both Antonese aspiration and retroflexion, while relevant acoustic parameters will be mentioned in the next paragraph.

3.2 Subjective/Objective Bilingual Frequency in Favorable Context (SOBFFC) and its applications

This section contains a summary of the current status and future perspectives of a research in progress in Antona, aimed at testing the effectiveness of a usage-based approach centered on BFFC in the Italian bilingual situation.

A list of Italian words containing geminate laterals and voiceless plosives was extracted from itWaC (Baroni, Bernardini, Ferraresi & Zanchetta, 2009) through a case-unsensitive query. In order to slim it down, the words were administered to local informants (two men aged 25 and 30, and a woman aged 56) asking them to gray out those items whose dialect equivalents were hugely dissimilar (i.e., among others, *parpagòn* vs. It. *pipistrello* for "bat"; cf. Mosti, Nancesi, 2005: 378). The remaining words were provisionally[16] considered cognates of dialect forms, whose exact pronunciation was asked as well during this phase. The corpus frequencies of these Italian words were extracted for the BFFC computation and log-transformed (Baayen, 2001: 32-34), while values for their dialect counterparts will be retrieved as follows.

Antonese has a small written tradition, constituted by collections of prayers (Mariotti Giromella, 2006), proverbs and sayings (e.g. Bertuccelli, Casotti, 1982: 43-44). These works are widely known among the villagers: even though extracting

[16] As an anonymous reviewer rightly pointed out, speakers' age plays a major role in determining individual knowledge of the dialect lexicon (e.g. Wieling, Montemagni, Nerbonne & Baayen, 2014). For this reason, the absence of non-cognate alternatives for the extracted Italian words cannot be ascertained without a socially stratified lexicographic inquiry in the village. I should clarify that this was not the goal of this phase. While I was indeed trying to numerically reduce the list of items, the eventual presence of residual non-cognates can be spotted through the results of the rating scales and participants' comments (see below).

from them word frequency values is clearly unfeasible because of their smallness and lack of thematic variety, they created some level of orthographic agreement for non-standard phones, exploitable in testing conditions. The dialect component of BFFC will be gathered through subjective frequency estimation, adapting the guidelines of Balota *et al.* (2001; see above, 1.2). In particular, I will neglect their diamesic bipartition, since written dialect occurrences are still limited to the rare above-mentioned circumstances; moreover, a complexification of their focus on social differences will be needed, adding a sex variable and a classification based on attitudinal criteria. This will be calculated mixing scores taken from three short questionnaires, i.e. the Regionality Index, the Language Use Index and the Ethnic Orientation Index (Dollinger, 2015: 285-294). While the eventual fluctuation in the perception of frequencies depending on dialect attitudinal scores will be object of interest *per se*, these measures will also be exploited for the interpretation of the results of the testing phase. Dialect words will be presented in their orthographic adaptations. This methodological choice will inevitably lead to some level of individual reinterpretation; however, since Myers, Li (2009) proved subjects' sensibility to phonetic details in subjective frequency assessments, proposing acoustic stimuli taken from *ad hoc* elicitations of local informants would have resulted in a large number of uncontrollable factors derived from potential idiolects.

FFC formulae are concerned with weighting the frequency of exposure to context promoting variation against word overall frequency. Note that both retroflexion and aspiration could be analyzed from this point of view. In fact, retroflexion may occur also in across-boundaries contexts, triggered by syntactic doubling (Loporcaro, 1997) in specific word combinations (e.g. [aˈɖːɛtːə] vs. It. [aˈlːɛtːo] "to bed"; Savoia, 1980: 284-285); on the other hand, we saw that singleton voiceless plosives in intervocalic context can either become sonorized or aspirated, depending on social factors. In order to avoid difficulties in representing compositionality and vague extralinguistic coordinates in subjective estimations, the target dialect words will present the inquired phenomena in unambiguous contexts, namely word-internal for retroflexion and in geminates for the aspiration of voiceless plosives, as already coded in the itWaC queries for their standard cognates. By doing so, BFFC will be reformulated as the population mean of the word subjective frequency estimate divided by itself plus the objective frequency of its Italian cognate. I refer to this revised measure as Subjective/Objective Bilingual Frequency in Favorable Context (SOBFFC). Figure 2 illustrates the transition from FFC to SOBFFC.

Figure 2 - *The FFC family of formulae. These equations are synthetic representations of methodological choices adopted in Brown (2004) for FFC, in Brown (2015) for BFFC and in this study for SOBFFC*

$$\boldsymbol{FFC} = \frac{frequency\_in\_variating\_contexts}{overall\_frequency}$$

$$\boldsymbol{BFFC} = \frac{Spanish\_frequency\_in\_variating\_contexts}{Spanish\_overall\_frequency + English\_cognate\_overall\_frequency}$$

$$\boldsymbol{SOBFFC} = \frac{dialect\_mean\_subjective\_frequency}{dialect\_mean\_subjective\_frequency + Italian\_cognate\_overall\_frequency}$$

SOBFFC values should positively correlate with the probability of one word to be produced in its dialect form. Two psycholinguistic methodologies will be exploited in a second session of data collection in order to obtain indirect measures of use. Picture naming tasks are a common procedure for testing the effects of cognitive representations of cognates. Reaction times are usually interpreted as an approximation of the ease in word retrieval which can either be enhanced or impeded by the existence of cognates. However, both in single language (e.g. Hoshino, Kroll, 2008) and language-switching designs (Broersma, Carter & Acheson, 2016) cognate pairs are matched not only by their frequencies, but also by other semantic (e.g. imageability) and lexical (e.g. age of acquisition) properties, which condition the speed of retrieval (e.g. Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004). While normed corpora of images for Italian picture naming tasks do exist (e.g. Navarrete, Arcara, Mondini & Penolazzi, 2019), the same cannot be said for dialects. Since our inquiry is based on potential attrition between lexical parameters (such as frequency), assuming that the Italian values are valid also for the dialects could be counterproductive. For the time being, at the cost of granularity of observation, we should refrain from testing the processes involved in lexical retrieval and limit our analysis to the contents of the retrieval itself. In other words, subjects will be asked to name the pictures representing the selected words in a simulated context generating potential task ambiguity, such as "in Antona". Through this specification, I intend to stimulate exemplars coded for the communal practices of the village, for which SOBFFC should simulate the degree of dialect advergence to the standard variety. By doing so, the effect of SOBFFC on the rate of retroflex and aspirated responses will be evaluated.

Aspiration can be represented as a temporal interval between the release of a plosive and the beginning of signal periodicity of the following vowel (i.e. VOT: Lisker, Abramson, 1964). On the other hand, the acoustic definition of retroflexion is far more complex. Hamann (2003) mainly focused on F4 and F3 lowering in neighboring vowels transitions (see Fig. 1) and short closure and release phases[17].

[17] As I mentioned above, this property finds a peculiar counterexample in Italian dialects.

Other studies also searched for regularities in spectral moments, i.e. high skewness and kurtosis and low center of gravity and standard deviation. These tendencies depend on the lengthening of the front oral cavity during tongue retraction (see e.g. Tabain, Butcher, Breen & Beare, 2014). In the study previewed here, SOBFFC represents the effect of lateral *continuous* productions on the emergence of retroflex *stops.* For this reason, differently from aspiration, retroflexion will not be treated as a gradient property, but as a class of responses featuring voicing and occlusion. In other words, SOBFFC will not try to predict the degree of retroflexion, but its absolute presence (a "dialectal" response) against its absence (an "Italian" response). While linear mixed effect statistical models can account for continuous response variables (in our case, aspiration), binary responses imply a logistic regression model with mixed effects. Therefore, in the case of retroflexion, SOBFFC will be treated as a fixed effect in a generalized linear mixed model.

Since Pierrehumbert (1994), acceptability judgements found their way into the representation of probabilistic knowledge of sound patterns. However, to date, this methodology has predominantly been used in usage-based syntactic and morpho-syntactic studies, with the underlying assumption that different levels of entrenchment entail proportional perceptions of grammaticality (Bybee, Eddington, 2006: 349). In this framework, researchers questioned the actual correspondence between frequency and acceptability, finding significant discrepancies at the low end of the frequency spectrum (Bermel, Knittl, 2012). Divjak (2017) specifically tried to clarify this issue, finding out that both word frequency and n-gram frequency are worst predictor of acceptability than conditional probabilities, i.e. measures that weight the frequency of occurrence of the inquired element in a specific construction against the overall frequency of the construction (attraction) or of the element (reliance). As this discovery is in line with the FFC tradition, I would argue that an application of these formulae to phonetic acceptability judgments could hold promising results. Research designs interested in observing correspondences between subjects' intuitions and actual use often rely on naturalistic utterances submitted as stimuli (e.g. Bresnan, Ford, 2010). In order to counterbalance potential effects of sentence compositionality beyond the inquired construct, common practices advise for the selection of multiple phrasings containing the test items; as Schütze, Sprouse (2013: 39) advocate eight or more different lexicalizations for syntactic structures, we may suppose that this quantification exponentially grows for small units of analysis, such as specific segments. Indeed, the problem is not posited in inter-categorial analyses, where specific phonetic variants are overall considered predictors of specific judgments against a (e.g., standard) baseline (e.g. Leivada, Papadopoulou, Kambanaros & Grohmann, 2017). However, in our case, the object of interest is intra-categorial, since the posited effect of frequency applies in a gradual manner to the judgements of the same category of items: in other words, the higher the SOBFFC value of a dialect word, the higher this should score in acceptability ratings. The design explicated in Kawahara (2011) could be adapted to this research. Raters provided judgments of naturalness of words containing the inquired phonetic variation with a 5-point scale ranging from very natural to very unnatural. Stimuli were proposed in the template *given [the baseline form], how natural would you find it to pronounce it as [the varied form]?*. Through this phrasing, baseline scores remain implicit; nonetheless, we may assume that competing Italian exemplars are activated and computed against the dialect form, as both are explicitly mentioned in the questionnaire. As shown in Kawahara (2011), participants' ratings can be investigated regressing them against a linear mixed effect model including frequency (in our case, SOBFFC) among its fixed effects.

4. *Conclusions*

In this paper, I outlined the theoretical background of a research in progress in the Italian village of Antona (MS). The research benefits from the acquisitions of usage-based bilingual phonetics: in particular, I tried to adapt the BFFC formula proposed in Brown (2015) to the analysis of dialect-standard advergence. Dialects do not usually have frequency corpora. For this reason, I suggested the integration of subjective frequency estimates into the representation of the competing dialect-standard exemplar aggregations. The selection of geminate lateral retroflexion and voiceless aspiration as target phenomena will set a test bench for both the under-tested assumption that frequency effects account also for non-reducing variation and the plausible application of BFFC to analogical change. For this purpose, I discussed the application of two psycholinguistic methodologies, namely picture naming and acceptability rating, to this specific research, not without proposing solutions to potential discrepancies between common practices and the current design.

Bibliography

Abramowicz, Ł. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). In *University of Pennsylvania Working Papers in Linguistics,* 13, 27-37.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U.H., & Segui, J. (2004). Predictors of picture naming speed. In *Behavior Research Methods, Instruments, & Computers*, 36(1), 140-155.

Alfonzetti, G. (1992). *Il discorso bilingue. Italiano e dialetto a Catania.* Milano: FrancoAngeli.

Ambrosi, A.C. (1956). Gli attuali limiti dell'area fonetica cacuminale nelle Alpi Apuane. In *Convegno di studi apuani. Rendiconti dei lavori della terza giornata in Carrara – 28 giugno 1955.* Carrara: D. Sanguinetti e figli, 33-40.

Ambrosi, A.C. (1966-1967). Le principali varietà dialettali della Lunigiana. In *La Spezia. Rivista del Comune*, 67, 121-126.

Anderwald, L., Szmrecsanyi, B. (2009). Corpus linguistics and dialectology. In Lüdeling, A., Kytö, M. (Eds.), *Corpus Linguistics. An International Handbook. Vol. II.* Berlin/New York: De Gruyter Mouton, 1126-1140.

Auer, P., Schwarz, C. (2014). Dialect-to-Standard Advergence. The Relevance of Compound Borrowing. In Torres Cacoullos, R., Dion, N. & Lapierre, A. (Eds.), *Linguistic Variation. Confronting Fact and Theory*. London: Routledge, 263-281.

Baayen, R.H. (2001). *Word frequency distributions*. Boston: Kluwer.

Backus, A. (2015). A usage-based approach to code-switching: The need for reconciling structure and function. In Stell, G., Yakpo, K. (Eds.), *Code-switching Between Structural and Sociolinguistic Perspectives.* Berlin/Boston: De Gruyter, 19-38.

Balota, D.A., Pilotti, M., & Cortese, M.J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. In *Memory & Cognition*, 29, 639-647.

Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In *Language Resources and Evaluation*, 43(3), 209-226.

Barras, W., Honeybone, P. & Trousdale, G. (2007). Is word-frequency a factor in phonological demergers? The sociophonology of SQUARE and NURSE in Lancashire English. Paper presented at the *2nd International Conference on the Linguistics of Contemporary English*, Toulouse.

Bermel, N., Knittl, L. (2012). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. In *Corpus Linguistics and Linguistic Theory*, 8(2), 241-275.

Berruto, G. (1987). Lingua, dialetto, diglossia, dilalìa. In Holtus, G., Kramer, J. (Eds.), *Romania et Slavia adriatica. Festschrift für Žarko Muljačić*. Hamburg: Buske, 57-81.

Berruto, G. (1989). Tra italiano e dialetto. In Holtus, G., Metzeltin, M. & Pfister, M. (Eds.), *La dialettologia italiana oggi. Studi offerti a Manlio Cortelazzo*. Tübingen: Narr, 107-122.

Berruto, G. (1997a). Code-switching and code-mixing. In Maiden, M., Parry, M. (Eds.), *The Dialects of Italy*. London/ New York: Routledge, 394-400.

Berruto, G. (1997b). Linguistica del contatto e aspetti dell'italianizzazione dei dialetti: appunti di creolistica casalinga. In Holtus, G., Kramer, J. & Schweickard, W. (Eds.), *Italica et Romanica. Festschrift für Max Pfister zum 65. Geburtstag*. Tübingen: Max Niemeyer Verlag, 13-30.

Berruto, G. (2004). Su restrizioni grammaticali nel codemixing e situazioni sociolinguistiche. Annotazioni in margine al modello MLF. In *Sociolinguistica*, 18, 54-72.

Berruto, G. (2005). Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In Auer, P., Hinskens, F. & Kerswill, P. (Eds.), *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 81-95.

Berruto, G. (2011). Interazione e compenetrabilità di grammatiche nel contatto linguistico. Il caso dell'enunciazione mistilingue. In Bombi, R., D'Agostino, M., Dal Negro, S. & Franceschini, R. (Eds.), *Lingue e culture in contatto. In memoria di Roberto Gusmani. Atti del 10° Congresso dell'Associazione Italiana di Linguistica Applicata*. Perugia: Guerra, 47-70.

Berruto, G. (2018). The languages and dialects of Italy. In Ayres-Bennett, W., Carruthers, J. (Eds.), *Manual of Romance Sociolinguistics*. Berlin: De Gruyter, 494-525.

Bertuccelli, G., Casotti, P. (1982). *Antona. Storia-Leggenda-Arte-Folklore-Architettura-Geografia-Tradizioni-Lavoro-Musica-Dialetto-Medicamenti.* Massa: Assessorato Pubblica Istruzione e Cultura.

Bonin, E. (1952). Beiträge zur Mundart und Volkskunde von Gorfigliano (Garfagnana) und Nachbarorte. Ph.D. Dissertation, University of Munich.

Bottiglioni, G. (1911). Dalla Magra al Frigido. Saggio fonetico. In *Revue de Dialectologie Romane*, 3, 77-143.

Bottiglioni, G. (1926). La penetrazione toscana e le regioni di Pomonte nei parlari di Corsica (Saggio di ricostruzione storico-linguistica). In *L'Italia Dialettale*, 3, 2-69.

Bottiglioni, G. (1930). Di alcune particolari tendenze fonetiche nell'italico (Continuazione e fine). In *Athenaeum. Studii periodici di letteratura e storia dell'antichità*, 8(1), 3-26.

Bresnan, J., Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. In *Language*, 86(1), 186-213.

Broersma, M. (2009). Triggered codeswitching between cognate languages. In *Bilingualism: Language and Cognition*, 12(4), 447-462.

Broersma, M., Carter, D. & Acheson, D.J. (2016). Cognate Costs in Bilingual Speech Production: Evidence from Language Switching. In *frontiers in Psychology*, 7, 1-16.

Broersma, M., Carter, D., Donnelly, K. & Konopka, A (2019). Triggered codeswitching: Lexical processing and conversational dynamics. In *Bilingualism: Language and Cognition*, ahead of print, 1-14.

Broersma, M., De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. In *Bilingualism: Language and Cognition*, 9(1), 1-13.

Brown, E.L. (2004). Reduction of syllable-initial /s/ in the Spanish of New Mexico and Southern Colorado: A usage-based approach. Ph.D. Dissertation, University of New Mexico.

Brown, E.L. (2013). Word classes in studies of phonological variation: Conditioning factors or epiphenomena?. In Howe, C., Blackwell, S. & Lubbers Quesada, M. (Eds.), *Selected Proceedings of the 15th Hispanic Linguistics Symposium*. Somerville: Cascadilla Proceedings Project, 179-186.

Brown, E.L. (2015). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. In *International Journal of Bilingualism*, 19(4), 387-406.

Brown, E.L. (2018). Cumulative exposure to phonetic reducing environments marks the lexicon: Spanish /d-/ words spoken in isolation. In Smith, K.A., Nordquist, D. (Eds.), *Functionalist and Usage-based Approaches to the Study of Language: In honor of Joan L. Bybee*. Amsterdam: John Benjamins, 127-153.

Brown, E.L., Amengual, M. (2015). Fine-grained and probabilistic cross-linguistic influence in the pronunciation of cognates: Evidence from corpus-based spontaneous conversation and experimentally elicited data. In *Studies in Hispanic and Lusophone Linguistics*, 8(1), 59-83.

Brown, E.L., Harper, D. (2009). Phonological Evidence of Interlingual Exemplar Connections. In *Studies in Hispanic and Lusophone Linguistics*, 2(2), 257-274.

Brown, E.L., Raymond, W.D. (2012). Explaining the distribution of Spanish f-/h- words. In *Diachronica*, 29(2), 139-161.

Bybee, J.L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J.L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. In *Language Variation and Change*, 14, 261-290.

Bybee, J.L. (2006). From usage to grammar. The mind's response to repetition. In *Language*, 82, 711-733.

Bybee, J.L., Beckner, C. (2009). Usage-based theory. In Heine, B., Narrog, H. (Eds), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 827-855.

Bybee, J.L., Eddington, D. (2006). A Usage-based Approach to Spanish Verbs of Becoming. In *Language*, 82(2), 323-355.

Calamai, S. (2017). Tuscan between standard and vernacular: a sociophonetic perspective. In Cerruti, M., Crocco, C. & Marzo, S. (Eds.), *Towards a New Standard. Theoretical and Empirical Studies on the Restandardization of Italian*. Boston: De Gruyter, 213-241.

Carpitelli, E. (1995). Description des systèmes des voyelles toniques de quelques dialectes de la Toscane nord-occidentale. In *Géolinguistique*, 6, 43-73.

Carpitelli, E. (2017). Esempi di variazione fonetica nell'Appennino settentrionale. Il caso dei vocalismi instabili. In Regueira, X.L., Fernández Rei, E. (Eds.), *Estudos sobre o cambio lingüístico no galego actual*. Santiago de Compostela: Consello da Cultura Galega, 215-240.

Carroll, J.B. (1971). Measurement properties of subjective magnitude estimates of word frequency. In *Journal of Verbal Learning and Verbal Behavior*, 10, 722-729.

Celata, C. (2006). Analisi dei processi di retroflessione delle liquide in area romanza con dati sperimentali dal còrso e dal siciliano. Ph.D. Dissertation, Scuola Normale Superiore di Pisa.

Cenceschi, S. Trivilini, A. Sbattella, L. & Tedesco, R. (2019). Collecting Italian spontaneous social media speech: the WAsp2 project. Poster presented at the *XV AISV Conference (Audio Archives at the Crossroads of Speech Sciences, Digital Humanities and Digital Heritage)*, Arezzo, February 14-16 2019.

Cerruti, M. (2016). L'italianizzazione dei dialetti: una rassegna. In *Quaderns d'Italià*, 21, 63-74.

Cerruti, M. (2018). Code-switching in Italo-Romance: a variationist study of convergence in bilingual speech. In *Lingue e Linguaggio*, 1/2018, 87-106.

Cerruti, M., Regis, R. (2005). 'Code switching' e teoria linguistica: la situazione italoromanza. In *Italian Journal of Linguistics*. *Rivista di Linguistica*, 17(1), 179-208.

Cho, T., Ladefoged, P. (1999). Variation and Universals in VOT: Evidence from 18 Languages. In *Journal of Phonetics*, 27, 207-229.

Clark, L. (2008). Re-examining vocalic variation in Scottish English: a Cognitive Grammar approach. In *Language Variation and Change*, 20(3), 255-273.

Clark, L., Watson, K. (2011). Testing claims of a usage-based phonology with Liverpool English t-to-r. In *English Language and Linguistics*, 15(3), 523-547.

Clopper, C.G., Mitsch, J.F. & Tamati, T.N. (2017). Effects of phonetic reduction and dialect variation on vowel production. In *Journal of Phonetics*, 60, 38-59.

Clyne, M. (1967). *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia.* The Hague: Martinus Nijhoff.

Coetzee, A.W., Kawahara, S. (2013). Frequency biases in phonological variation. In *Natural Language and Linguistic Theory*, 30(1), 47-89.

Coseriu, E. (1980)."Historische sprache" und "dialekt". In Göschel, J., Ivić, P. & Kehr, K. (Eds.), *Dialekt und dialektologie. Ergebnisse des Internationalen Symposions „Zur Theorie des Dialekts", Marburg/Lahn 5. - 10. September 1977.* Wiesbaden: Steiner, 106-122.

Dal Negro, S., Vietti, A. (2011). Italian and Italo-Romance Dialects. In *International Journal of the Sociology of Language*, 210, 71-92.

De Schryver, J., Neijt, A., Ghesquière, P. & Ernestus, M. (2008). Analogy, Frequency, and Sound Change. The Case of Dutch Devoicing. In *Journal of Germanic Linguistics*, 20(2), 159-195.

Della Sala, A.M. (1986). Caratteristiche morfologiche e sociali del linguaggio di Antona. In *Le Apuane. Rivista di cultura-storia-etnologia*, 11, 37-56.

Dinkin, A.J. (2008). The real effect of word frequency on phonetic variation. In *University of Pennsylvania Working Papers in Linguistics*, 14, 97-106.

Divjak, D. (2017). The Role of Lexical Frequency in the Acceptability of Syntactic Variants: Evidence From *that*-Clauses in Polish. In *Cognitive Science. A Multidisciplinary Journal*, 41(2), 354-382.

Dollinger, S. (2015). *The Written Questionnaire in Social Dialectology, History, theory, practice.* Amsterdam/Philadelphia: John Benjamins.

Drager, K. (2011). Sociophonetic variation and the lemma. In *Journal of Phonetics*, 39, 694-707.

Eddington, D., Channer, C. (2010). American English Has Goʔ A Loʔ Of Glottal Stops: Social Diffusion and Linguistic Motivation. In *American Speech*, 85(3), 338-351.

Forrest, J. (2017). The dynamic interaction between lexical and contextual frequency: A case study of (ING). In *Language Variation and Change*, 29, 129-156.

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. In *Laboratory Phonology*, 1(1), 5-39.

Ghimenton, A. (2013a). Quantitative Approaches to Italian / Dialect Family Interactions: Considerations of Methodology and Language Transmission in a Contact Situation. In *Journal of Language Contact*, 6, 106-133.

Ghimenton, A. (2013b). Pathways to multilingual acquisition in Veneto: a usage-based perspective of code choices in the input and output in a language contact situation. In Reif, M., Robinson, J. & Pütz, M. (Eds.), *Variation in Language and Language Use. Linguistic, Socio-Cultural and Cognitive Perspectives*. Bern: Peter Lang, 59-72.

Giannarelli, D. (1912). *Caratteri generali dei dialetti lunigianesi compresi fra la Magra e l'Appennino Reggiano.* Tortona: Tipografia Peila.

Giannelli, L. (2000). *Toscana*. Pisa: Pacini.

Giannelli, L., Savoia, L.M. (1979-1980). L'indebolimento consonantico in Toscana. In *Rivista Italiana di Dialettologia. Scuola società territorio*, 3-4, 38-101.

Habib, R. (2011). Frequency effects and lexical split in the use of [t] and [s] and [d] and [z] in the Syrian Arabic of Christian rural migrants. In *Journal of Historical Linguistics*, 1(1), 77-105.

Hamann, S.R. (2003). *The Phonetics and Phonology of Retroflexes.* Utrecht: LOT.

Hay, J., Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. In *Language*, 92(2), 298-330.

Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and /ay/: Lexical frequency, referee design, and style. In *Proceedings of the 14th international congress of phonetic sciences*, 1389-1392.

Hoshino, N., Kroll, J.F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script?. In *Cognition*, 106, 501-511.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. In *Journal of Experimental Psychology*, 48(2), 106-112.

ISTAT (2017). *L'uso della lingua italiana, dei dialetti e delle lingue straniere (anno 2015).* National survey.

Johnson, K. (1997). Speech perception without speaker normalization. An exemplar model. In Johnson, K., Mullennix, J.W. (Eds.), *Talker variability in speech processing*. San Diego: Academic Press, 145-165.

Kawahara, S. (2011). Aspects of Japanese loanword devoicing. In *Journal of East Asian Linguistics*, 20, 169-194.

Langacker, R. (1987). *Foundations of cognitive grammar. Vol. 1. Theoretical prerequisites.* Stanford: Stanford University Press.

Lee, O., Kapatsinski, V. (2014). Frequency Effects in Morphologisation of Korean /n/-Epenthesis. Paper presented as a poster at the *89th annual meeting of the Linguistic Society of America* (Portland, Oregon, January 2015).

Leivada, E., Papadopoulou, E., Kambanaros, M. & Grohmann, K.K. (2017). The Influence of Bilectalism and Non-standardization on the Perception of Native Grammatical Variants. In *frontiers in Psychology*, 8, 1-11.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W.J., Marchal, A. (Eds.), *Speech production and speech modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

Lisker, L., Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. In *Word*, 20(3), 527-565.

Loporcaro, M. (1997). Lengthening and raddoppiamento fonosintattico. In Maiden, M., Parry, M. (Eds.), *The Dialects of Italy*. London/New York: Routledge, 41-51.

Maffei Bellucci, P. (1977). *Lunigiana*. Pisa: Pacini.

Maiden, M., Parry, M. (Eds., 1997). *The Dialects of Italy*. London/New York: Routledge.

Mariotti Giromella, N. (2006). Preghiere in dialetto antonese. In *Le Apuane. Rivista di cultura-storia-etnologia*, 51, 140-144.

Mattheier, K.J. (1996). Varietätenkonvergenz: Überlegungen zu einem Baustein einer Theorie der Sprachvariation. In *Sociolinguistica*, 10, 31-52.

Mello, E., Miozzo, M. & Peressotti, F. (2017). Decision making in a not-so-foreign language: is a there a reduction of decision biases in Venetian dialect? Poster presented at the *XX European Society for Cognitive Psychology Conference (ESCOP)*, Postdam, Germany.

Merlo, C. (1956). Liguri o mediterranei gli odierni Apuani?. In *L'Italia Dialettale*, 20, 86.

Mioni, A.M. (1992). Il punto di vista della sociolinguistica. In Foresti, F., Sobrero, A.A. (Eds.), *Dialetti realtà ricerca. Atti del colloquio "I dialetti e la dialettologia negli anni Novanta"*. Bologna: CLUEB, 17-34.

Muldner, K., Hoiting, L., Sanger, L., Blumenfeld, L. & Toivonen, I. (2019). The phonetics of code-switched vowels. In *International Journal of Bilingualism*, 23(1), 37-52.

Myers, J., Li, Y. (2009). Lexical frequency effects in Taiwan Southern Min syllable contraction. In *Journal of Phonetics*, 37, 212-230.

Nadeu, M. (2016). Phonetic and phonological vowel reduction in Central Catalan. In *Journal of the International Phonetic Association*, 46(1), 33-60.

Navarrete, E., Arcara, G., Mondini, S. & Penolazzi, B. (2019). Italian norms and naming latencies for 357 high quality color images. In *PLoS ONE*, 14(2), 1-17.

Nusbaum, H.C., Pisoni, D.B. & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research in Speech Perception. Progress Report. Indiana University*, 10, 357-375.

Pellegrini, G.B. (1980^2). *Carta dei Dialetti d'Italia*. Pisa: Pacini.

Phillips, B. (2006). *Word frequency and lexical diffusion*. New York: Palgrave.

Pierrehumbert, J.B. (1994). Syllable structure and word structure: a study of triconsonantal clusters in English. In Keating, P.A. (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 168-188.

Pierrehumbert, J.B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J., Hopper, P. (Eds), *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins, 137-157.

Pierrehumbert, J.B. (2002). Word-specific phonetics. In Gussenhoven, C., Warner, N. (Eds.), *Laboratory phonology VII*. Berlin: Mouton de Gruyter, 101-140.

Raymond, W.D., Brown, E.L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In Gries, S., Divjak, D. (Eds.), *Frequency effects in language. Vol. 2. Learning and processing*. Berlin: Mouton de Gruyter, 35-52.

Raymond, W.D., Brown, E.L., & Healy, A.F. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. In *Language Variation and Change*, 28, 175-202.

Regis, R. (2013). Può un dialetto essere standard?. In *Vox Romanica*, 72, 151-169.

Regis, R. (2017). La nozione coseriana di dialetto e le sue implicazioni per l'area italoromanza. In *Revue Romane. Langue et littérature. International Journal of Romance Languages and Literatures*, 52(2), 137-169.

Ringeling, J.C.T. (1984). Subjective Estimations as a Useful Alternative to Word Frequency Counts. In *Interlanguage Studies Bulletin*, 8, 59-69.

Rossi, M. (1976). *Contribution à la méthodologie de l'analyse linguistique avec application à la description phonétique et phonologique du parler de Rossano (province de Massa, Italie): thèse* *presentée devant l'Université de Paris III le 10 juin 1974.* Lille/Paris: Atelier Reproduction des thèses / Librairie Honoré Champion.

Sanga, G. (1985). La convergenza linguistica. (Alla memoria di Benvenuto Terracini). In *Rivista Italiana di Dialettologia. Società scuola territorio*, 9, 7-41.

Savoia, L.M. (1980). Fonologia delle varietà apuane e garfagnine: consonantismo. In *Studi Urbinati di Storia, Filosofia e Letteratura*, 2, 233-293.

Schmid, H.J. (2000). *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin/New York: Mouton de Gruyter.

Schmitt, J.M., Auer, P. & Ferstl, E.C. (2019). Understanding fairy tales spoken in dialect: an fMRI study. In *Language, Cognition and Neuroscience*, 34(4), 440-456.

Schütze, C., Sprouse, J. (2013). Judgment data. In Podesva, R.J., Sharma, D. (Eds.), *Research methods in linguistics*. Cambridge: Cambridge University Press, 27-50.

Sobrero, A.A. (1996). Italianization and Variations in the Repertoire: The Koinai. In *Sociolinguistica. International Yearbook of European Sociolinguistics*, 10, 105-111.

Song, J., Dalola, A. (2019). Linguistic entrenchment and the effect of subjective lexical familiarity in Korean /n/-insertion. In *Proceedings of the Linguistic Society of America*, 4(34), 1-8.

Sorianello, P. (1996). Indici fonetici delle occlusive sorde in cosentino. In *Rivista Italiana di Dialettologia*, *Lingua dialetti società*, 20, 123-159.

Tabain, M., Butcher, A., Breen, G. & Beare, R. (2014). A spectral analysis of laterals in three Central Australian languages. In Hay, J., Parnell, E. (Eds.), *Proceedings of the 15th Australasian Speech Science and Technology Association*, 98-101.

Tagliamonte, S.A. (2013). Comparative Sociolinguistics. In Chambers, J.K., Schilling, N. (Eds.), *The Handbook of Language Variation and Change. Second Edition*. Boston: John Wiley & Sons, 128-156.

Thompson, G.L., Desrochers, A. (2009). Corroborating biased indicators: Global and local agreement among objective and subjective estimates of printed word frequency. In *Behavior Research Methods*, 41(2), 452-471.

Thorndike, E.G., Lorge, L. (1944). *A teacher's word book of 30,000 words*. New York: Columbia University Press.

Todd, S., Pierrehumbert, J.B. & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. In *Cognition*, 185, 1-20.

Toglia, M.P., Battig, W.F. (1978). *Handbook of semantic word norms*. Hillsdale: Erlbaum.

Trifone, P. (1992). *Roma e il Lazio*. Torino: UTET.

Wieling, M., Montemagni, S., Nerbonne, J. & Baayen, H. (2014). Lexical differences between Tuscan dialects and Standard Italian: A sociolinguistic analysis using generalized additive mixed modeling. In *Language*, 90(3), 669-692.

Yan, H. (2018). *The Nature of Variation in Tone Sandhi Patterns of Shanghai and Wuxi Wu*. Singapore: Springer/Peking University Press.

Yan, H., Zhang, J. (2017). The nature of variation in the tone Sandhi patterns of Wuxi Wu. In *The Journal of the Acoustical Society of America*, 142, 2519.

Zhang, J., Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. In *Journal of Phonetics*, 54, 169-201.

DOI: 10.17469/O2106AISV000020

DANIELE SANTINI, GIOVANNA LENOCI, GIOVANNA MAROTTA, PIER MARCO BERTINETTO

Repeated Vowels in Italian: An exploratory study

This paper addresses the articulatory realization of repeated identical vowels in Italian, such as they exist in root-internal position (e.g. *coorte* 'cohort', *Sahara* 'id.') or across a morpheme boundary (*atee* 'atheists.F', *coordinare* 'to coordinate'). In structural terms, these vowel sequences give rise to a hiatus, but their actual realization is variable. At fast speech rate, they may collapse into a single vowel also in terms of total duration. In careful speech, however, they retain a much longer duration than their single vowel counterparts (cf. *corte* 'court', *Sara* 'id.'). In addition, and crucially for our purpose, they may exhibit some hints at rearticulation at the boundary between the first and the second vowel. Some evidence that this may indeed occur was found by means of the UTI technique in a corpus of carefully produced speech.

Key words: repeated identical vowels, hiatus, Ultrasound Tongue Imaging (UTI), articulation.

1. *Introduction*

The Italian phonotactics admits quite a number of abutting vocoids, i.e. both true vowels and vocoidal approximants. In Italian, the difference between these two kinds of segments is not particularly large in terms of phonetics, but is quite remarkable vis-à-vis their function in syllable structure, where (for most theorists) only true vowels can be part of the nucleus. Regarding phonemic interpretation, one has to distinguish between hiatuses and diphthongs (or, more rarely, triphthongs, as in *continuiamo* /kon.ti.ˈnwja.mo/). The latter canonically involve the approximants /j w/, although not all sequences of a high vocoid and a vowel should be so analyzed. In particular, whenever a high vocoid is stressed, it must be a syllable nucleus, instead of an on/off-glide (e.g., *sfarfallio* 'flickering' and *bue* 'ox', with stress on, respectively, /i/ and /u/). Actually, as various authors have noted (Bertinetto, 1981: 158; Marotta, 1985; 1987; 1988; 2010a; 2010b; Bertinetto, Loporcaro, 2005), the decision about the actual phonemic interpretation can in some cases depend on speech rate, with faster rates favouring the diphthongization of a lento speech hiatus. This may create, for some words, a sort of incumbent ambiguity, because the speed parameter is inherently gradual, despite the all-or-none nature of the phonemic status. Indeed, the possible ambiguity between diphthong and hiatus has often been exploited in traditional Italian poetry, by means of the opposing devices of synaloepha and dieresis, respectively turning a hiatus into a diphthong, and a diphthong into a hiatus. The latter is, admittedly, a more artificial phenomenon, only pertaining to the domain of metrical performance, yet it is massively present in the versification of the major Italian poets. Consider, for example, the following two lines from Foscolo's sonnet *Alla sera* 'To the evening', which feature dieresis on *quiete* 'stillness' and synaloepha on *reo* 'guilty' and *lui* 'he/him':

Forse perché della fatal quïete (line 1)
questo reo tempo, e van con lui le torme (line 11)

It is important to observe that synaloepha can arise even in the absence of high vocoids (as in *reo* above), which would be the natural candidates to the role of vocoidal approximant. A special case of often ambiguous treatment can also be found at the boundary between a vowel-ending prefix and a vowel-initial root, where the propensity to give rise to a diphthong, once again irrespective of the nature of the vocoids concerned, depends (besides speech rate) on the degree of familiarity (Bertinetto, Gili Fivela, 1999: 134). Indeed, *coabitare* 'to coinhabit' has a higher chance to be considered a four-syllabic word than *coibentare* 'to insulate (w.r.t. temperature)', despite the presence of an unstressed high vocoid in the latter word.

Vocoidal sequences occur with considerable frequency in Italian. Marotta (1985) reports the results of a count based on the 1984 edition of the Ragazzini Dictionary of the Italian language, which yielded 17121 lemmas with two or more adjacent vocalic graphemes, disregarding purely orthographic (hence not phonemic) <i>. This corresponds to slightly more than 30% of the whole data-base used, consisting of 56359 entries. Salza (1986), as reported in Salza (1991), states that the Italian lexicon contains up to 167 different vocoidal combinations (106 bi-, 45 tri-, e 16 tetra-vocoidal combinations).

Needless to say, the various combinations of vowels occur with different frequency, and some of them are definitely dispreferred. In particular, and in agreement with a strong universal tendency, no Italian diphthong consists of homorganic vocoids (hence, */ji ij wu uw/). More generally, according to Chiari (2002: 221), the overall number of hiatuses in Italian is not very large. Despite this, the Italian lexicon admits the existence of hiatuses in a number of cases, of which Marotta (1987: 870) itemizes the most relevant ones. Such sequences can of course also arise postlexically at word boundaries, with variable phonetic consequences, such as elision or creation of (possibly non-canonical) diphthongs (Marotta, Sorianello, 1997).

The special topic addressed in this investigation is the phenomenon of abutting identical vowels, that we shall call 'repeated vowels', whose actual realization is a matter of performance variability. They can also be created postlexically at word boundaries (as in *buona amica* 'good.F friend.F'), often giving rise to elision. In this paper we concentrate, however, on word-internal repeated vowels, of the sort that one can find root-internally in words such as *Sahara* 'id.' and *coorte* 'cohort', or across a morphological boundary in words such as *atrii* 'entrance halls' and *atee* 'atheists.F'. The frequency of such sequences varies (Viviani, 2011):

– <aa> and <uu>[1] are definitely rare, with the latter one only found in words of Latin origin (*perpetuum*);

[1] In the first part of this paper we transcribe repeated middle vowels as orthographic sequences (by means of angled brackets), because their phonemic implementation can vary, both between words and across Italian dialects. But see §2.1 for further qualifications.

- <oo> is slightly more frequent, but this is mostly due to the word *zoo* 'id.' in its function as either prefixoid (*zoologico* 'zoological') or as member of a compound (*protozoo* 'protozoa'); in addition, it can be found in words containing the prefix *co-* (*cooperare* 'to cooperate', *cooptare* 'to coopt');
- <ee> is comparatively more frequent, considering the plural of feminine nouns ending in *-ea* (*maree* 'tides') or the feminine plural of adjectives with masculine singular in *-eo* (*atee* 'atheists.F');
- similarly, <ii> is mostly found in the plural of words ending in *-io*, provided <i> marks a true vocoid rather than being a mere orthographic marker of palatality, and to the extent that this sequence is not reduced, which is often the case in unstressed position (cf. *negozio/negozi* 'shop/-s', as opposed to *zio/zii* 'uncle/-s'); moreover, <ii> can occur in the preterite first person singular of third conjugation verbs (*partii* 'I left').

As the last examples show, stress location can have a strong impact on the phonetic realization of a repeated vowels sequence. Stress on the first <i> of *zii* and *partii* contributes to hiatus preservation, whereas lack of (primary) stress on <oo> in *zoologico* and *cooperare* may pave the way, at very fast rate, to reduction to a single vowel, which is indeed the normal fate of *negozio/negozi* 'shop/-s' and similar words (*monopolio* 'monopoly', *cappio* 'noose' etc.).

When a sequence of identical vocoids consists of true vowels, it gives rise, structurally speaking, to a hiatus, i.e. a disyllabic sequence. Actually, Valesio (1967) interpreted such sequences as phonologically long vowels, but to our knowledge this position has not been shared by any later study, since the absence of phonemic vowel quantity in the Italian phonology is unanimously assumed as a solid fact. But although the phonological interpretation of identical repeated vowels is not an issue, their phonetic realization remains a poorly investigated topic. As far as we know, this is the first study addressing it from the articulatory point of view.

In a nutshell, the research questions that we set to ourselves were:
RQ1: Do repeated vowels differ in terms of duration vis-à-vis the corresponding single vowels (at least in carefully produced speech)?
RQ2: Are there any hints of a sort of 'rearticulation' at the boundary between two repeated vowels?
RQ3: Can an Ultrasound Tongue Imaging (henceforth, UTI) analysis effectively enhance our knowledge about sequences of identical vowels?

2. *Method*

2.1 Corpus

As precondition of this work, a list of words containing repeated vowels (thus, excluding diphthongs) was built, with each word contrasted with a similar one presenting a single vowel in the same (or at least very similar) phonotactic environment (e.g.: *Sahara* 'id.' ~ *Sara* 'id.', *coorte* 'cohort' ~ *corte* 'court', *ree* 'guilty.F.PL' ~ *re* 'king'). This allows one to compare the acoustic and articulatory production of /VV/ ~/V/ (where /VV/ should be read, here and in the remaining of this paper, as /V_xV_x/). In addition, with respect to stress position, the list contained three types of /VV/ hiatus: Unstressed (N 10), Stressed on V1 (N 10), and Stressed on V2 (N 8). Table 1 reports the whole corpus. Since all target sequences consist of true vowels, from now on we will place them within slashes to underline their phonemic status.

It was initially decided to analyze one pair for each type in Table 1. Where we had to make a choice, we gave precedence to the most frequent and familiar words (for example, *sciita* ~ *gita* was preferred to *piissimo* ~ *Trissino*). In just one case we selected two pairs (*veemente* ~ *demente*, *linee* ~ *trine*) because their frequency and familiarity was considered roughly comparable. Among these initially selected items, we took into further consideration only those in which the comparative evolution of the articulatory gesture for the single and the repeated vowels presented significant differences. These were detected through visual inspection of the splines generated by the UTI equipment; all authors took part in this procedure. We thus selected the following 13 pairs: *corte* ~ *coorte* (speakers 2, 3, 6), *là* ~ *Laa* (speakers 2, 3, 4, 6), *re* ~ *ree* (speakers 2, 4, 6), *spinetta* ~ *lineetta* (speakers 4, 6), *atri* ~ *patrii* (speaker 6), *numero* ~ *duumviro* (speakers 4, 6). Considering that each pair was produced by 6 speakers, our selection consisted of 13% of the initial corpus. Due to space limitations, in this paper we will only present the pairs *là* ~ *Laa* (speaker 2); *re* ~ *ree* (speaker 4); *atri* ~ *patrii* (speaker 6); *corte* ~ *coorte* (speaker 3); *numero* ~ *duumviro* (speaker 4).

2.2 Participants

The list of stimuli was read by 6 participants, all born and raised in northern Italy. This choice was made because speakers of central Italy distinguish the quality of stressed middle vowels according to etymological criteria, as e.g. in *ree* [ˈrɛe] 'guilty.F.PL' or *ninfee* [niɱˈfɛːe] 'water lilies', whereas for speakers of northern Italy (except the Veneto area) the quality of such vowels merely depends on phonotactic reasons and, crucially for our purposes, does not vary between the stressed and the unstressed vowel in the /ˈVV/ and /VˈV/ sequences of our corpus. This precaution allows us to confidently use, in the remaining of this paper, a trustworthy phonemic representation also for the middle vowels /e o/. The participants (5 females and 1 male) were from Lombardia (3), Piemonte (2) and Liguria (1). Their ages at the time of the recordings were in the range 21 to 24 years. They all had university education; they had a clear pronunciation and no one presented heavy dialect features. Since, however, the performance of two of them was judged not perfectly satisfactory on all tokens produced, the analyses were only carried out on the 4 best speakers.

The subjects were instructed to read the whole list three times at comfortable pace, with no emphasis whatsoever. They also read twice the same target words as included in short sentences, although the articulatory analysis of repeated *vs* single vowels in sentence context has not yet been carried out. Hence, the duration data reported below for /VV/ ~/V/ refer to isolated words (see § 3.1).

Table 1 - *Experimental list of words contrasting repeated (left) vs single vowels (right), divided by stress configuration*

	UNSTRESSED HIATUS			UNSTRESSED VOWEL
/ii/	*patrii* 'pertaining to motherland (M.PL)', *arbitrii* 'arbitrary acts'	*vs*	/i/	*atri* 'entrance halls', *ladri* 'thieves'
/ee/	*linee* 'lines', *veemente* 'vehement'	*vs*	/e/	*trine* 'ornamental laces', *demente* 'demented'
/aa/	*sahariano* 'Saharian', *Nausicaa* '(person name)'	*vs*	/a/	*Mariano* '(person name)', *Marica* '(person name)'
/oo/	*cooptare* 'to coopt', *coordinare* 'to coordinate', *Alcinoo* '(person name)'	*vs*	/o/	*lottare* 'to fight', *scorporare* 'to extract'
/uu/	*duumvirato* 'duumvirate'	*vs*	/u/	*duplicato* 'duplicate(d)'
	STRESSED HIATUS ON V1			**STRESSED VOWEL**
/ii/	*sentii* 'I felt', *addii* 'farewells', *pii* 'pious.M.PL', *trii* 'trios', *rinvii* 'deferments'	*vs*	/i/	*sentì* 's/he felt', *addì* 'on the date', *pipì* 'pee'
/ee/	*ree* 'guiltiy.F.PL', *cee* 'eel's newborns', *dee* 'goddesses'	*vs*	/e/	*re* 'king', *c'è* 'there is'
/aa/	*Laa* '(person name)'	*vs*	/a/	*là* 'there'
/oo/	*Coo* '(person name)'	*vs*	/o/	*Po* '(river name)'
/uu/	...	*vs*	/u/	...
	STRESSED HIATUS ON V2			**STRESSED VOWEL**
/ii/	*piissimo* 'very pious', *sciita* 'Shiite'	*vs*	/i/	*Trissino* '(person name)', *gita* 'excursion'
/ee/	*lineetta* 'small line'	*vs*	/e/	*spinetta* 'spinet', *Linetta* 'little Lina (person name)'
/aa/	*Sahara* 'id.'	*vs*	/a/	*Sara* 'id.'
/oo/	*coorte* 'cohort', (io) *coopto* 'I coopt', *Laocoonte* '(person name)'	*vs*	/o/	*corte* 'court', *copto* 'coptic', arconte 'archon'
/uu/	*duumviro* 'duumvir'	*vs*	/u/	*numero* 'number'

2.3 Experimental Procedure

The recordings were carried out in the anechoic booth of the ELiTe section of the SMART lab of Scuola Normale Superiore, Pisa. Prior to that, the participants were familiarized with the task and the experimental procedure. They were assisted to wear a helmet (Articulate Stabilisation Headset, produced by Articulate Instruments Ltd., 2008) to which a micro-convex ultrasound probe was attached (Mindray probe 65EC10EA). This guaranteed controlled distance and orientation of the probe with respect to the palate of the speaker independently of head movements. The probe was held under the chin and yielded a mid-sagittal configuration of the tongue (Fig. 1).

The ultrasound images generated by a Mindray UTI system were acquired at a frequency rate of 60 MHz and synchronized with the audio signal through a synchronization unit (Synch Bright-up unit). The speakers were asked to read the list of the stimuli, with three repetitions of each item in pseudo-random order. The UTI recording session lasted approximately 40 mins and all subjects successfully completed the task. The ultrasound data were recorded and analyzed by the Articulated Assistant Advanced software (AAA; Articulate Instruments Ltd., 2007).

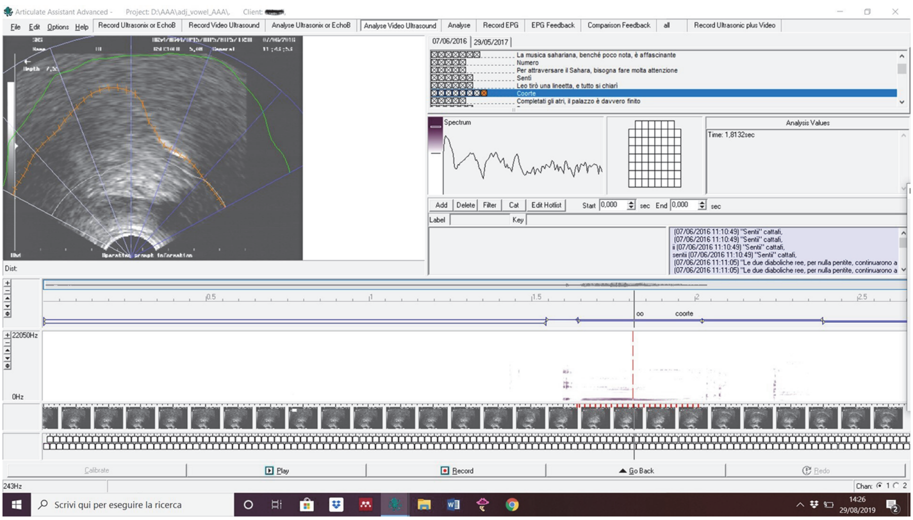
2.4 Analysis

For each participant and each /VV/~/V/ sequence, all items were first auditorily inspected for correct pronunciation. No production had to be rejected.

First, by means of the Praat facility (Boersma, Weenink, 2001), all target vowels were segmented on the acoustic signal and labelled appropriately. Conventional manual segmentation (Salza, 1990) criteria were used to measure vowel duration, taking into account the first and last glottal pulse of the relevant intervals.

The annotations were then imported into the AAA software for articulatory processing. At each sampling point within the acoustic interval of the target vowels, a semi-automatic tongue contour (splining), with subsequent manual correction, was fitted to the UTI images. The splines were plotted into the AAA workspace for qualitative analysis of the dynamic evolution of the tongue gesture. Vowel durations were computed by means of the AAA 'Export data' function.

Figure 1 - *AAA window showing, at the bottom, the spectrogram with phonetic labelling and segmentation of an /oo/ sequence; at the top, the ultrasound frame (and the corresponding spline) of the tongue at the selected UTI frame, as produced by a single speaker*



3. *Results*

The research questions of the present work, as itemized at the end of §1, converge on the goal of finding out whether there is any difference in the phonetic realization of repeated *vs* single vowels, at least in carefully produced laboratory speech. To this purpose, two kinds of analyses were designed:

1. an acoustic analysis of the vowel interval durations (§ 5.1);
2. a qualitative examination of UTI images (§ 5.2), to find out whether there are any hints at a sort of 'rearticulation' within the repeated vowels interval.

3.1 Measure of durations

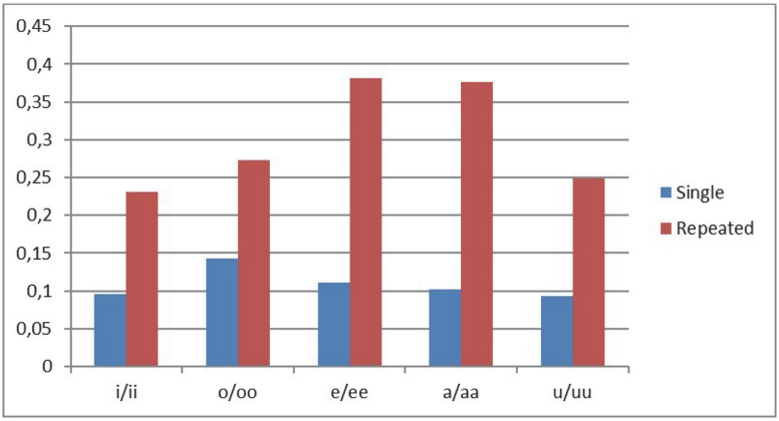
For each speaker, we calculated the mean vowel duration (in ms) of the three repetitions of each isolated item. To allow the comparisons between /VV/~ /V/, the acoustic data of all participants were pooled together in the analysis (Figure 2) and compared within each single *vs* repeated vowel pair (Table 2). An anonymous reviewer suggested to us to normalize the individual absolute durations. However, considering the small number of observations and, most importantly, the consistent convergence of the individual behaviours, we prefer to leave this move to a future expansion of this investigation.

Table 2 - *Number of tokens (3 repetitions for each of the 4 selected speakers), mean and standard deviation of /V/ and /VV/ durations (in ms) for each vowel pair. P-value indicate the significance level of the t-test*

/V/ ~ /VV/	N	Mean (ms)	*p-value*
/a/	12	0.102 ± 0.035	
/aa/	12	0.378 ± 0.081	***
/e/	12	0.105 ± 0.034	
/ee/	12	0.381 ± 0.058	***
/i/	12	0.096 ± 0.045	
/ii/	12	0.231 ± 0.055	***
/u/	12	0.094 ± 0.029	
/uu/	12	0.249 ± 0.028	***
/o/	12	0.143 ± 0.037	
/oo/	12	0.272 ± 0.030	***

With the partial exception of /oo/ ~/o/, the average duration of the single vowels is less than half the average duration of their repeated counterpart (Fig. 2). The t-test analyses show that these differences are statistically significant ($p < .001$).

Figure 2 - *Compared average duration for single (blu) and repeated vowels (red)*



This result is confirmed by the analysis performed on the same items as produced by the same speakers in sentence context: the average /V/ ~ /VV/ ratio across all vowels and speakers is 0,54 in sentence context, as against 0,38 in isolated words. Hence, although the duration contrast between repeated *vs* single vowels is predictably less sharp in sentence context, it is nevertheless quite robust in the corpus used.

3.2 Articulatory analyses

In this section we present the results of the qualitative inspection of the selected UTI images collected from the individual speakers. For space reasons, we only report a few representative graphs of the comparisons between the /V/ ~ /VV/ tongue profiles. Since the UTI and audio signals were synchronized, the tongue profiles follow the temporal evolution of the gesture: the blue and green lines refer, respectively, to the initial and final phase of the articulation, while the red lines refer to the middle phase. Furthermore, the blue and green thick dotted lines indicate the absolute first and absolute last profile in the vowel interval. The respective number of profiles differs for single *vs* repeated vowels, because the /V/ interval is of course significantly shorter. However, in order to make the visual inspection easier, in the repeated vowels figures we only reported one spline every second one. For this reason, the distance between two successive profiles is 16 ms in the /V/ figures, and 32 ms in the /VV/ ones. To help the reader in the inspection of the actual gestural evolution, we added numbers in the repeated vowel figures to mirror the progression of the actual splines (only odd numbers are shown, because we report one profile every second one).

Figure 3 presents the tongue splines for the pair /e/ ~ /ee/ in the words *re* 'king' (top) and *ree* guilty.F.PL' (bottom). As can be seen, the tongue tip and body remains approximately stable during the entire articulatory gesture of the single vowel (*re*), while the root shows progressively less peripherality, although the displacement is relatively small. The overall gesture of the repeated vowels (*ree*) appears to be more mobile and, above all, it exhibits detectable signs of rearticulation. While the tongue tip gesture shows progressive centralization, the body expands at first its peripherality, but then turns to a slightly more centralized position; similarly, and most significantly, the tongue root reaches its most centralized position in the middle portion of the /VV/ interval to then exhibit a kind of backward movement towards the end, witnessing a tendential rearticulation of the gesture.

Figure 3 - *Tongue splines (with tongue tip to the right of the images and tongue root to the left) for the single vowel /e/ (top) and the repeated vowels /ee/ (bottom) as produced by speaker 4. The numbers in the repeated vowel figure, indicate the temporal progression of the tongue gestures. See the beginning of § 3.2 for further details*

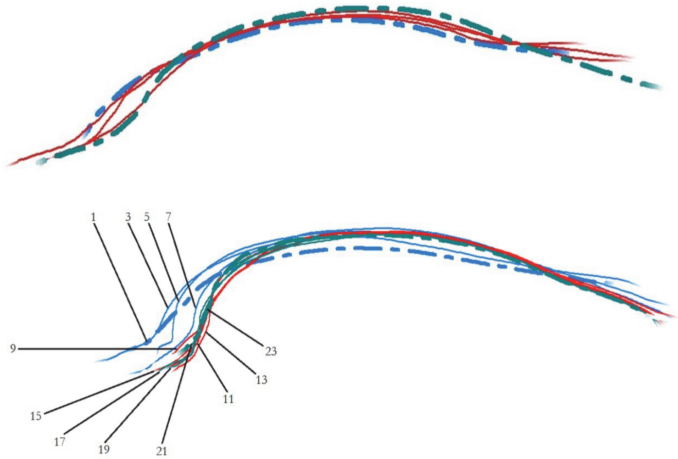


Figure 4 compares the tongue dynamics for /a/ ~ /aa/ in *là* 'there' (top) and *Laa '(person name)'* (bottom). The tongue gesture for the single vowel appears to be quite dynamic: the tongue tip is initially higher, due to coarticulatory effects, to lower and remain stable throughout the middle and final part of the gesture, while the root, and even more the body, rise to a peripheral position. In the repeated vowels, the tongue tip follows the same pattern as the single vowel, again owing to coarticulatory reasons, while the body and the root are very dynamic and show some hint at rearticulation. The body starts from a lower position, travels towards the peripheral space in the central phase of the gesture and then slightly moves back during the final phase; the root also starts from a more central position and reaches topmost peripherality during the central phase, to then return to the initial position in the final phase.

Figure 4 - *Tongue splines (tongue tip to the right) for the single vowel /ɑ/ (top) and the repeated vowels /ɑɑ/ (bottom) as produced by speaker 2*

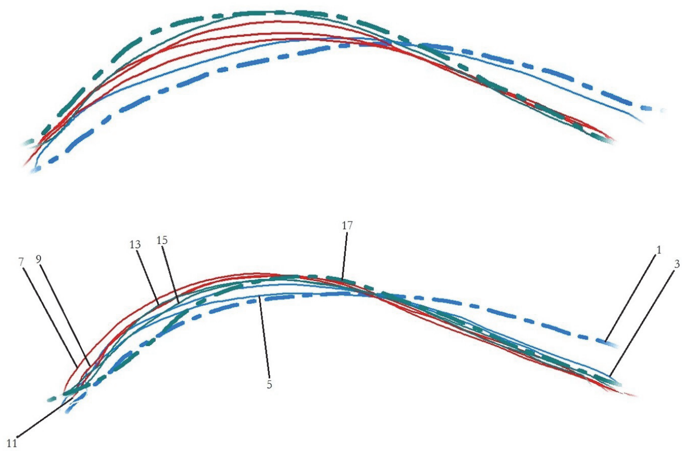


Figure 5 compares the tongue dynamics for the pair /i/ ~ /ii/ in *atri* 'entrance halls' (top) and *patrii* 'pertaining to motherland (M.PL)' (bottom). For the single vowel, a limited tongue movement can be observed in the root, with more peripherality in the initial phase, which then decreases as the gesture unfolds. As for the body and tip, they remain relatively stable during the whole gesture. A similar trend can be observed for the repeated vowels counterpart: the tongue root shows the highest dynamism, moving from a more peripheral position to a less peripheral one in the middle of the interval, to then rise again at the end (unlike its single counterpart). The body and tip of the tongue show a rather stable pattern during the whole gesture, with however a slightly higher degree of peripherality in the middle of the interval, which confirms the tendency towards rearticulation.

Figure 5 - *Tongue splines (tongue tip to the right) for the single vowel /i/ (top) and the repeated vowels /ii/ (bottom) as produced by speaker 6*

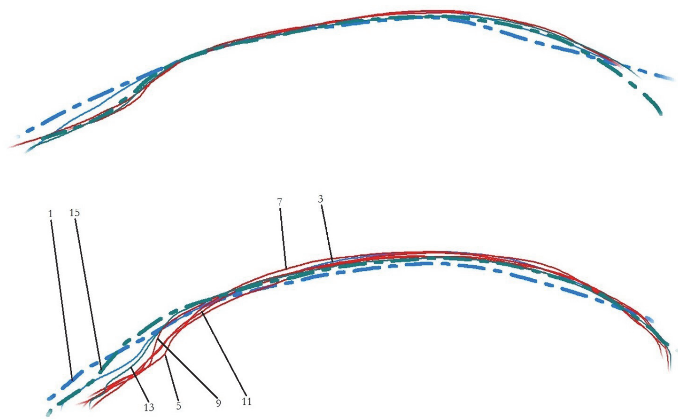
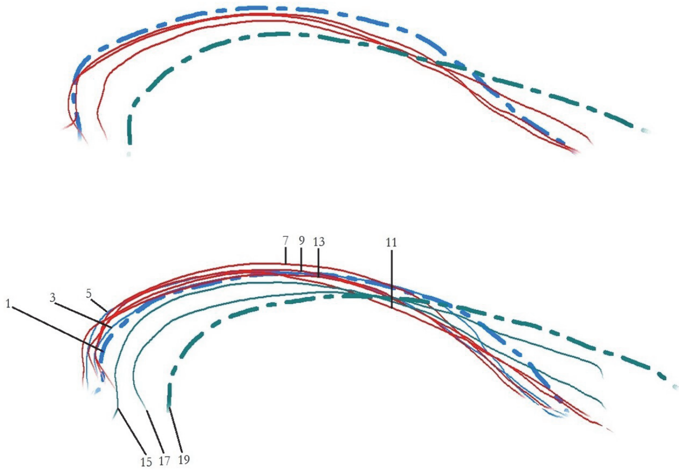


Figure 6, comparing the tongue dynamics for the pair /o/ ~ /oo/ in *corte* 'court' (top) and *coorte* 'cohort' (bottom), offers the most striking example of gestural rearticulation. With the single vowel /o/, the tip of the tongue shows a stable movement from less to more peripheral, coupled with a symmetrically opposite movement of the body and the root. With the repeated vowels, the gestural dynamics exhibits a similarly opposite movement of the tip as opposed to body and root, but in addition it shows a remarkable rearticulation in the middle phase of the gesture, which overcomes the absolute initial position (see the red central profiles as contrasted with the blue dotted line).

Figure 6 - *Tongue splines (tongue tip to the right) for the single vowel /o/ (top) and the repeated vowels /oo/ (bottom) as produced by speaker 3*



4. *Discussion and conclusion*

The durational data reported in § 3.1 provide a positive answer to the first research question (RQ1) announced at the beginning of this paper. Similarly, the data reported in § 3.2 add a positive answer to RQ2, lending support to the hypothesis that the repetition of two adjacent identical vowels may give rise, in Italian, to a detectable gestural rearticulation, approximately occurring in the middle portion of the whole vowel interval. When this occurs, the speaker adjusts the articulators so as to partially restart the gesture. We can thus safely state that, at least in careful pronunciation, the gestural dynamics underlying the production of Italian repeated identical vowels should not be interpreted as the mere prolongation of the corresponding single vowel gesture.

Needless to say, not all production tokens are equally explicit in this respect and this lack of systematicity invites some caution. First of all, the corpus used for our investigation consists of carefully produced speech. At faster rates, and above all in hypoarticulated speech, it is unlikely that the same results would arise. Besides, even in our carefully articulated speech not all instances showed the same amount of rearticulation. As observed in § 2.1, the most remarkable instances amounted to a mere 13% of the produced tokens. Although this should not be intended in the sense that 87% of the tokens exhibited no evidence thereof, it is nevertheless apparent that not all rearticulation gestures were equally notable. Nevertheless, since the one at stake is a gradual phenomenon, we surmise that, by means of more refined methods of investigations, the articulatory contrast underlying /VV/~/V/ may be effectively measured in a much larger proportion of cases. What one has to consider, at any rate, is the fact that the articulatory difference between repeated and single vowels should not be seen in terms of dynamic *vs* static articulation, but rather in the presence *vs* absence of a specific type of articulatory dynamics in the middle portion of the vowel interval. The extent to which this behaviour is typical of the Italian speakers, as opposed to speakers of other languages, is a matter to which we cannot provide an answer at the present stage, in the absence of specifically designed contrastive studies.

A further matter of caution concerns the variable phonotactic structure of our experimental items, with regard to both the preceding and the following consonant (or consonant cluster). The only way to overcome this problem would consist in using a set of non-words, with strictly controlled phonotactics. However, this would introduce a high degree of artificiality, which would deplete the very essence of our research enterprise. As is well-known, real words give rise to deep-rooted articulatory routines, unlikely to be activated by non-words. This is especially the case for frequent words, as some influential authors have shown (see at least Bybee, Hopper, 2001 and Pierrehumbert, 2002).

The limited goal of this investigation was, at any rate, to prove that the gesture underlying repeated vowels should not be understood as a mere prolongation of the corresponding single vowel gesture, and from this point of view the results can be seen as quite conclusive. Although mere gestural prolongation cannot be excluded in all instances, especially so in fast speech, we found solid evidence that Italian speakers show some tendency to produce two successive vowel gestures, rather than simply keeping a single gesture for a longer duration. However, we prefer to speak of 'rearticulation' instead of referring to two separate articulatory targets, because all we can detect is a partial gestural reorientation, rather than two successive gestures.

Since the UTI technique proved successful in detecting observable differences between /VV/ and /V/ words, we can conclude that also RQ3 received a positive answer. However, one should observe that the experimental technique used here only allows inspection of the supralaryngeal activity. Further research should investigate what happens at the laryngeal level, where specific adjustments might occur. We hope to be able to address this issue in future work by synchronizing UTI with laryngographic analysis.

In addition, it will also be advisable to extend the investigation to words uttered in (semi-)spontaneous speech. Finally, it is necessary to investigate the role of stress. As the reported examples show, stress may fall on the first vowel (e.g. *ree*), on the second one (e.g. *coorte*), or on neither (e.g. *atrii*). This may have consequences on the overall gestural evolution, and indeed we have some initial evidence that this might be the case. Since, however, we could only compare different vowels in different phonotactic and prosodic positions, for the time being we cannot draw any solid generalization.

Note, finally, that when the second repeated vowel coincides with a morpheme marker (as in *re-e* 'guilty-F.PL'), there might be a stronger tendency to protect its articulatory salience, as compared to morpheme-internal /VV/ sequences. Ideally, within a larger corpus, also based on more speakers and more repetitions per speaker, one could make use of mixed statistical models to pinpoint the individual contribution of the various factors involved.

Bibliography

Articulate Instruments LTD (2007). *Articulate assistant advanced user guide: Version 2.07.* Edinburgh, UK: Author.

Articulate Instruments LTD (2008). *Ultrasound stabilisation headset user's manual: Revision 1.13.* Edinburgh, UK: Author.

Bertinetto, P.M. (1981). *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici.* Florence: Accademia della Crusca.

Bertinetto, P.M., Gili Fivela, B. (1999). Incontri vocalici tra prefisso e radice (iato o dittongo?). In *Archivio glottologico italiano*, 84, 129-172.

Bertinetto, P.M., Loporcato, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. In *Journal of the International Phonetics Association*, 35, 131-151.

Bybee, J., Hopper, P.J. (Eds.) (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Boersma, P., Weenik, D. (2001). PRAAT, a system for doing phonetics by computer. In *Glot International*, 5, 341-345.

Chiari, I. (2002). *Ridondanza e linguaggio. Un principio costitutivo delle lingue, prefazione di T. De Mauro.* Rome: Carocci.

Marotta, G. (1985). Syllabification rules. Discrimination between hiatus and diphthong. In *Esprit Project*, 64.

Marotta, G. (1987). Dittongo e iato in Italiano: una difficile discriminazione. In *Annali della Scuola Normale Superiore di Pisa*, 17(3), serie iii, 847-887.

Marotta, G. (1988). The Italian diphthongs and the autosegmental framework. In Bertinetto, P.M., Loporcato, M. (Eds.), *Certamen Phonologicum. Papers from the 1987 Cortona phonology meeting.* Turin: Rosenberg & Sellier, 389-420.

Marotta, G. (2010a). Dittongo. In *Enciclopedia dell'italiano online, Treccani:* http://www.treccani.it/enciclopedia/dittongo_(Enciclopedia-dell'Italiano).

Marotta, G. (2010b). Iato. In *Enciclopedia dell'italiano online*, *Treccani*: http://www.treccani.it/enciclopedia/iato_(Enciclopedia-dell'Italiano).

Marotta, G., Sorianello, P. (1997). Vocali contigue a confine di parola, Unità fonetiche e fonologiche. Produzione e percezione. In Bertinetto, P.M., Cioni, L. (Eds.), *Atti delle VIII giornate di studio del Gruppo di fonetica sperimentale (AIA)*. Pisa: Scuola Normale Superiore, 101-113.

Pierrehumbert, J. (2002). Word-specific phonetics. In *Laboratory Phonology,* VII, 101-139.

Salza, P.L. (1986). La durata dei suoni nelle nelle sequenze vocaliche dell'italiano. In *Elettronica e Telecomunicazioni*, 35(1), 27-34.

Salza, P.L. (1990). La problematica della segmentazione del segnale vocale. In *Atti della 1^ Giornata di Studio di G.F.S.* Padua, 3-6 Novembre 1990, 23-48.

Salza, P.L. (1991). Misura elettroacustica della durata segmentale in dittongo e iato dell'Italiano, Parte I-II. In *Rivista italiana di acustica*, 15, 3-19, 35-55.

Valesio, P. (1967). *Strutture dell'allitterazione*. Bologna: Zanichelli.

Viviani, A. (2011). Doppie lettere [prontuario]. In Simone, R., con la collaborazione di Berruto, G. (Eds.), *Enciclopedia dell'italiano*. Roma: Treccani. http://www.treccani.it/enciclopedia/doppie-lettere-prontuario_%28Enciclopedia-dell%27Italiano%29/.

Acknowledgments

The authors wish to thank Silvia Calamai for initial assistance in building the experimental corpus, as well as Chiara Bertini and Irene Ricci for indispensable assistance in the use of the lab equipments.

DOI: 10.17469/O2106AISV000021

MARTINA TURCONI, GIOVANNA MAROTTA

Speech and music rhythm: A pilot study on English and Italian songs[1]

Instrumental and vocal compositions agree with the mother tongue rhythm of the composer, proving that the same rhythmic categorization operates on both linguistic and music levels. In this study we aimed to verify how strong the mother tongue rhythm of the singers is in relation to the mother tongue rhythm of the composers, and if this is also able to influence the music rhythm of their songs. The results of the analysis of original English and Italian songs along with their foreign versions suggest that on the music level the *stress-timed* rhythm exerted a greater influence than the *syllable-timed* one, being able to influence the music rhythm of the Italian version of the English songs. On the contrary, on sung level both timing categories seem to be equally strong.

Key words: rhythm, sung speech, music rhythm, timing.

1. *Preliminaries*

The notion of "rhythm" is intrinsic to several areas of natural phenomena and human activity (e.g. wave sounds, the flow of the tide, and all those human activities that make use of rhythmic pattern, such as speaking and singing, but also playing music, dancing, ice-skating and so on). The English term is derived from the Greek word ῥυθμός (*rhythmós*), cognate with ῥέω (*réo*), which means "to flow". It was originally used by Ionic authors to mean "schema", whereas in the Attic age, Plato gave it a new meaning: "rhythm is order in movement" (Law, II book: 664e-665a). This definition immediately evokes a sense of periodicity, hierarchic regularity, and dynamic recurrences of durations. However, "rhythm is more than mere clock time, the invariant sequence of evenly spaced intervals, and yet of such regularity is rhythm born" (Cummins, 2015: 158). It is in music that the concept of rhythm in this sense, is most at home. In fact, a stable underlying beat is almost always present in music, and this allows for the organization of periodicity on multiple time scales in order to create musical meter (see Patel, Daniele, 2003a) and, undoubtedly, speech also involves dynamic recurrences of durations. Are the principles that drive rhythm in speech and music the same, or do they change?

This question has given rise to a number of studies that have aimed to uncover the underlying principles of rhythmical manifestations. One such principle is that both in music and linguistics, rhythm refers to patterns of duration (see London,

[1] Even though the article has been jointly discussed, the authorship of the paragraphs is intended to be divided as follows: M. Turconi § 1, 3, 4.1, 4.2, 5; G. Marotta § 2, 4.3, 6.

2001). Furthermore, each domain shares the rhythmic configurations of the *grouping structure*: both tones and words are grouped into phrases (see Selkirk, 1984; Nespor, Vogel, 1986, 2007), in "a 'configurational' patterning of stressed vs. unstressed syllables" (Patel, Daniele, 2003a: 140).

In the last decades of the Eighteen century, Joshua Steele (1775) applied the musical theory to the analysis of the English language, in an attempt to explore the common belief that English rhythm tends to form sequences of regular beats: "there is a general tendency to make the stress – points of stressed syllables follow each other at equal intervals of time, but [...] this general tendency is constantly interfered with by the variations in the number and nature of the sounds between successive stress – points" (Steele, 1775). The first person to attempt to empirically prove this rhythmical intuition was Classé (1939). Using a kymograph, he was able to represent the intensity wave form of English speech. The results demonstrated a tendency to isochronism and Classé hypothesized that the perceptive system might also play a key role in the perception of unequal intervals as being even. A few years later, Pike (1945) published an investigation on the intonation of American English, where he identified two rhythmic dichotomous units: *stress-timed* and *syllable-timed* rhythmic units. This pattern of temporal organization was then extended to all the natural languages by Abercrombie (1967). A third rhythmical category was later postulated: the *mora-timed* class. This class takes as its rhythmic unit the mora, which recurs in isochronous intervals in languages such as Japanese (Bloch, 1950; Han, 1962; Ladefoged, 1975; Nespor, Shukla & Mehler, 2011).

In the last part of the XX century, several research studies were conducted to confirm this timing classification. Lehiste (1977), Dauer (1983), and Bertinetto (1989) to name but a few, proved through various methods of analysis that the postulated isochronism did not correspond to reality, but that it is the perceptive mechanism that influences us and leads us to state that the speech rhythm units occur in isochronous intervals. Lehiste (1977) demonstrated that there is a perceptive threshold, that ranges from 30-100 milliseconds, within which a difference in duration can be perceived. If the interval is shorter than 30 milliseconds, it cannot be discriminated, and it can be considered equivalent to either a larger or smaller interval. Isochrony is, after all, a purely perceptual phenomenon. Later, Bertinetto (1989), who analyzed two Italian speakers, and Dauer (1983), who measured interstress intervals from readings of texts in English, Thai, Spanish, Italian, and Greek, both argued that it might not be possible to categorize the languages into two opposite and discrete rhythmical categories, but that they might rather be distributed along a continuum, whose poles are the pure stress-timed language and the pure syllable-timed language, which, actually, do not exist. "In ordinary speech and everyday prose [...in fact] the rhythmic effect is a purely automatic consequence of linguistic circumstances" (Dauer, 1983: 60).

The linguistic scenario drastically changed at the turn of the last century. Nazzi, Bertoncini, and Mehler (1998) had proven that newborns could discriminate two languages if they belonged to two different rhythmic classes (for example they could discriminate English from Japanese, but not English from Dutch). Given those results, Ramus, Nespor & Mehler (1999) reviewed the timing model of classification, stating that feet, syllables, and morae were not the most suitable rhythmic units, but that languages could be rhythmically classified according to the comparison of the proportion of vocalic intervals within a sentence (*%V*), and the standard deviation of the vocalic and intervocalic intervals within a sentence (*ΔV* and *ΔC)*. Indeed, through these metrics the authors could demonstrate that stress-timed languages were characterized by more complex syllable structures and greater vowel reduction than syllable-timed ones.

Later on, other researchers proposed new coefficients in order to normalize the speech rates not only between different languages, but also different speakers of the same language, as with the *Pairwise Variability Index* of Grabe and Low (2002), or the *varco* coefficient of variation. This latter variation was applied by Dellwo, Wagner (2003) and Dellwo (2006) to improve Ramus' coefficients. In their paper, Grabe & Low (2002) took into consideration many more languages than ever before and the results proved the continuum hypothesis of Dauer (1983) and Bertinetto (1989). In fact, many of these languages, such as Greek, Welsh, and Polish, could not be classified either as stress-timed (e.g. English, German or Dutch) or syllable-timed (e.g. Romance languages) because of their 'mix' character.

A few years later, Bertinetto, Bertini (2008) proposed a new model of rhythmic classification capable to capture the rhythmic tendencies of natural languages: the *Control/Compensation Hypothesis*. Based on the consideration that all languages have syllables and prominence contrasts, the authors agreed that what makes the difference, in rhythmic terms, is the prosodic structure built on this common ground. Indeed, the *Control/Compensation Index* (*CCI*) "aimed at describing the intra-syllabic behavior, which in turn affects (or is possibly affected by) the overarching accentual alternation" (Bertinetto, Bertini, 2008: 427).

Even though new recent perceptual experiments have confirmed the rhythmic classification of languages based on quantitative measurement methods (e.g. Hannon, 2009), part of the scientific community pointed out not only the inability of all metrics to successfully and unequivocally classify the rhythmically non-prototypical languages, such as Greek and Polish (see Grabe, Low, 2003), but also that metrics "are very sensitive to the effects that methodological choices have on the durations of consonantal and vocalic intervals" (Arvaniti, 2012: 354) and that the inter-speaker variability in timing is more extensive than previously suspected, mainly in the spontaneous speech. Thus, rethinking the validity both of metrics and rhythm theories, Tilsen, Arvaniti (2013) initiated a different approach to analyse speech rhythm that was not dependent on syllable structure. Using the technique known as *empirical mode decomposition* (*EMD*, see Huang, Shen, Long, Wu, Shih, Zheng, Yen, Tung & Liu, 1998)[2], a spectral analysis of the envelope of speech waveform, they attempted to identify rhythmic stretches of speech and compare rhythmic patterns across languages. The envelope metrics[3] the au-

[2] "This procedure extracts non-orthogonal basis functions from the envelope, each of which captures oscillations on a different time-scale" (Tilsen, Arvaniti, 2013: 629).

[3] Tilsen and Arvaniti (2013) elaborated three types of envelope metrics: the *power distribution metrics* are able to capture the relative power in syllabic vs. supra-syllabic oscillations in the envelope; the *rate*

thors used seem to be able to dissociate rate from power distributions, capturing different information from amplitude envelopes, and to exhibit weak dependence on syllable structure, being less dependent on syllable phonotactics than interval metrics. For these reasons Tilsen, Arvaniti (2013) judge their metrics to be an alternative approach to describe speech rhythm. Moreover, the investigation of other acoustic dimensions, such as pitch other than duration, could help in understanding the cognitive mechanisms of rhythm in speech production and perception.

Also, Brow and colleagues (2017) argue that the durational coefficients describe just the surface level of rhythm organization, not providing enough information about the relative duration of the syllables: "they are not able to specify the actual rhythm of any given utterance within a language" (Brown, Pfordresher & Chow, 2017: 97). In order to fill these gaps, Brown and colleagues (2017) proposed an effective but upgradeable musical model capable to predict the rhythmic representations of a sentence. In contrast with the previous studies, the authors recovered one of the characteristics of musical meter to analyse speech rhythm: although meter often remain stable across periods of time, it can occasionally change creating *heterometric sentences* (sentences where a change in focus-word changes the sentence rhythm)[4], which could well represent the complexity of speech timing along with *isodurational* (stable meter and invariant syllable durations, such as the nursery rhyme "Twinkle Twinkle little star. How I wonder what you are") and *isometric sentences* (stable meter but variable timing of syllables within measures, such as the sentence "Pamela purchased beautiful flowers Saturday morning all through the year", in which there is an alternation between 3-syllable and 2-syllable groupings). Comparing the timing of *prominence groups*[5] between an a priori rhythmic transcriptions of a set of sentences and their spoken versions by calculating the *coefficient of variation* (*CV*)[6], the authors found out that "most of the transcriptions were borne out by the productions, suggesting that metricality in speech can be measured reliably" (Brown et al., 2017: 108). In fact, musical transcriptions, thanks to its metrical structure, can be adapted to all the different realizations of a given sentence.

2. *Rhythm in music and speech*

At the very same time, some scholars returned to the roots of the timing debate, rediscovering the powerful link that connects speech and music. In fact, the quantitative measurement methods used in the linguistic research on rhythm were successfully applied by Patel, Daniele (2003a) to the analysis of musical compositions,

metrics, such as w_1 and w_2, register the frequencies of those oscillations; and the *rhythmic stability metrics* capture the stability of oscillations.

[4] For example, in the sentence "Miguel bought two yellow shirts at the men's store by the bay" either "yellow" or "two" could be accented, changing the meter of the sentence.

[5] The authors' notion of prominence group is similar to the concept of an 'inter-stress interval' that can be found in previous research on speech rhythm, such as Dauer (1983), Cummins, Port (1998).

[6] The *coefficient of variation* is the Standard Deviation of normalized prominence groups (Brown et al., 2017).

in order to verify if the mother tongue rhythm of a composer could influence the music rhythm of their compositions. They considered the *Pairwise Variability Index* (by Grabe, Low, 2002) the most suitable variable for musical corpora, and, using this, were able to quantify the degree to which successive units differ in duration. Therefore, to analyze music rhythm variability, the unit to be taken into consideration should be the note. After having computed all the variables on their instrumental music corpus[7] and having compared the results with the data of English and French speech by Ramus (2002), they were able to affirm that the mother tongue rhythm of the composers overlapped with the music rhythm of their compositions.

This pioneering study generated an interesting debate as well as a new branch of studies which aimed to prove these findings in extended corpora (see Huron, Ollen 2003; Patel, Daniele, 2003b; and London, Jones, 2011). In recent years, research has addressed new linguistic areas (e.g. Polish, see Jekiel, 2014) and new musical genres such as blues (Larroque, 2012), popular music and Japanese vocal music (Sadakata, Desian, Honing, Patel & Iversen, 2004), which all appear to confirm Patel, Daniele (2003) hypothesis.

Our study proceeds along the path traced by the studies of Patel, Daniele (2003) and fellow musicologists (Huron, Ollen, 2003; Patel, Daniele, 2003b; London, Jones, 2011) as well as linguistic studies on speech rhythm (Jekiel, 2014; Larroque, 2012; Sadakata et al., 2004). Firstly, we aim to prove that, even in modern vocal compositions, the mother tongue rhythm of the composer influences the music rhythm of their songs. Moreover, we want to verify how strong the mother tongue rhythm of the singers is in relation to the mother tongue rhythm of the composers, and if this is also able to influence the music rhythm of their songs. To do so we decided to resort to a particular type of vocal composition: original songs that had been rearranged in a foreign language version. Therefore, we considered also to attempt to analyse the rhythm of the same song sung in two different language types. Indeed, using the linguistic quantitative measurement methods proposed by Grabe, Low (2002), and Dellwo, Wagner (2003), we were able to analyse the *sung speech rhythm*, which can be considered as the timespace where music and speech rhythm are fully involved in each other[8]. Assuming that general rhythmic principles drive the rhythmic motor behavior of human beings on multiple levels, and since it has been demonstrated that there is a "significant phase-locked neural response to rhythm in both music and speech domains" (Harding, Sammler, Henry, Large & Kotz, 2019: 98), we argue that the rhythm of sung speech could also be classified along with the traditional rhythmic continuum, as it was postulated by Dauer (1983) (see §1).

[7] Patel and Daniele (2003a) selected English and French compositions written by author lived between the Nineteenth- and Twentieth-century. English composers: A. Bax, F. Delius, J. Ireland, E. Elgar, R.V. Williams and G. Holst; French composers: C. Debussy, V. D'Indy, G. Fauré, A. Honneger, J. Ibert, D. Milhaud, F. Poulenc, M. Ravel, A. Roussel and C. Saint-Saëns.

[8] During the second part of the XX century some of the characteristics of the singing voice were investigated, e.g. melody structure by Sundberg, Lindblom (1976), vowel quality and consonants palatalization by Oussilova (1995), and spectral modifications of French vowels by Florig (1995).

3. *Corpus*

Focusing on a type of musical composition that exhibits a robust relationship between speech and music rhythm (i.e. songs), we decided to analyze vocal music as opposed to instrumental music for its tight relationship between speech and music rhythm. The choice of the language was guided by the aim of having two languages with different timing classifications. Therefore, we opted for American English, because it has been traditionally classified as a stress-timed language, and Italian, a syllable-timed language. Regarding the music period, we chose songs from the 60s and the 70s, which allowed us to select original songs along with their foreign version. In those years the practice of creating a new original version of a well-known foreign song, while maintaining an almost unaltered melodic line, was very popular. Hence, we chose three Italian and three English songs and their corresponding foreign version. The list is presented in Table 1:

Table 1 - *Italian and English songs and their corresponding version*

Original Italian songs	**English versions**
Azzurro – Adriano Celentano (1968)	*Blue skies (Azzurro)* – Ken Dodd (1970)
Mi ritorni in mente – Lucio Battisti (1969)	*Wake me, I am dreaming* – The Love Affair (1971)
Balla Linda – Lucio Battisti (1968)	*Bella Linda* – The Grass Roots (1969)
Original English songs	**Italian versions**
Stand by me – Ben E. King (1961)	*Pregherò* – Adriano Celentano (1962)
I'm a believer – Niel Diamond (1966)	*Sono bugiarda* – Caterina Caselli (1967)
Elenore – The Turtles (1968)	*Scende la pioggia* – Gianni Morandi (1968)

All the songs were performed in an "Andante" tempo, neither too fast nor too slow, and the performance didn't overly alter the temporal organization of the songs.

It was easy to find a number of recordings of all the songs of our *corpus*, so we selected the best audio track from the ones available[9]. On the contrary, it was difficult to source a printed score for some of the songs, so we decided to transcribe the missing scores from the recording ourselves. We also noticed that some official tran-

[9] Italian songs: L. Battisti, *Mi ritorni in mente*, "*RememberingItaly*", track 15, © 2016 Codex, 2016; L. Battisti, *Balla Linda*, "*I Successi di Lucio Battisti*", track 8, © 1998 DV More, 2009; C. Caselli, *Sono bugiarda*, "*Qualcuno mi può giudicare*", track 8, © 1997 Sugar Srl, 2006; A. Celentano, *Azzurro*, "*…Adriano*", disk 3, track 11, © 2013 Clan Celentano Srl, 2013; A. Celentano, *Pregherò*, "*Milva, Mina & More from Italy*", track 5, © 2013 OVC Media Limited, 2013; G. Morandi, *Scende la pioggia*, "*C'era un ragazzo*", track 11, 2002. English songs: N. Diamond, *I'm a believer*, "*The Bang Years 1966-1968*", track 7, Universal Music Italia Srl, 2011; K. Dodd, *Blue skies*, "*All the songs I love*", disk 2, track 7, Parlophone Records Ltd, 2005; B.E. King, *Stand by me*, "*Stand by me*", track 1, The Web Engine, 2016; The Grass Roots, *Bella Linda*, "*Greatest Hits: The Grass Roots*", track 9, © 2012 One Media iP Ltd, 2012; The Love Affair, *Wake me, I am dreaming*, https://www.youtube.com/watch?v=2T9pmRzRPl4 [21/03/2019]; The Turtles, *Elenore*, "*The turtles present the Battle of the Bands*", track 3, © 1968 Flo & Eddie, Inc, 2005.

scriptions, such as "Balla Linda" by Lucio Battisti, adopted the common normalized version for the "swing" style: instead of writing the song in triplets, as it should be performed, the transcribers wrote it in a binary beat division. Therefore, we had to create a new written version using a ternary beat division so that the rhythm of the score adhered, as far as possible, to the rhythm of our audio track[10].

4. *Analysis of the music rhythm*

4.1 Method

Following the method proposed by Grabe, Low (2002) and adopted by Patel, Daniele (2003a), we considered the *Pairwise Variability Index* (PVI)[11] to be the most suitable variable to analyze a musical corpus. According to Patel, Daniele (2003a), in fact, the value of a note is easily comparable to a vocal interval. Given that, we decided to calculate the *normalized PVI* (*nPVI*) (see Figure 1) of the melodic lines from the scores to stay as much as possible in line with the method previously used by the authors who conducted similar studies (Patel, Daniele, 2003a and London, Jones, 2011, among others), and in order to compare their results with ours, especially for the English values. Since the scores played such an important role in our analysis, we decided to rewrite the ones that couldn't reflect the exact rhythm of the acoustic performance (see § 2).

Figure 1 - *Normalized Pairwise Variability Index (Grabe, Low, 2002)*

$$n\,PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

The variability index was computed for the first verse and the first refrain/chorus, where it was possible. Some foreign versions didn't use the entire original song: "Wake me I am dreaming" was composed only on the musical line of the verse of "Mi ritorni in mente"[12]; others were performed so freely compared with the scores, that we were forced to discard those parts of the score that was rhythmically too far from the audio track.

We calculated the *nPVI* assigning an arbitrary duration of 1 to the first note of each selected theme; consequently, the durations of the remaining notes were expressed as multiples or submultiples of the value of the first note. When a rest was part of the theme (in this musical genre it is a common characteristic), we didn't

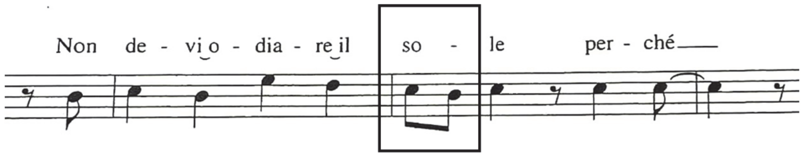
[10] See *Appendix*.

[11] The *Pairwise Variability Index* expresses the level of durational variability in successive measurements. The *normalized* version of *PVI* "is compiled by calculating the difference in duration between each pair of successive measurements, taking the absolute value of the difference and dividing it by the mean duration of the pair" (Grabe, Low, 2002: 523).

[12] See *Appendix*.

take into consideration the value of the rest in the algorithm (as in London, Jones 2011). On the contrary we did take into consideration the durations of all the notes making up the music line, even the ones belonging to the so called "melismatic configurations" (see Johnson, Huron & Collister, 2014)[13]. An example of such melismatic configuration is given in Figure 2: the syllable "so-" of the word "sole" in the song "Pregherò" by Adriano Celentano.

Figure 2 - *Example of a melismatic configuration*

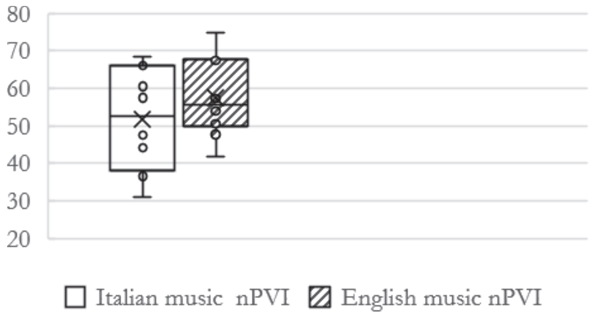


Lastly, to statistically prove our results, we carried out a *t-test analysis*.

4.2 Results

Figure 3 presents the box and whisker plot of the data collected: the average, the median and the range of the measurement of the Italian and English songs. Despite the wide-ranging results, it seems that there is a difference between Italian and English values.

Figure 3 - *Musical* nPVI *values for Italian (in white) and English songs (in black stripes)*



But how reliable is this account? The descriptive statistics of Table 2 as long as the conducted *t-test* ($t(16.577) = -1.0713$, $p = 0.299$) do not prove such a difference between Italian and English music *nPVI*.

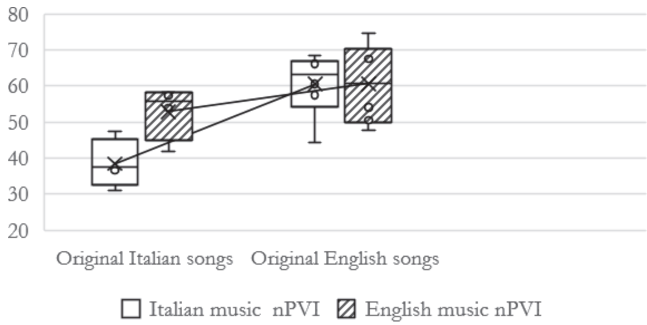
[13] In vocal compositions there are two types of note configuration: melismatic configurations and syllabic ones. In this second type, one single note is assigned one syllable, whereas in the melismatic configuration, each syllable can be sung as two or more notes.

Table 2 - *Mean, median and standard deviation of the Italian and English music nPVI, and their respective minimum and maximum values*

	Italian music nPVI	English music nPVI
Mean	51.68	57.48
Median	52.47	55.71
Std. Deviation	13.79	10.20
Minimum	31.04	41.97
Maximum	68.49	74.77

Looking in more depth at the data, the *nPVI* of the original Italian songs shows a noticeable difference from the original English songs *nPVI* (see Figure 4). But, if we look at the two groups, "Original Italian songs" and "Original English songs", we can see a different trend. The Italian versions of the original English songs do not show, as expected to be, a syllable-timed rhythm: the *nPVI* values (distributed in the white box at the right of Figure 4) suggest that they are carachterised by a stress-timed rhythm, as their English counterparts. Indeed, the difference between Italian orginal English songs and English original English songs is not statistically proven ($t(9.5609) = -0.00911$, $p = 0.992$), whereas it is proven the difference between the Italian original Italian songs and the Italian original English songs ($t(7.7364) = -4.399$, $p = 0.002$). On the other side, the English versions of the original Italian songs do not follow the syllable timing of Italian, mantaining their native stress-timed speech rhythm ($t(5.9447) = -2.8403$, $p = 0.029$).

Figure 4 - *The box and whisker plot shows the* nPVI *results for the first subgroup "Original Italian songs" on the left and the "Original English songs" on the right*



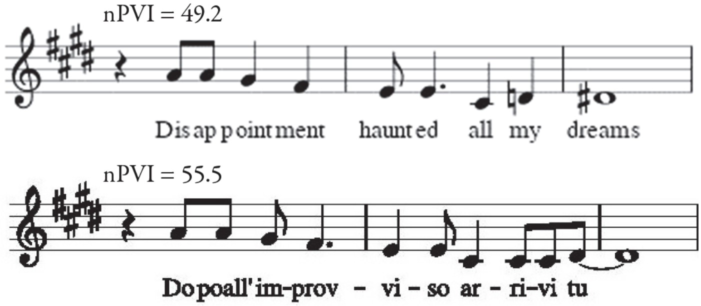
4.3 Discussion

The data collected allow us to state that, in our corpus, the rhythm of the English songs tends towards the stress-timed typology, because of the higher *nPVI* values. The situation regarding the Italian song rhythms, however, is much more complicated. In fact, with regard to the original Italian compositions (mean 38.45; Standard Deviation 6.8), the music rhythm tends towards the syllable-timed typology, but it does not do so for the Italian versions of the original English songs (mean 60.5;

Standard Deviation 8.9). It seems that only the English mother tongue rhythm is able to influence the rhythmic structure of the melodic line, thus making it more variable. Conversely, the Italian speech rhythm does not manage to reduce the variability of the rhythmic pattern, actually making it, at times, even more complex.

An example of this increased complexity can be found in the Italian version "Sono bugiarda" of the original English song "I'm a believer" by Niel Diamond, in which not only the rhythmic melodic structure is changed, but also the musical and linguistic metric one. Musically, the rhythmic structure of the English verse has a lower *nPVI*, and therefore the rhythm is less variable. In Figure 5 we present the verse "Disappointment haunted all my dreams" from the original English song "I'm a believer" by Niel Diamond at the top, followed by the verse "Dopo all'improvviso arrivi tu" from its Italian version "Sono bugiarda" by Caterina Caselli.

Figure 5 - *Comparison between two verses*



In the metric structure, the second accent falls on the beat in the English verse, whereas in the Italian one, it is anticipated on the upbeat through a syncope, which greatly alters the metric structure:

– Italian verse: /do.poal.lim.prov.ˈ**vi**.zo.ar.ri.vi.ˈ**tu**/
– English verse: /di.sə.pɔɪnt.mənt.ˈ**hɔ:n**.tɪd.ɔ:l.maɪ.ˈ**dri:mz**/[14]

5. *Analysis of sung speech rhythm*

The speech data regarding English speech and Romance languages (Ramus et al., 1999; Grabe, Low, 2003; Dellwo, Wagner, 2003) and the music data (Patel, Daniele 2003a) suggest that the same rhythmic categories operate not only on the linguistic level, but also on the musical one. Our data also support the hypothesis that the mother tongue rhythm of the composer almost always overlaps with the music rhythm of the composition.

What happens when the speech rhythm interacts with the music rhythm? How does speech rhythm behave when it is sung? Does the mother tongue rhythm of

[14] The accented syllables are written in bold letters, and the syllables where the second musical accent falls is underlined.

the singer overlap with the music rhythm of the songs? We have attempted to answer these questions by analysing the sung speech rhythm of our corpus. In order to analyze the rhythm of the sung speech we adopted two variables: the *normalize Pairwise Variability Index* by Grabe, Low (2003) and the *varcoΔV* and *varcoΔC* by Dellwo, Wagner (2003)[15]. These coefficients allowed us to normalize the tempo variability of the songs. In actual fact, even though we chose only songs, whose mood was "Andante", the term refers not to an exact speed rate, but to a range of rates.

After isolating the sentences of each song first verse and refrain with the software *Audacity*, we segmented the sonogram in vocalic and intervocalic intervals[16] using *Praat*. Then, we computed the variables with the software *Correlatore* and processed a statistical analysis to identify any possible outlier. Afterwards, we remove all the outliers to obtain a less scattered data. Lastly, to statistically prove our results, we carried out a *t-test analysis*.

5.1 Results and discussion

The results obtained for both the variables show a similar trend; but, only the results concerning the *nPVI* coefficients show a statistically significant difference between Italian and English rhythm (*vocalic nPVI*: $t(58.632) = -2.9077$, $p = 0.005$; *intervocalic nPVI*: $t(59.839) = 2.2641$, $p = 0.027$)[17]. The vocalic intervals are more variable for English sung speech, whereas the intervocalic intervals seem not to be subject to such a variation. Indeed, we have observed in various cases that the syllable structure of English sung speech changed, enduring reduction and deletion phenomena of consonants, mostly at the end of the phrases and the sentences. This could be an explanation of the low intervocalic values of English songs (see Table 3 and Table 4). With such preliminary overview, our hypothesis seems to be corroborated by our *nPVI* data. However, given that we took into consideration a corpus not adhering to stringent criteria such as the same number of syllables for each sentence, further analysis on syllable structure as well as syntactic and morphological analyses are needed.

Lastly, through in-depth analysis of the data, there is no trace of the contrast between the original and the foreign versions that was evident in the musical analysis[18].

---

[15] The *varco* coefficient is the percentage of ΔC of the mean value for consonantal intervals and the percentage of ΔV of the mean value for vocalic intervals (Dellwo, Wagner, 2003; Dellwo, 2006).

[16] To segment the vocalic and intervocalic intervals we followed the definition of Grabe, Low (2002: 520): "Vocalic intervals were defined as the stretch of signal between vowel onset and vowel offset, characterised by vowel formants, regardless of the number of vowels included in the section (a vocalic section could contain a monophthong, a diphthong, or, in some cases, two or more vowels spanning the offset of one word and the onset of the next). Intervocalic intervals were defined as the stretch of signal between vowel offset and vowel onset, regardless of the number of consonants included."

[17] The differences between English and Italian *varcoΔV* ($t(59.018) = -1.1$, $p = 0.275$) and *varcoΔC* ($t(53.535) = 1.2854$, $p = 0.207$) are not statistically proven.

[18] Data of the vocalic *nPVI* computed for the four groups: mean of the vocalic *nPVI* of original Italian songs: 57.90; mean of the vocalic *nPVI* of the English versions of the original Italian songs: 74.89; mean of the vocalic *nPVI* of the original English songs: 69.09; mean of the vocalic *nPVI* of the Italian versions of the original English songs: 61.00.

In fact, the hypothesized configuration occurs both for the originals and the transpositions; for example, as it can be seen in Figure 6, the English vocalic *nPVI* is higher than the Italian one in the group of "Original Italian songs" ($t(15.784) = -2.8194$, $p = 0.012$), whereas there is no significative difference between English and Italian vocalic *nPVI* of the group of "Original English songs" ($t(31.617) = -1.4258$, $p = 0.163$), as it was for the musical data (see § 4.2). Moreover, *t-test* analyses confirm that there is not a significative difference neither between Italian vocalic *nPVI*s ($t(24.441) = 0.63665$, $p = 0.5303$) nor between English vocalic *nPVI*s ($t(21.597) = 0.86575$, $p = 0.3961$). Consequently, English and Italian sung speech rhythm remain stable regardless the origin of the musical theme.

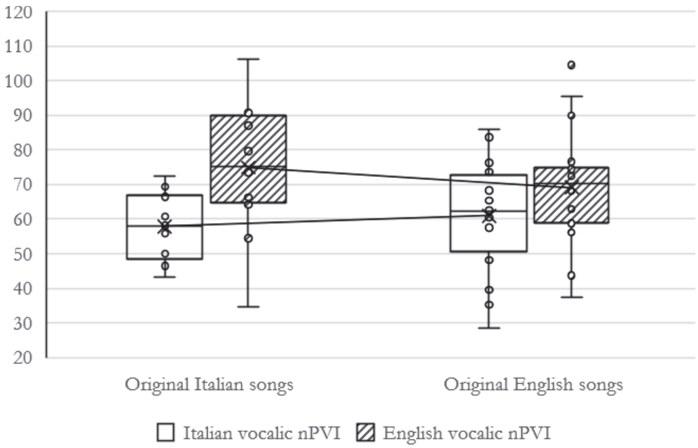
Table 3 - *Mean, median and standard deviation of the Italian and English vocalic and intervocalic nPVI, and their respective minimum and maximum values*

	Italian voc nPVI	**English voc nPVI**	**Italian interv nPVI**	**English interv nPVI**
Valid	29	30	30	28
Missing	2	1	1	3
Mean	59.61	71.41	67.90	55.88
Mediana	60.49	73.19	67.14	55.95
Std. dev.	13.57	17.47	19.98	15.62
Minimum	28.45	34.68	36.64	29.90
Maximum	86.03	106.3	107.6	85.39

Table 4 - *Mean, median and standard deviation of the Italian and English varcoΔV and varcoΔC coefficients, and their respective minimum and maximum values*

	Italian varcoΔV	**English varcoΔV**	**Italian varcoΔC**	**English varcoΔC**
Valid	29	30	30	29
Missing	2	1	1	2
Mean	61.67	69.63	56.26	52.68
Mediana	59.77	66.74	57.45	54.72
Std. Dev.	16.97	22.04	16.43	15.01
Minimum	30.74	28.08	29.04	22.79
Maximum	94.58	117.7	93.39	82.60

Figure 6 - *The box and whisker plot shows the vocalic nPVI results for the first subgroup "Original Italian songs" on the left and the "Original English songs" on the right*



6. *Conclusion*

Although our research might be considered a pilot study on sung speech rhythm, the data collected suggest that the mother tongue of the composer, as well as that of the singer, is mostly coherent with the rhythm of their music and songs. Italian versions of original English songs are in some way different from the original Italian songs inasmuch they retrieve the English rhythm. We know American music has exerted a greater influence on Italian composers of the last century, not allowing them to completely maintain their mother tongue rhythm in the compositions. On the contrary, this influence does not extend to sung speech rhythm, which mostly adhere with speech rhythm. However, larger corpora are needed in order to confirm this hypothesis, and an in-depth analysis of sung speech is required to understand which principles drive rhythm in songs and how sung speech behaves.

A question remains unanswered: is rhythm a prosodic property of speech which can be measured in the signal in simple numerical terms? "The quantitative algorithms do not allow us to relate acoustic data to phonological constraints, although they are widely used because of their explicitness and predictive value" (Marotta 2012). Moreover, Ramus (1999), Grabe, Low (2002), and Dellwo, Wagner (2003) rhythm metrics reflect only the duration and the clustering of segments, but we know that the complexity of the syllable structure and the reduction of unstressed vowels are fundamental to the perception of speech rhythm (see Dauer, 1983; Gil, 1986; Nespor et al., 2011). Besides, recent works pointed out the multidimensional complexity of the perception of rhythm, requiring a multimodal analysis of speech, but also of music, signal (Cummins, 2009).

In conclusion, speech rhythm as well as music rhythm could be epiphenomena driven by more general motor behaviors of human beings, that are biologically constrained.

Acknowledgment

We would like to thank Dr. Alessandro Cecchi, researcher in musicology, for his tireless support and assistance throughout this research.

Bibliography

Abercrombie, D. (1965). *Studies in Phonetics & Linguistics*. London: Oxford University Press, 16-25.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. In *Journal of Phonetics*, 40, 351-373.

Bertinetto, P.M. (1989). Reflections on the Dicotomy "Stress- vs. Syllable-timing". In *Revue de Phónetique Appliquée*, 91/93, 99-130.

Bertinetto, P.M., Bertini, C. (2008). On modelling the rhythm of natural languages. In *Proceedings of the 4th Speech Prosody Conference*, University of Campinas, 427-430.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. In *Language*, 26, 86-125.

Brown, S., Pfordresher, P.Q., Chow, I. (2017). A musical model of speech rhythm. In *Psychomusicology: Music, Mind, and Brain*, 27, 2, 95-112.

Cummins, F. (2009). Rhythm as an Affordance for the Entrainment of Movement. In *Phonetica*, 66, 15-28.

Cummins, F. (2015). Rhythm and speech. In Redford, M.A. (Ed.), *The Handbook of Speech Production*. Wiley, 158-177.

Cummins, F., Port, R. (1998). Rhythmic constraints of stress timing in English. In *Journal of Phonetics*, 26, 145-171.

Dauer, R.M. (1983). Stress-timing and syllable-timing. In *Journal of Phonetics*, 11, 51-62.

Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. In Karnowski, P., Szigeti, I. (Eds.), *Language and Language-processing*. Frankfurt am Meim: Peter Lang, 231-241.

Dellwo, V., Wagner, P. (2003). Relations between language rhythm and speech rate. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of 15th International Congress of Phonetic Sciences*, 471-474.

Florig, E. (1995). An acoustic study of French vowels in speech and singing voice. In Elenius, K., Branderud, P. (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences. ICPhS 1995. Stockholm, Sweden. 13-19 August 1995, vol. 1*. Stockholm: The Congress Organisers at KTH and Stockholm University, 210-213.

Gill, D. (1986). A prosodic typology of language. In *Folia Linguistica*, 20, 165-231.

Grabe, E., Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In *Papers in Laboratory Phonology*, 7. Cambridge: Cambridge University Press, 515-546.

Han, M.S. (1962). The feature of duration in Japanese. In *Onsei no kenkyuu*, 10, 65-80.

Hannon, E.E. (2009). Perceiving speech rhythm in music: Listener classify instrumental songs according to language of origin. In *Cognition*, 111, 403-409.

Harding, E.E., Sammler, D., Henry, M.J., Large, E.W. & Kotz, S.A. (2019). Cortical tracking of rhythm in music and speech. In *NeuroImage*, 185, 96-101.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C. & Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proc. R. Soc. London*, Ser. A 454, 903-995.

Huron, D., Ollen, J. (2003). Agogic contrast in French and English themes: further support for Patel and Daniele (2003). In *Music Perception: An Interdisciplinary Journal*, 21, 2, 267-271.

Jekiel, M. (2014). Comparing rhythm in speech and music: the case of English and Polish. In *Yearbook of the Poznań Linguistic Meeting* 1. De Gruyter Open, 55-71.

Johnson, R., Huron, D. & Collister, L. (2014). Music and lyrics interaction and their influence on recognition of sung words: an investigation of word frequency, rhyme, metric stress, vocal timbre, melisma and repetition priming. In *Empirical Musicology Review*, 9, 1, 2-20.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Larroque, P. (2012). The Rhythm of English and Blues Music. In *American International Journal of Contemporary Research*, 2, 5, 123-136.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lehiste, I. (1977). Isochrony reconsidered. In *Journal of Phonetics*, 5, 253-263.

London, J. (2017). Rhythm. In *New Grove dictionary of music and musicians*. Oxford: Oxford University Press, 21, 277-309.

London, J., Jones, K. (2011). Rhythmic refinements to the nPVI measure: a reanalysis of Patel & Daniele. In *Music Perception: an Interdisciplinary Journal*, 29, 1, 115-120.

Marotta, G. (2012). Piedi metrici e sillabe orfane nella prosodia dell'italiano. In Schafroth, E., Selig, M. (Eds.), *Testo e ritmi. Zum Rhythmus in der italienischen Sprache. Studia Romanica et linguistica, 35*. Frankfurt a.M.: Lang, 89-103.

Nespor, M., Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Nespor, M., Vogel, I. (2007). *Prosodic Phonology*: *with a new foreword*. Berlin-New York: Mouton De Gruyter.

Nespor, M., Shukla, M. & Mehler, J. (2011). Stress-timed *vs.* Syllable-timed languages. In Van Oostendorp, M., Ewen, C.J., Hume, E. & Rice, K. (Eds.), *The Blackwell companion to phonology II*. Oxford: Wiley-Blackwell, 1147-1159.

Oussilova, E. (1995). Effect of vowel modification on the phonemic accuracy of vowels and palatalization of the consonants in Russian vocalized speech. In Elenius, K., Branderud, P. (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences. ICPhS 1995. Stockholm, Sweden. 13-19 August 1995*, *vol. 1*. Stockholm: The Congress Organisers at KTH and Stockholm University, 210-213.

Patel, A.D., Daniele, J.R. (2003a). An empirical comparison of rhythm in language and music. In *Cognition*, 87, B35-B45.

Patel, A.D., Daniele, J.R. (2003b). Language, music, syntax and the brain. In *Nature Neuroscience*, 6, 674-681.

Pike, K.J. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Publications.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *1st International Conference on Speech Prosody*, France, 115-120.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. In *Cognition*, 73, 3, 287-288.

Sadakata, M., Desian, P., Honing, H., Patel, A.D. & Iversen, J.R. (2004). A cross-cultural study of the rhythm in English and Japanese popular music. In *Proceedings of the International Symposium on Musical Acoustics (ISMA)*, Nara, 41-44.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure.* Cambridge, MA: The MIT Press.

Steele, J. (1775). *An essay toward Establishing the Melody and Measure of Speech*. London: Printed by Bowyer, W., Nichols, J., for Almon, J.

Sundberg, J., Lindblom, B., (1976). Generative Theories in Language and Music Description. In *Cognition*, 4, 99-122.

Tilsen, S., Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. In *The Journal of the Acoustical Society of America*, 134, 1, 628-639.

Appendix

Balla Linda – Lucio Battisti

Bella Linda – The Grass Roots

Mi ritorni in mente – Lucio Battisti

Wake me I am dreaming – The Love Affair

DOI: 10.17469/O2106AISV000022

CLAUDIO ZMARICH, ERICA BONATO, ROBERTO BOVO,
VINCENZO GALATÀ, ALESSANDRA PINTON

A new phonological discrimination test for children aged 48-72 months[1]

This contribution is part of a wider project aiming at the creation of a phonological discrimination test for preschool subjects (48- to 72-months-old). Children with primary language impairment are usually affected by deficits in speech production and/or in the phonological representation of speech sounds causing discrimination disorders. The proposed test is designed to evaluate the ability to discriminate pairs of non-words through a "same/different" judgment. The test items combined in pairs examine all the possible contrasts' simplifications (i.e. "phonological processes") that may accompany the speech production of typical developing preschool children even if these processes are not always (and all) present during language development, and even if there is currently no consensus in the literature on which processes belong to typical or atypical development.

Key words: child language development, speech perception, non-words discrimination, phonological processes, AX testing.

1. *Introduction*

1.1 The development of speech perception and its clinical implications

We know from the literature that: i) a correct perception (and production) of the sound system of a language is the *sine qua non* condition to be able to access the other levels of the spoken language (Saffran, Werker & Werner, 2006; Werker, 2018); ii) speech and language disorders with a phonetic-phonological component are an important, if not the main, portion of the caseloads of pediatric speech language therapists (Law, Boyle, Harris, Harkness & Nye, 2000; Joffe, Pring, 2008; Eadie, Morgan, Ukoumunne, Ttofari Eecen, Wake & Reilly, 2015).

A vast body of literature has shown the role of auditory discrimination in language disorders, both in children and in adolescents (Sussman, Steinschneider, Lee & Lawson, 2015; Kujala, Leminen, 2017). Some authors hypothesize that the underlying cause to such disorders has to be sought in a deficit in the auditory elaboration of the rapid formant transitions determining the difficulty to quickly process the input (Tallal, 1980; Moore, Rosenberg & Coleman, 2005). Other authors, on the other hand, claim that these disorders are determined by a deficit in the ability to represent the phonological form in a stable and well-formed manner: the per-

[1] Authorship note: while the paper is the result of a joint collaboration and discussion between the five authors, the main research idea is to be attributed to CZ, VG and AP.

formance of these children is influenced by the degree of phonological similarity leading to significant difficulty in the discrimination of minimal pair stimuli (opposed in terms of place, manner and sonority in both a silence and noise condition), rather than by the rapid spectral change in the components of the auditory signal (Ziegler, Pech-Georgel, George & Lorenzi, 2011; Nittrouer, Shune & Lowenstein, 2011; Leybaert, Macchi, Huyse, Champoux, Bayard, Colin & Berthommier, 2014). According to others, these difficulties would be the core of specific language disorders and would represent the link to the risk of learning disorders to which they are so frequently associated (e.g. Vandewalle, Boets, Ghesquière & Zink, 2012).

According to Preston, Irwin & Turcios (2015) the perception of acoustic details is important as they are eventually mapped onto phonological and motoric representations that govern speech production. This information is also clinically relevant and many clinical remedial interventions tackling speech sound disorders exploit perception (e.g. Hodson, 2010; Bowen, 2014; Williams, McLeod & McCauley, 2010; Rvachew, Brosseau-Lapré, 2012; Kaiser, Scherer, Frey & Roberts, 2017) in order to facilitate an accurate production of speech sounds. As reported by Lof and Synan (1997), the possible simultaneous occurrence of speech perception problems is known since a quite long time in the articulation/phonology literature: as stated in early clinical writings by Van Riper (1939), "all children with speech sound production errors should routinely have their discrimination tested by having them compare the clinician's imitation of the misproduced sound to a correct production" (Lof, Synan, 1997: 63).

Although both electrophysiological (Gonçalves, Wertzner, Samelli & Matas, 2011) as well as brain imaging studies (Preston, Felsenfeld, Frost, Mencl, Fulbright, Grigorenko, Landi, Seki & Pugh, 2012) have evidenced underlying neural abnormalities in auditory and perceptual speech processing of children with phonological disorders, these techniques are clinically impractical. Hence, it is important to determine if behavioural measures of speech perception can sensitively support the identification of perceptual differences in school-age children with phonological disorders. Speech perception differences could include problems at different levels such as sound discrimination, categorical perception, and goodness judgments of phonetically acceptable and unacceptable productions of words with reference to the speaker's native language and dialect. A speech sound discrimination task typically requires children to recognize differences between phonetically similar items; a task based on categorical perception requires the child to classify synthetic speech tokens into phoneme categories; goodness judgement tasks require children to determine if a token is phonetically acceptable or unacceptable for a given category (Preston et al., 2015). Since we chose to test speech sound discrimination in 4- to 6-years-old children, we will briefly describe what is currently known about the speech perceptual capacities of children of these ages and the characteristics of a speech discrimination task.

1.2 The perceptual skills of pre-school children

The development of speech perception abilities in children up to 5 years of age is documented by a huge amount of literature (for some recent surveys the interested reader is referred to Choi, Black & Werker, 2018; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008; Saffran et al., 2006; Walley, 2005; Werker, 2018). We focus here on 4- to 6-years-old pre-school children because this is the age-range at which it is possible to gather reliable responses by applying adult-like testing. Starting with the ability to process the acoustic dimensions of speech, Jensen, Neff (1993) demonstrated that children tested at 4 years of age and re-tested 12-18 months later, improved speech discrimination skills beginning with variations in intensity, followed by changes in frequency and finally by changes in duration. However, at the final assessment, for many of them, discrimination in the domain of frequency and duration was still poorer than adults' discrimination. This delay in sensitivity maturation with reference to temporal information is due to both a central level processing and a working memory capacity. From a developmental point of view, the perception of consonants is, generally speaking and until 5-6 years of age, less categorical in nature and more influenced by the context with respect to adults' perception. According to Walley (2005), all these outcomes are compatible with the hypothesis that 4- to 5-years-old children are more dependent from a global and syllabic representation of speech rather than from a segmental one.

1.3 Phonological discrimination tests

In broader terms, phonological discrimination refers to that process of categorical perception through which differences that unfold along a physical continuum (of frequency, intensity, duration) are traced to discrete categories. Phonological discrimination represents an essential part of a normal speech perception development and systematically improves till 10 years of age (Edwards, Fox & Rogers, 2002), although the cornerstones for a correct discrimination are already laid down by the age of five (Tamashige, Nishizawa, Itoda, Kasai, Igawa & Fukuda, 2008). The children's normal development can fortunately be tested starting from age 4-5 by using the same methods used for adults (Polka, Jusczyk & Rvachew, 1995). Phonological discrimination tests represent an important procedure for assessing proficiency in speech acquisition, and any alteration in the ability to discriminate "similar" sounds could contribute to unveil the onset of speech and/or language disorders (Lof, Synan, 1997; Pascoe, Rossouw, Fish, Jansen, Manley, Powell & Rosen, 2016).

Phonological discrimination tests may vary with respect to both the form and the content (Vance, Rosen & Coleman, 2009). Regarding the form, i.e. the procedural paradigm used to test the phonological discrimination skills, the AX or "same/different" paradigm is to be preferred, notwithstanding a possible bias towards a "same" judgment (see Gerrits, Schouten, 2004, although their subjects were adults), because of less taxing the working-memory of the younger children in comparison to more sophisticated designs (Polka et al., 1995). Regarding the content, non-words (vs. words) stimuli are to be preferred because they are independent from previous lexical knowledge, thus engaging only the perceptual system and/or the phonological memory, but not the lexical/semantic system. Anyway, according to McAllister Byun (2015), by 5 years of age the children's discrimination skills are essentially adult-like.

It is important to emphasize the distinction between the ability to discriminate two sounds (minimal pairs) from the ability to use this contrast in a phonologically relevant way (to learn new sounds), since the two skills can have different time courses. In fact, production errors in older children who have a speech disorder may reflect either motor problems or an inadequate phonemic representation (Rvachew, Ohberg, Grawburg & Heyding, 2003; Gierut, 1998; Pascoe et al., 2016; Stackhouse, Pascoe & Gardner, 2006). At present, the question of how linguistic perception and production are interrelated is still unresolved. During development, it is possible that the child perceives speech at almost the adult level, but that he does not yet have the motor skills to achieve a certain target (McAllister Byun, Tiede, 2017). On the other hand, it is possible that the child has adequate motor skills, but still a too wide auditory-perceptive representation of the target, with the consequence of not being able to receive the error feedback that would lead him to modify his motor planning (Shiller, Rochon, 2014).

Regrettably, the understanding of the perception-production link in speech acquisition is hampered by the theoretical and historical division, since the first 70s, of the research field into two separate traditions, which Dunbar, Idsardi (2010) call as "Child Phonology", and "Infant Child Perception". The first one is embraced by the linguists' community and originated by the pioneering studies on speech production by Jakobson (1941, but first published in English in 1968), as documented by parents' diaries. The second one is embraced by most developmental psychologists and could be traced to the seminal work of Eimas, Siqueland, Jusczyk & Vigorito (1971) who shifted the focus from observational studies of production to laboratory work in perception. These two traditions elaborated different views about the primitives of phonology, the phonological features. According to the traditional linguistic theory, the distinctive features are viewed as abstract cognitive entities that characterize a certain sound in the mind of the speaker/listener (Chomsky, Halle, 1968), and they are generally be assumed to be part of universal grammar, the innate language faculty underlying chomskyan generative theory. The innateness' assumption exempted the linguists from looking for a natural history of features acquisition, but most of the developmental psychologists, together with the functional linguists, have provided serious challenges to this view, by arguing that features are learned and language specific, rather than innate and universal. In a sense, they are inherent, since they are biologically grounded as opposed to arbitrary, but as emergent from an intricate interplay among auditory-acoustic input signal, the child's developing cognitive capacity and the developing articulatory capacity: they do not stand as pre-experiential cognitive given (Menn, Vihman, 2011).

In considering the acquisition of phonological features, it is necessary to acknowledge that they perform two main functions:

– "distinctive" (Cristià, Seidl & Francis, 2011) or "lexical" (Ridouane, Clements, 2011): they are used to distinguish sounds in lexical contrast with each other (i.e. an acoustic difference can lead to a change in meaning, as shown by minimal pairs: e.g. It. /ˈpane/, Eng. "bread" ~ It. /ˈkane/, Eng. "dog");
– "classifying" (Cristià et al., 2011) or "phonological" (Ridouane, Clements, 2011): they determine the classes of sounds based on common characteristics, which may be subject to the same phonological rule. 7- to 9-months-old children quickly learn phonotactic patterns in auditory input and generalize the constraints to new sequences, although phonetic similarity is probably necessary to form a sound category (Saffran, Thiessen, 2003; Cristià et al., 2011).

One of the best evidences that features do not depend on the presence of minimal pairs in the to-be-learned child's lexicon but on the environment instead, comes from their relative acoustic salience which makes them more easily discriminable: for instance, while voiced and voiceless stops are discriminated by 1 to 4 months of age (Eimas et al., 1971), the contrast between a stop and a voiced dental fricative is still undergoing development at 10-12 months (Polka, Colantonio & Sundara, 2001). Furthermore, the fact that 14-months-old children are not able to learn minimal pairs differing in voicing (Stager, Werker, 1997) underlines the importance of distinguishing the ability to discriminate two sounds from the ability to use this contrast in a phonologically relevant manner.

We would like to conclude this paragraph by quoting the final summary by Cristià et al. (2011: 14): "experimental research in infancy does not support the hypothesis that a single construct, an abstract phonological feature set, underlies speech sound discrimination, sound pattern learning, and word learning. On the contrary, this research suggests a separation between discrimination of acoustic contrasts, the distinctive function, and the classificatory function of features [...]".

A more in-depth understanding of the relationships between perception and production in the course of development would help better understand the enormous variability of production capacity observed in children. A logical way to investigate this relationship is to provide measures of speech perception and production within the same child, and we adopted this strategy.

1.4 The need for a new phonological discrimination test

We started our endeavour in the creation of this test moving from the consideration that existing tests for Italian do not evaluate the phonological discrimination of late pre-schoolers (e.g. 48- to 72-months-old) in a satisfactory way. To be precise, there are only two norm-referenced tests that make use of non-words' pairs in an AX paradigm. These tests are proposed as a part of a battery (BVN 5-11, Tressoldi, Vio, Gugliotta, Bisiacchi & Cendron, 2005, also in Pinton, Zanettin, 1998; CMF, Marotta, Trasciani & Vicari, 2008), but they present a number of problems that can be summarised as follows:

- they do not systematically test all the phonological contrasts of Italian nor do they adopt a selection of the most significant contrasts to test the children's phonological abilities with reference to their age;
- there is no apparent attention to the phonetic context;
- the length opposition is not tested;
- it is not clear how the consonantal groups are evaluated;
- the administration procedure is neither automated nor randomized;
- the tests are not proposed in a playful way (or at least no further indication is provided on how to propose the test to the children);
- the verbal stimuli have to be pronounced anew each time by the clinician/therapist (i.e. they are not pre-recorded as they are in Pinton, Zanettin, 1998);
- the test items are written in orthography rather than in IPA.

By capitalizing on advantages and disadvantages of these previous tests and moving from a preliminary test prepared and used by Galatà and Zmarich (2011a; 2011b) to investigate the phonological discrimination of immigrant children acquiring Italian as L2, we started to elaborate a brand new test.

2. *Experimental section*

2.1 The rationale behind the new phonological discrimination test and its implementation

The proposed test development consists of two sets. The first set of stimuli is based on contrasts involving processes causing *systemic* simplifications (see, among many others: Grunwell, 1987: 221; Pinton, 2018: 105-110). These processes hold this name because they mainly consist of phones' substitution, the final effect being the reduction of the phonological inventory/system. The second set of stimuli is based on contrasts involving processes effecting *structural* simplifications (Grunwell, 1987: 212; Pinton, 2018: 105-110). These processes hold this name because they mainly consist in phones' harmony or phones' addition/deletion, the final effect being the alteration of the original phonological shape/structure. The test follows the AX paradigm (more on this later) with test items proposed in pairs. For each set, a number of control stimuli pairs have also been devised (with completely equal or completely different pair members), in order to individuate and exclude those children who are not able to accomplish the task. To this regard, a pre-test session precedes both test sets with the aim of familiarizing the children with the task.

The first set consists of 40 minimal pairs with 'VCV non-words, testing all the 10 distinctive features of the Italian phonological system (as exposed by Schmid, 1999: 134) involved in as many *systemic* processes (each process is tested with four pairs of stimuli):

1. Gliding [±consonantal]
2. Stopping/Liquid [±sonorant]
3. Stopping/Frication [±continuous]

4. Affrication/Deaffrication [±delayed released]
5. Devoicing/Voicing [±voice]
6. Nasalization/Denasalization [±nasal]
7. r/l - l/r Substitution [±lateral]
8. Fronting/Backing [±coronal]
9. Fronting/Backing [±anterior]
10. Fronting/Backing [±posterior]

Each distinctive feature is represented by two different consonant pairs for each of the two vocalic contexts (ˈaCa, ˈiCi), with each pair opposing a [+*feature*] consonant against a [–*feature*] consonant. Crucially, the two pairs make use of consonants which are different one from another but at the same time are sharing the same featural opposition, as in the case [+continuous] ~ [–continuous]. If we postulate the first consonant to be the target, and the second to be the output of a phonological process, the phonological process in the examples provided in (1) is called *stopping*:

(1) /ˈasa/~/ˈata/; /ˈisi/~/ˈiti/; /ˈava/~/ˈaba/; /ˈivi/~/ˈibi/

In this way we can be more confident that the child's response will depend on the common feature shared by the consonants rather than on the consonants *per se*.

The second set consists of 56 non-words' minimal pairs testing 14 *structural* processes (four pairs of stimuli for each process):

1. Weak syllable deletion;
2. Consonant omission;
3. Vowel omission;
4. Consonant metathesis;
5. Syllable metathesis;
6. Epenthesis;
7. Diphtong reduction;
8. Consonant harmony;
9. Vowel harmony;
10. Voiceless assimilation;
11. Degemination;
12. Homosyllabic cluster reduction (from 3 to 2 consonants);
13. Homosyllabic cluster reduction (from 2 to 1 consonant);
14. Heterosyllabic cluster reduction (from 2 to 1 consonant).

Each of the 14 processes is represented in two different pairs for each of the different vocalic contexts /a/ or /i/, with each pair opposing a target against the same target as affected by a particular process. The examples in (2) are relative to the process of homosyllabic cluster reduction (from 2 to 1 consonant):

(2) /ˈspata/~/ˈpata/; /ˈprata/~/ˈpata/; /ˈspiti/~/ˈpiti/; /ˈpriti/~/ˈpiti/

As written before, the particular system of distinctive features that we chose for describing Italian is the one presented by Schmid (1999), which adapted the theoretical framework proposed by Chomsky and Halle (1968) to Italian. Since our aim is to assess the children's perception of stimuli mirroring the typical articulatory distortions produced by late pre-schoolers, the nature of the distinctive features has to be articulatory based as in the Chomsky and Halle system.

As for the stimuli testing the *systemic* phonological processes, given that each phonological process describes a pattern of error that involves groups of phonemes sharing a critical feature (the one that defines a certain natural class, based on articulatory features), for the selection of the four consonants (two for each vocalic context) among the many potentially eligible which represent that class, we chose the consonants that:

- within each pair, are in opposition (+/–) possibly on just the articulatory feature defining the particular process, and at the same time are acoustically as similar as possible (on the basis of Jakobson's distinctive features, adapted for Italian by Muljacic, 1972, and revised by Mioni, 1983);
- do not give origin to real words;
- whenever possible, are nor marked for frequency in the language (Goslin, Galluzzi & Romani, 2014) neither for age of acquisition (Zanobini, Viterbiori & Saraceno, 2012).

The last two criteria were also used to create the stimuli testing the *structural* processes, with a supplementary criterion dictated by the particular complexity of some stimuli, as for example for the process of *weak syllable deletion* in (3):

(3) /pakaˈpata/~/kaˈpata/; /pikiˈpiti/~/kiˈpiti/; /fakaˈfata/~/kaˈfata/; /fikiˈfiti/~/kiˈfiti/

This additional criterion requires all the consonants constituting the stimuli not to be marked on a distinctive feature which is not involved in the process and which is shared by the consonants: for instance, all the consonants have to be equal for the [±voice] feature by taking the "–" (minus) value. In fact, the unvoiced (not aspirated) consonants are less marked than the voiced ones because they are more frequent in the languages of the world, less articulatory complex and learned earlier by the children.

A final criterion common to both *systemic* and *structural* processes, requires that each pair of stimuli must test a single process (i.e. the same pair cannot be used to test two different processes). This is necessary to avoid interpretation ambiguity in case of children failing the discrimination of the proposed pair.

Before moving further, we would like to clarify and address an important methodological *caveat*. We are perfectly aware that if a child does not distinguish two stimuli which are different, one cannot determine through the discrimination test alone the direction of the assimilation: for instance, confusion between [+continuous]~[–continuous] stimuli could be due to a perception of both consonants as stops (e.g. *stopping*) or fricatives (e.g. *frication*). To individuate the source of misperception, this test needs to be integrated with other tests: our project includes the use of an articulation test (see paragraph 2.2) which can offer an indication about the articulation errors in production, but in the next future we foresee the use of a forced choice word perception test, to further selectively investigate the critical features or processes evidenced by the test we propose here.

The test stimuli were recorded in a soundproof cabin and were pronounced aloud within a carrier sentence by the second author[2]. Each of the test stimuli has then been isolated and normalized for intensity by means of Praat. A list of test pairs was compiled for both sets and implemented by the fourth author by means of Praat's *ExperimentMFC* object in order to facilitate the administration and responses collection on a notebook. The stimuli pairs in each set (with the two members of each pair separated by one second silence and the target stimuli always in the same position) were automatically randomized and proposed to the children in a playful setting using two puppets from a popular story (i.e. the *Talking Cricket*, It. "il Grillo Parlante", and *Pinocchio*, see Figure 1). The experimenter presented the test to the children as follows: «Now I'll introduce you to two characters. This is the *Talking Cricket* (indicate): he is very good at saying words, especially when the word is "magic". This, instead, is *Pinocchio* (indicate): he wants to learn to speak like the *Cricket*, and to learn the magic words pronounced by the cricket. Let's see if he's good. Now the *Cricket* will say a word and *Pinocchio* will repeat it: if he says the word correctly, you will tell him *Bravo!* and you will give him a token». Alternatively, depending on the child, the answer requested after listening to the two items was one of the two options: "right/wrong" (e.g. «Was the word repeated by *Pinocchio* right or wrong?»), or "same/different" (e.g. «Did *Pinocchio* repeat the same word or a different one?»). In some cases, we also used small coloured cards, on which the features of a smiling face were drawn, as a means by which the child could express his own judgment of equality between the pair's non-words. Before starting with the experimental session, the child was asked to listen and to answer to some "easy" stimuli pairs consisting of real words in order to accustom him/her to the task, followed by some instances of non-words, exemplifying the presence of *systemic* as well as *structural* processes.

Figure 1 - *The setting of the test administration*

[2] In Gaiotto (2016) we experimentally tested and verified the usage of both synthetic and natural stimuli in a preliminary version of the proposed discrimination test. We found normal stimuli to be superior to synthetic stimuli created by a dynamically generated voice using the Mivoq Voice Synthesis Engine2 (which represents the state of the art of Italian synthesis; Tesser, Sommavilla, Paci & Cosi, 2016): children better perceive normal over synthetic stimuli.

Once the child declared to be ready, and was judged so also by the experimenter, the test started with the presentation of the experimental pairs. For each proposed pair the experimenter registered the child's answer by pushing one of two predefined keys ("same/different") on the notebook. If requested by the child, each proposed pair was played back again for no more than two times. If the child did not express any evaluation, the answer was given the "uncorrect" value, and the successive pair was proposed after a short break. After completing half of the list of item pairs, the procedure stopped for a short pause. At the end of the whole session, the experimenter saved and exported the collected responses to a text file. The whole procedure and administration setting were adapted from a previous project (see Galatà, Zmarich, 2011a, b; Galatà, Angonese & Zmarich, 2017).

2.2 Subjects, materials and procedure

The two sets of the test were administered by students graduating in speech therapy at the University of Padova to two groups of monolingual Italian children (balanced per gender) recruited and tested in various kindergarten in the Veneto region (Italy):
- the first group (56 subjects) was assessed with the first set of the test, namely the test for *systemic* processes (Bonato, 2016);
- the second group (61 subjects) was assessed with the second set of the test, namely the test for *structural* processes (Rossi, 2017; Marchetti, 2017; Degano, 2018).

According to their parents, who gave consent and compiled a questionnaire, all the children were normal under the psychophysical profile. Subsequently, the children were tested for hearing integrity through a *Starkey WRA* audiometer equipped with *Qualitone* headphones, and an illustration depicting a wooded landscape and some animals playing different instruments, attached to the instrument's flap cover. The test consisted of 9 stimuli, played to the children within a range of intensity between 20 and 60 dB with a 5 dB step for each of the following frequencies: 500, 1000, 2000 and 4000 Hz. Each child was invited to indicate when no sound was perceived anymore: the last perceived frequency was manually recorded as a threshold value. The hearing thresholds provided for exclusion from the sample were 40 dB for 500 Hz, 30 dB for 1000 Hz and 20 dB for 2000 and 4000 Hz. In a separate session all the children have also been tested with an articulation test (a modified version of Test Fonetico per la Prima Infanzia, TFPI, not published; see Zmarich, Fava, Del Monego & Bonifacio, 2012) during which they were asked to name a set of pictures. The data from the articulation test – not addressed here and used in a separate study – will be used to compare speech production and perception proficiency within each child.

The first half of each of the two groups was offered a test administration in two rounds: in the first round the audiometric test and the first part of the phonological discrimination test were performed, while in the second round the TFPI was performed – followed by the second part of the phonological discrimination test. For the second half of each of the two groups, the administration of the tests followed a reverse order: TFPI, first part of the phonological discrimination test, audiometric test and the second part of the phonological discrimination test.

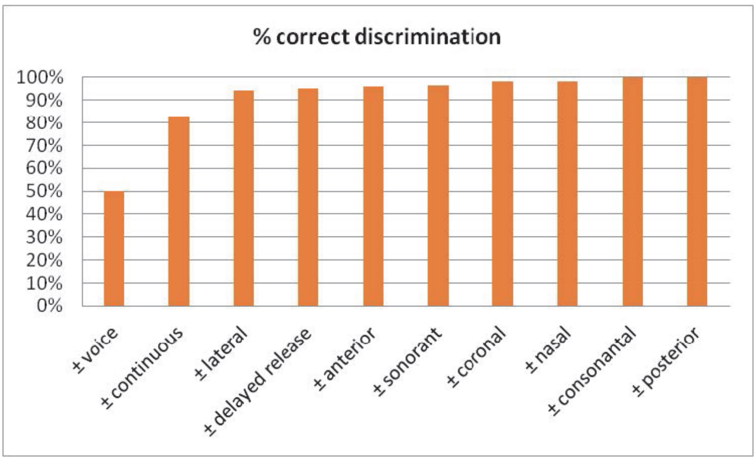
Two subjects were excluded from the first group because they did not understand the discrimination task. Three subjects were excluded for the same reason from the second group with an additional one excluded for non collaborative behavior. Moreover, a number of children exhibited some selected deficits to the audiometric test (11 for the first group, 15 for the second group). However, since none of the subjects showed deficits to both ears, nor to all frequencies, and since our study made use of stimuli in the open field, in this preliminary phase these subjects were included in the sample and coded with a special label for future analyses.

At the end, the number of children considered for the analyses were 54 subjects for the first group tested for *systemic* processes and 57 for the second group tested for *structural* processes.

2.3 Results

With reference to the analyses for the first test set (e.g. *systemic* contrasts), we considered distinctive features (ten levels, see Figure 2), age (four groups spanning six months each) and gender (male and female) as between variables and vocalic context as within variable in a repeated measures ANOVA. Percentage of correct discrimination for each feature was converted to arcsine scores ($y = 2 \arcsin \sqrt{x}$) and entered as dependent variable. No significant differences were found for age group ($F_{3,80} = 1.859$; $p = .143$) or gender ($F_{1,80} = 0.345$; $p = .558$). Distinctive features resulted highly significant ($F_{9,80} = 18.687$; $p < .000$).

Figure 2 - *Systemic phonological processes: discrimination percentages for distinctive features*



A post-hoc pairwise comparison test (with Tuckey correction for multiple comparisons) showed that [±voice] feature contrasted significantly ($p < .001$) with all the others. Indeed, this feature obtained only 50% of correct discrimination (see

Figure 2). The within-subject variable (vocalic context) resulted highly significant ($F_{1,80} = 30.793$; $p < .001$): discrimination was almost overall better in /a/ context (93%) than /i/ context (87%). Furthermore, the vocalic context interacted significantly with the distinctive features ($F_{9,80} = 8.364$; $p < .000$): a post-hoc pairwise comparison test (with Tuckey correction for multiple comparisons) showed for the continuous feature (and more specifically for /ˈava/~/ˈaba/; /ˈivi/~/ˈibi/) a reverse pattern for the vocalic context with the second pair better discriminated than the first one.

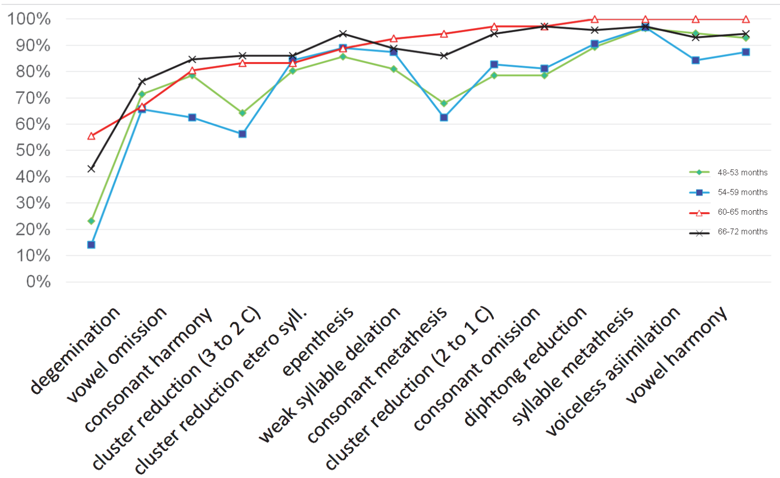
Since each distinctive feature is involved in a characteristic phonological process, a second repeated measures ANOVA was performed on the same scores as before as a dependent variable, this time considering phonological processes (seven levels) instead of features, and maintaining age (four groups spanning six months each) and gender (male and female) as between variables, with vocalic context as within variable. No significant differences were found for age group ($F_{3,104} = 1.890$; $p = .136$) or for gender ($F_{1,104} = 0.122$; $p = .728$). Phonological processes resulted significant ($F_{6,104} = 29.173$; $p < .001$). A post-hoc pairwise comparison test (with Tuckey correction for multiple comparisons) showed that the *voicing* phonological process contrasted significantly ($p < .001$) with all the others. Adding to this, stopping contrasted significantly with backing ($t_{104} = 3.285$; $p = .023$) as well as with gliding ($t_{104} = 3.139$; $p = .035$). The within-subject variable (vocalic context) resulted significant ($F_{1,104} = 38.766$; $p < .001$; discrimination was almost overall better in /a/ context than /i/ context), and the vocalic context interacted significantly with phonological processes ($F_{6,104} = 9.191$; $p < .000$). A post-hoc pairwise comparison test (with Tuckey correction for multiple comparisons) showed that the *voicing* phonological process contrasted significantly ($p < .001$) with all the others in both vocalic contexts (/a/, /i/). Moreover, l/r substitution process in /a/ context contrasted significantly with the same process in /i/ context ($t_{104} = 3.931$; $p = .011$), with the /a/ stimuli better discriminated than the /i/ stimuli.

As a final exploration, we performed an ANOVA to determine whether the discrimination scores improved as a function of the number of distinctive features distinguishing the pairs' stimuli. The variables proved to be highly significant ($F_{3,316} = 29.483$; $p < .000$), with the pairs having only one contrasting feature to be significantly different to all the others ($p < .000$, post-hoc pairwise comparison test with Tuckey correction for multiple comparisons). To this regard, the percentages of correct discrimination were 77.10% for one feature, 94.74% for two features, 100% for three features and 98.61% for four features.

With reference to the analyses for the second test set (e.g. *structural* contrasts), we factorized in an ANOVA statistical design the phonological processes (14 levels, see Figure 3), age (four groups spanning six months each) and gender (male and female) and their interactions. Percentage of correct discrimination was entered as dependent variable. Figure 3 shows the discrimination percentages for the 14 categories of *structural* phonological processes by age group.

For the *structural* contrasts tested with the second set of stimuli the following factors have been found statistically significant: phonological processes ($F_{13,686}$ = 26.299; p < .000), age ($F_{3,686}$ = 19.938; p < .000), gender by age interaction ($F_{3,686}$ = 9.148; p < .000). A Tukey's Honestly-Significant-Difference post-hoc test on phonological processes scores, using least squares means and MSE model of 0.463 with 686 *df*, allowed to discover that *degemination* scores were significantly lower than all other processes (p < .000); *vowel omission* was lower than *consonant omission* (p < .000), *diphthong reduction* (p < .000), *epenthesis* (p < .000), *syllable metathesis* (p < .000), *vowel harmony* (p < .000), *cluster reduction* (2-to-1 consonant) (p = .001) and *weak syllable deletion* (p = .003); *cluster reduction* (3-to-2 consonants) was lower than *diphthong reduction* (p < .000), *syllable metathesis* (p < .000), *voiceless assimilation* (p < .000), *vowel harmony* (p = .000), *epenthesis* (p = .020) and *consonant omission* (p = 0.021); *syllable metathesis* was higher than *consonant harmony* (p = .000), *consonant metathesis* (p < .000) and *heterosyllabic cluster reduction* (p = .011); *consonant harmony* was lower than *diphthong reduction* (p = .002), *vowel harmony* (p = .003) and *voiceless assimilation* (p = .004); *consonant metathesis* was lower than *vowel harmony* (p = .006) and *voiceless assimilation* (p = .009).

Figure 3 - *Structural phonological processes: discrimination percentages for categories by age group*



A Tukey's Honestly-Significant-Difference post-hoc test on the age variable, using least squares means and MSE model of 0.463 with 686 *df*, showed the first and the second group (respectively 48-53 months and 54-59 months) to be significantly different (p < .000) from the third and the fourth (respectively 60-65 months and 66-72 months). No significant difference was found for gender, whereas the vocalic context was not considered because not pertinent in a consistent proportion of stimuli pairs.

2.4 Discussion

Overall, considering the targeted age bands, *systemic* simplifications are managed better than *structural* ones. Comparing the performances of correct discrimination for *systemic* processes with those referred to *structural* processes in the whole sample of children, the resulting picture shows that *systemic* simplifications go beyond 90% of correctness (apart from those involving the [±voice] and [±continuous] features); analysing the *structural* simplifications, on the other hand, only four processes reach 90% of correctness (respectively *voiceless assimilation* 92.11%, *vowel harmony* 92.98%, *diphthong reduction* 93.42% and *syllable metathesis* 97.37%), with *degemination* obtaining only around 20% of correct responses for the first two age groups. The comparison of the children's performances across the different age bands (Figure 3) shows that the competences related to *structural* processes are under development given that the older children are significantly better than the younger ones. Anyway, by 5 years of age there are still a number of contrasts characterized by *degemination*, *cluster reduction (3-to-2 consonants)* and *consonant/vowel omission* that are not well discriminated by a consistent number of children.

The results from the discrimination of *systemic* processes reflect only partially what characterises production. From data on the phonetic inventory for Italian and for other languages, during typical development children seem to need more time to be able to properly master the [±voice] feature: studies on English, Canadian French and Italian show that voiceless stops, voiceless fricatives and voiceless affricates are mastered before the voiced homorganic ones (Dodd, Holm, Hua & Crosbie, 2003; McLeod, Sutton, Trudeau & Thordardottir, 2011; Zanobini et al., 2012; Viterbori, Zanobini & Cozzani, 2018; Zmarich, Bonifacio, 2005). However, perception and production don't seem to fully overlap. While voicing and devoicing processes in production are overcome by the children by the age of 3 and a half, in our sample these processes are not yet well discriminated. Affricates show the opposite case: even though they are mastered in production later with respect to other consonants' classes and the processes of affrication and deaffrication are present till older age, the discrimination of the opposition involving the [±delayed release] feature seems to be better mastered by the younger children in our sample. The question on which is the relation between perception and production remains open, e.g. whether there is a causal connection between the two or whether one of the two influences the other (McAllister Byun, 2015). A similar picture emerges if we observe the responses on the discrimination of some structural processes that can be compared to data on production in Scottish, English and Italian children (Cohen, Hodson, O'Hare, Boyle, Durrani, McCartney, Mattey, Naftalin, Watson, 2005; Dodd et al., 2003; Zanobini et al., 2012). While *cluster reduction* from 3-to-2 consonants at around 60 months of age seems to be problematic both in production and in perception (this process reaches 71.93% of correctness in our sample – ranging between 55 and 85% in the comparison across age groups – see Figure 3 – demonstrating that this process is still evolving), other processes seem to have a dissonant trend between production and perception. For example, according to the studies previously mentioned, the processes of *harmony* seem to disappear in production after 42 months of age, while in our sample children at 48 months of age (see Figure 3) show some weakness in the discrimination of contrasts involving *consonant harmony*.

Some children showed a reduced hearing threshold. However, their performance is on average fully in line with the performance of age matched peers. This demonstrates that the task we proposed to the children reflects a categorical perception: differently from auditory detection tasks, where the signal detection is achieved based upon the signal's intensity, in a discrimination task the recognition is based upon already mapped perceptual categories according to which distinctive features are detected even in noisy situations where the signal is degraded (see Ziegler et al., 2011).

3. *General discussion and conclusion*

The rationale and the choices adopted in the construction of the test can be so far be considered satisfactory allowing us to trace the path for future improvements of the test itself. In the following, we try to list a number of conclusions together with some questions still open.

The significant difference between the two vocalic contexts confirms the relevance of the vocalic context, at least for what concerns the discrimination of *systemic* contrasts. It is therefore relevant for a discrimination test to take into account such a factor.

As to the very low percentages of correctness for the *degemination* process, we can not exclude that this may be a consequence of the well known tendency to weaken the gemination contrast in production put in place by people living in the Veneto region (and therefore the children, too; see Telmon, 1997; Vietti, 2019).

Since we represented all the *systemic* and *structural* processes in a systematic way, based on the gathered data it is possible to identify those processes that better describe the children's abilities.

The structure of the proposed test was accepted by the children who fully accomplished the task (only 5 subjects have been excluded from the sample because of their scarce collaboration or non-comprehension of the proposed task): the possibility to use the test in its current form is confirmed and is in line with Polka et al. (1995) suggesting adult-like methods to test perceptual discrimination abilities in children.

However, the division of the current test in two parts, makes it a little cumbersome, long and potentially annoying the child, especially the younger ones. For this reason we are continuing working along the same lines described here to reduce the task load of the proposed test (currently about 20 + 20 minutes long) without reducing its validity and, according to preliminary results not addressed here, we are on the way to achieve a one session test lasting less than 30 minutes (Degli Agostini, 2019). Other aspects that need to be taken into consideration in future work are: i) whether the order of the target items in the pairs might have biased the results (the target items occurred always in the same position: for example, when testing the *structural* processes the "difficult" item was presented first and the "reduced" one next); ii) whether using the same voice for the stimuli could have biased the subjects' answers.

Future work is planned to further refine the proposed test as done in Degli Agostini (2019) by also increasing the number of children to be tested (including also children younger than 48 months of age and older than 72 months of age, as well as an adequate number of adults) and taking into account socioeconomic variables as well as geographical variables (e.g. regional variants of Italian) in order to be able to standardize the final and future version of the test. Additionally, we will consider the opportunity to run Mixed Effect Models rather than ANOVAs to incorporate the effect of the variability between subjects and item (random factors) in the model.

As already argued, since in case of speech errors it is not easy to ascribe them to production or perception factors, we started to investigate the correlations between the results from perception capacities (the present test) and production capacities (the phonetic test), by comparing for every child the mean percentage of correctness scores with the number of phones resulting from the phonetic inventory (Parisio, 2017; Boscolo-Nata, 2018). These results will be further addressed and proposed in future contributions.

Acknowledgments

To all the wonderful students (Alessia, Elisa, Erica, Lucia, Mariavittoria, Sabrina, Silvia M., Silvia B.N.) who graduated in Speech Therapy by using the present test and to all the children they administered it.

Bibliography

Bonato, E. (2016). La percezione fonologica in età pre-scolare. Costruzione e applicazione di un test di discriminazione uditiva. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Boscolo-Nata, S. (2018). Percezione e produzione fonetico-fonologica in età prescolare. Uno studio sperimentale. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Bowen, C. (2014). *Children's speech sound disorders*. Oxford: Wiley-Blackwell.

Choi, D., Black, A.K. & Werker, J.F. (2018). Cascading and multisensory influences on speech perception development. In *Mind, Brain, and Education*, 12(4), 212-223.

Chomsky, N., Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cohen, W., Hodson, A., O'Hare, A., Boyle, J., Durrani, T., McCartney, E., Mattey, M., Naftalin, L. & Watson, J. (2005). Effects of computer-based intervention through acoustically modified speech (Fast ForWord) in severe mixed receptive—expressive language impairment: Outcomes from a randomized controlled trial. In *Journal of Speech, Language, and Hearing Research*, 48(3), 715-729.

Cristià, A., Seidl, A. & Francis, A. (2011). Phonological features in infancy. In Clements, N., Ridouane, R. (Eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives*. Amsterdam: John Benjamins, 303-326.

Degano, A. (2018). Contrasti fonologici in età evolutiva: una prova di discriminazione uditiva. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Degli Agostini, L. (2019). Proposta di miglioramento di un nuovo test di discriminazione fonologica e sua verifica sperimentale. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Dodd, B., Holm, A., Hua, Z. & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. In *Clinical Linguistics & Phonetics*, 17(8), 617-643.

Dunbar, E., Idsardi, W. (2012). The acquisition of phonological inventories. In Lidz J., Snyder W. & Pater J. (Eds.), *The Oxford Handbook of Developmental Linguistics*. Oxford: Oxford University Press, 7-26.

Eadie, P., Morgan, A., Ukoumunne, O.C., Ttofari Eecen, K., Wake, M. & Reilly, S. (2015). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. In *Developmental Medicine & Child Neurology*, 57(6), 578-584.

Edwards, J., Fox, R.A. & Rogers, C.L. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. In *Journal of Speech, Language, and Hearing Research*, 45(2), 231-242.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. In *Science*, 171(3968), 303-306.

Gaiotto, M. (2016). L'influenza della struttura acustica dello stimolo nella discriminazione fonologica in bambini di età prescolare. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Galatà, V., Zmarich, C. (2011a). Una proposta per valutare l'influenza fonetico-fonologica della lingua di origine dei bambini figli di immigrati sull'acquisizione dell'italiano. In Bruno, G.C., Caruso, I., Sanna, M. & Vellecco, I. (Eds.), *Percorsi migranti*. Milano: McGraw-Hill, 24, 301-317.

Galatà, V., Zmarich, C. (2011b). Le non-parole in uno studio sulla discriminazione e sulla produzione dei suoni consonantici dell'italiano da parte di bambini pre-scolari. In Gili Fivela, B., Stella, A., Garrapa, L. & Grimaldi, M. (Eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua*. Atti del 7° convegno AISV, 26-28 Gennaio 2011, Università del Salento – Lecce. Roma: Bulzoni Editore, 118-129.

Gerrits, E., Schouten, M.E.H. (2004). Categorical perception depends on the discrimination task. In *Perception & psychophysics,* 66(3), 363-376.

Gierut, J.A. (1998). Treatment Efficacy: Functional Phonological Disorders in Children. In *Journal of Speech, Language, and Hearing Research,* 41, 85-100.

Gonçalves, I.C., Wertzner, H.F., Samelli, A.G. & Matas, C.G. (2011). Speech and non-speech processing in children with phonological disorders: an electrophysiological study. In *Clinics*, 66(2), 293-298.

Goslin, J., Galluzzi, C. & Romani, C. (2014). PhonItalia: a phonological lexicon for Italian. In *Behavior research methods*, 46(3), 872-886.

Grunwell, P. (1987). *Clinical Phonology*, 2° Ed.. London: Chapman & Hall.

Hodson, B. (2010). *Evaluating and enhancing children's phonological systems: Research and theory to practice*. Wichita, KS: Phonocomp Publishers.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Jensen, J.K., Neff, D.L. (1993). Development of basic auditory discrimination in preschool children. In *Psychological Science*, 4(2), 104-107.

Joffe, V., Pring, T. (2008). Children with phonological problems: A survey of clinical practice. In *International Journal of Language & Communication Disorders*, 43(2), 154-164.

Kaiser, A.P., Scherer, N.J., Frey, J.R. & Roberts, M.Y. (2017). The effects of enhanced milieu teaching with phonological emphasis on the speech and language skills of young children with cleft palate: A pilot study. In *American journal of speech-language pathology*, 26(3), 806-818.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). In *Phil. Trans. R. Soc. B*, 363, 979-1000.

Kujala, T., Leminen, M. (2017). Low-level neural auditory discrimination dysfunctions in specific language impairment—a review on mismatch negativity findings. In *Developmental Cognitive Neuroscience*, 28, 65-75.

Law, J., Boyle, J., Harris, F., Harkness, A. & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. In *International Journal of Language and Communication Disorders*, 35, 165-188.

Leybaert, J., Macchi, L., Huyse, A., Champoux, F., Bayard, C., Colin, C. & Berthommier, F. (2014). Atypical audio-visual speech perception and McGurk effects in children with specific language impairment. In *Frontiers in psychology*, 5, 422.

Lof, G.L., Synan, S. (1997). Is there a speech discrimination/perception link to disordered articulation and phonology? A review of 80 years of literature. In *Contemporary Issues in Communication Science and Disorders*, 24(24), 63-77.

Marchetti, S. (2017). Processi fonologici in età evolutiva. Analisi della letteratura e costruzione di un test di discriminazione uditiva per l'età pre-scolare. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Marotta, L., Trasciani, M. & Vicari, S. (2008). *Test CMF. Valutazione delle competenze metafonologiche*. Trento: Erickson.

Mcallister Byun, T. (2015). Perceptual Discrimination across Contexts and Contrasts in Preschool-Aged Children. In *Lingua*, 160, 38-53.

Mcallister Byun, T., Tiede, M. (2017). Perception-Production Relations in Later Development of American English Rhotics. In *PloS one*, 12(2).

McLeod, A.A., Sutton, A., Trudeau, N. & Thordardottir, E. (2011). The acquisition of consonants in Québécois French: A cross-sectional study of pre-school aged children. In *International Journal of Speech-Language Pathology*, 13(2), 93-109.

Menn, L., Vihman, M.M. (2011). Features in child phonology: inherent, emergent, or artefacts of analysis? In Clements, N., Ridouane, R. (Eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives*. Amsterdam: John Benjamins, 261-301.

Mioni, A. (1983). Fonologia. In Croatto, L. (Ed.), *Trattato di Foniatria e Logopedia, Vol. 2, Aspetti linguistici della comunicazione.* Padova: La Garangola, 51-87.

Moore, D.R., Rosenberg, J.F. & Coleman, J.S. (2005). Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. In *Brain and language*, 94(1), 72-85.

Muljacic Z. (1972). *Fonologia della lingua italiana*. Bologna: Il Mulino.

Nittrouer, S., Shune, S. & Lowenstein, J.H. (2011). What is the deficit in phonological processing deficits: Auditory sensitivity, masking, or category formation? In *Journal of experimental child psychology*, 108(4), 762-785.

Parisio, E. (2017), Studio sperimentale sulle relazioni tra produzione e percezione fonetico-fonologica in età pre-scolare. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Pascoe, M., Rossouw, K., Fish, L., Jansen, C., Manley, N., Powell, M. & Rosen, L. (2016). Speech processing and production in two-year-old children acquiring isiXhosa: A tale of two children. In *South African Journal of Communication Disorders*, 63(2), 1-15.

Pinton, A. (2018). *I disturbi fonetici e fonologici nell'età dello sviluppo.* Roma: Carocci.

Pinton, A., Zanettin, F. (1998). Le abilità fonetiche e fonologiche in età prescolare. Un compito di discriminazione uditiva. In Frasson, S., Lena, L. & Menin, S. (Eds.), *Procedure e metodi di trattamento nei disordini della comunicazione*. Pisa: Del Cerro, 130-142.

Polka, L., Colantonio, C. & Sundara, M. (2001). A cross-language comparison of/d/–/ð/perception: evidence for a new developmental pattern. In *The Journal of the Acoustical Society of America*, 109(5), 2190-2201.

Polka, L., Jusczyk, P.W. & Rvachew, S. (1995). Methods for studying speech perception in infants and children. In Strange, W. (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium (Md): York Press, 49-89.

Preston, J.L., Felsenfeld, S., Frost, S.J., Mencl, W.E., Fulbright, R.K., Grigorenko, E.L., Landi, N., Seki, A. & Pugh, K.R. (2012). Functional brain activation differences in school-age children with speech sound errors: speech and print processing. In *Journal of Speech, Language, and Hearing Research*, 55(4), 1068-1082.

Preston, J.L., Irwin, J.R. & Turcios, J. (2015). Perception of speech sounds in school-aged children with speech sound disorders. In *Seminars in speech and language*, 36(4), 224-233.

Ridouane, R., Clements, G.N. (2011). Editors' overview. In Clements, N., Ridouane, R. (Eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives*. Amsterdam: John Benjamins, 3-11.

Rossi, S. (2017). Processi fonologici in età evolutiva. Uno studio di discriminazione uditiva in età pre-scolare. Tesi di Laurea in Logopedia, Università degli Studi di Padova.

Rvachew, S., Brosseau-Lapré, F. (2012). An input-focused intervention for children with developmental phonological disorders. In *Perspectives on Language Learning and Education*, 19(1), 31-35.

Rvachew, S., Ohberg, A., Grawburg, M. & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. In *American Journal of Speech-Language Pathology*, 12, 463-471.

Saffran, J.R., Thiessen, E.D. (2003). Pattern induction by infant language learners. In *Developmental psychology*, 39(3), 484.

Saffran, L.R., Werker, J.F. & Werner, L. (2006), The infant's auditory world: hearing, speech, and the beginnings of language. In Kuhn D., Siegler R. (Eds.), *Handbook of Child Psychology*, 6th edn, vol. 2: *Cognition, Perception, and Language*. New York: Wiley, 58-108.

Schmid, S. (1999). *Fonetica e fonologia dell'italiano*. Torino: Paravia Scriptorium.

Shiller, D.M., Rochon, M.L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. In *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1308-1315.

Stackhouse, J., Pascoe, M. & Gardner, H. (2006). Intervention for a child with persisting speech and literacy difficulties: A psycholinguistic approach. In *Advances in Speech Language Pathology*, 8(3), 231-244.

Stager, C.L., Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. In *Nature*, 388(6640), 381.

Sussman, E., Steinschneider, M., Lee, W. & Lawson, K. (2015). Auditory scene analysis in school-aged children with developmental language disorders. In *International Journal of Psychophysiology*, 95(2), 113-124.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. In *Brain and language*, 9(2), 182-198.

Tamashige, E., Nishizawa, N., Itoda, H., Kasai, S., Igawa, H.H. & Fukuda, S. (2008). Development of Phonemic Distinction in Japanese Preschool Children. In *Folia Phoniatrica et Logopaedica*, 60(6), 318-322.

Telmon, T. (1997). Varietà regionali. In Sobrero, A.A. (Eds.), *Introduzione all'italiano contemporaneo, Vol. II. La variazione e gli usi.* Bari: Laterza, 93-149.

Tesser, F., Sommavilla, G., Paci, G. & Cosi, P. (2016). Automatic Creation of TTS Intelligibility Tests. In Savy, R., Alfano, I. (Eds.), *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere.* Studi AISV 2, Milano: AISV, 289-304.

Tressoldi, P.E., Vio, C., Gugliotta, M., Bisiacchi, P.S. & Cendron, M. (2005). BVN 5-11: *Batteria di valutazione neuropsicologica per l'età evolutiva*. Trento: Erickson.

Van Riper, C. (1939). *Speech Correction: Principles and Methods*. Englewood Cliff (NJ): Prentice-Hall.

Vance, M., Rosen, S. & Coleman, M. (2009). Assessing speech perception in young children and relationships with language skills. In *International Journal of Audiology*, 48(10), 708-717.

Vandewalle, E., Boets, B., Ghesquière, P. & Zink, I. (2012). Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. In *Research in Developmental Disabilities*, 33(2), 635-644.

Vietti, A. (2019). Phonological variation and change in Italian. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Viterbori, P., Zanobini, M. & Cozzani, F. (2018). Phonological development in children with different lexical skills. In *First Language*, 38(5), 538-559.

Walley, A.C. (2005). Speech Perception in Childhood. In Pisoni D.B., Remez R.E. (Eds.), *The Handbook of Speech Perception.* UK: Blackwell Publishing, 449-468.

Werker, J.F. (2018). Perceptual beginnings to language acquisition. In *Applied Psycholinguistics*, 39(4), 703-728.

Williams, A.L., McLeod, S. & McCauley, R.J. (2010). *Interventions for Speech Sound Disorders in Children*. Baltimore, MD: Brookes Publishing Company.

Zanobini, M., Viterbori, P. & Saraceno, F. (2012). Phonology and language development in Italian children: An analysis of production and accuracy. In *Journal of Speech, Language, and Hearing Research*, 55(1), 16-31.

Ziegler, J.C., Pech-Georgel, C., George, F. & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. In *Journal of experimental child psychology*, 110(3), 362-372.

Zmarich, C., Fava, I., Del Monego, G. & Bonifacio, S. (2012). Verso un "Test Fonetico per la Prima Infanzia". In Falcone, M., Paoloni, A. (Eds.), *Atti dell'8° convegno nazionale AISV*. Roma: Bulzoni Editore, 25-27.

Zmarich, C., & Bonifacio, S. (2005). Phonetic inventories in Italian children aged 18-27 months: a longitudinal study. In *Interspeech*, 757-760.

DOI: 10.17469/O2106AISV000023

ENRICO ZOVATO, VITO QUINCI, PAOLO MAIRANO

Modelling Sentiment Analysis scores and acoustic features of emotional speech with neural networks: A pilot study

Abundant literature has shown that emotional speech is characterized by various acoustic cues. However, most studies focused on sentences produced by actors, disregarding more naturally produced speech due to the difficulty in finding suitable emotional data. In our previous work we had performed an analysis of audiobook data in order to see if sentiment analysis could be of help in selecting emotional sentences from read speech. A regression analysis with Linear Mixed Models had revealed small effects, and the power of the models was low. We propose here an analysis with a neural network classifier predicting sentiment on the basis of acoustic cues, given the success of such models in the speech literature. However, the accuracy of the output was merely +0.13 above chance levels, suggesting that the different components used to express emotions (acoustic and lexical) tend to be complementary rather than additive, at least in audiobooks.

Key words: emotional speech, sentiment analysis, prosody, voice quality.

1. *Introduction*

The acoustic cues of emotional speech have been extensively described in the literature (Banse, Scherer, 1996; Burkhardt, Sendlmeier, 2000; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001). However, most studies are affected by methodological issues, such as the use of acted speech, often produced by actors or other professional speakers. In this study, we contribute to the study of emotions on "non-explicitly-elicited" speech by investigating the correlation between Sentiment Analysis (SA) scores and acoustic cues of emotions in English audiobooks. In particular, we take in consideration a set of acoustic cues which have been associated to emotional speech in the literature, and investigate their relation with SA scores extracted from text. In fact, SA techniques apply to written text, but can of course be used with speech transcripts. Previous studies in this topic are largely missing, and the few studies that explored this topic found weak correlations between acoustic and lexical cues of emotions: Charfuelan and Schröder (2012) analyzed data of one audiobook and found mild but significant correlations. Mairano, Zovato and Quinci (2018) analyzed data from 251 audiobooks, extending the scope of Charfuelan and Schröder's experiment, and found significant effects which explained only small amounts of variance. In this contribution, we re-analyze the data used in our previous study (Mairano et al., 2018) using a neural network classifier. Neural networks are a powerful statistical tool that has recently been applied in many domains of the speech literature. Given the success of this technique in making accurate predictions of complex social phenomena, and in accounting for non-linear relationships among variables, we hope that their use will improve the accuracy and the power of our models.

For this work, we used the same data as Mairano et al. (2018), i.e. the *LibriSpeech* corpus, more precisely the subset containing data suitable for ASR models training. This set is called train_clean_100 and is composed of 100 hours of speech (28539 sentences) in English read by professional and non-professional speakers, male and female. Scripts are extracted from 251 books.

2. *Sentiment Analysis*

Most SA systems work at sentence or document level and are based on lexicons. In our framework, SA scores were extracted at sentence level by means of open source tools, namely Vader (Gilbert, Hutto, 2014) and SentiWordNet (Baccianella, Esuli & Sebastiani, 2012). Vader provides three values:

1. an index of positive polarity, in the range 0-1,
2. an index of negative polarity, in the range of 0-1,
3. a composite index that takes into account the previous values, in the range [1,1].

These scores are extracted by means of contextual rules and of the Vader Lexicon, which allows to assign a score to each word in the sentence. For example:

What a wonderful day, full of happiness and joy!
0 0 +4 0 0 0 +3 0 +3 → **10**

After being sacked, he felt horribly depressed.
0 0 0 0 0 -3 -2 → **-5**

One of the authors manually annotated a subset of 1000 sentences of the corpus as positive, negative or neutral, and compared the manual classification with Vader scores. By considering scores < 0.2 as negative, -0.2 < score +0.2 as neutral, and score > +0.2 as positive, the agreement between the manual classification and Vader scores was 72%.

Here are some examples of SA scores associated with sentences available in the *LibriSpeech* scripts:

Table 1 - *Examples of sentence SA scores*

King of glory king of peace hear the song and see the star welcome be thou heavenly king	+0.93
The juniors forever hurrah fans hurrah our team is a winner	+0.90
Now again I feel that bliss to love one's neighbors.	+0.84
Since she too is a lady whom I love most tenderly	+0.80
The most peaceful and lovely thing he had ever seen	+0.82
That no doubt was why she hated him	-0.84
But was defeated in a fierce battle on the banks of the Daban canal	-0.82
How they had cruelly robbed and murdered poor people	-0.91
How fearful and dizzy tis to cast one's eyes so low I'll look no more	-0.84
He was so enraged at this that he again began war	-0.81

3. *Acoustic Analysis*

As described in Mairano et al. (2018), we extracted acoustic features at sentence and word level. Most of the studies in this field focus on sentence-level analyses, however it is likely that the strong semantic emotional connotation of some words can be reflected by certain acoustic cues. Word level acoustic features were calculated isolating the portion of signal around the stressed vowel. This could be done because the whole dataset was forced-aligned at word, syllable and phoneme level. We analysed all words in the sentence except function words.

Audio data were segmented and phonetically aligned. In fact, we automatically transcribed all the text data with a General American English grapheme to phoneme tool. A speaker-independent HMM model was used to align the waveform data to their phonetic labels and; eventually, the output was converted into *Praat*'s textgrid format (Boersma, Weenink, 2018) for the successive analysis and processing. In summary, the acoustic analysis consisted of:

- Grapheme to Phoneme conversion;
- Forced alignment;
- Extraction of acoustic parameters;
- Data pruning.

Feature representation plays an important role in developing emotion related applications. Previous studies (Banse, Scherer, 1996; Burkhardt, Sendlmeier, 2000; Schröder, Cowie, Douglas-Cowie, Westerdijk & Gielen, 2001), which extensively analyzed acoustic features, led us to select and calculate the following pitch features (*f0*) with a custom Praat script:

- F0 mean (in semitones);
- F0 stdev, standard deviation of F0 in semitones;
- F0 range (0.05 – 0.95);
- F0 max (0.95);
- F0 min (0.05).

F0 data were extracted with the autocorrelation method available in Praat and was executed in two steps. During the first iteration, fixed values of minimum and maximum F0 were used (75-400 Hz). The results were used to determine the inter quartile range (IQR). In the second step, the same analysis was run using a range between -25% and + 50% of IQR. Apart from F0 range, F0 min and max were included to explicitly define the level of F0.

Additionally, the following spectral features (*spec*) were extracted:

- Shimmer, as the perturbation in the amplitude of consecutive periods in voiced sounds;
- Jitter, as the perturbation in the duration of consecutive periods in voiced sounds;
- HAM, Hammarberg Index, as the difference of peak energy in the bands 0-2 kHz and 2-5 kHz (Hammarberg, Fritzell, Gauffin, Sundberg & Wedin, 1980);
- Do1000, as the linear declination of energy above 1000 Hz;
- Pe1000, relative energy of frequencies above 1000 Hz with respect to energy below 1000 Hz (Sherer, 1989; Drioli, Tisato, Cosi & Tesser, 2003);
- HNR, Harmonic to Noise Ratio;
- NHR, Noise to Harmonic ratio.

Beyond the features mentioned above, further duration parameters (*dur*) were extracted at sentence level, namely:

- DUR, total duration of the sentence from the beginning of the first phoneme to the end of the last one;
- SR, speech rate. The number of phonemes divided by the total length of the sentence, including the pauses;
- AR, articulation rate. The number of phonemes divided by the total length of the sentence, excluding the pauses;
- PSR, pause/speech ratio.

All the extracted parameters were converted into z-scores for each speaker (i.e. normalized with respect to the average and standard deviation values of the corresponding speaker). Different combinations of the abovementioned set of features (*f0*, *dur*, *spec*) were used in later experiments in order to isolate their effect on the classification models.

We discarded utterances whose duration was below 3 seconds. Moreover, all the acoustic parameters exceeding 2.5 times the standard deviation were excluded, as likely detection errors.

4. *Modeling SA scores and acoustic features*

We analyzed the acoustic and SA scores with two techniques. The initial analysis (reported in Mairano et al., 2018), was performed with linear mixed effects regression, using R and the *lme4* library (Bates, Maechler, Bolker & Walker, 2014) and revealed a weak linear relation between SA scores and acoustic features of emotional speech. Some effects were indeed significant, but the effect size was often extremely small. This was reflected in the power of the models, which was also disappointingly low. In this contribution, we present a follow-up experiment where we trained a neural network SA classifier, whose input was composed of the acoustic features.

The output of the model was tested on a specific test-set, which had not been used for training.

4.1 Sentiment Analysis Neural Net Classifier

Neural networks are powerful parallel processors that can acquire knowledge through observed data. They can be used both for regression and classification tasks (LeCun, Boser, Denker, Henderson, Howard, Hubbard & Jackel, 1990). Their effectiveness lies in the capability to model complex non-linear transformations between input and target data. The deepness in layers helps abstract the input layer information at increasing levels of complexity.

In our experiment we opted for supervised learning applied to multi-layer feed-forward networks, given the nature of input and output data. Training the network means determining the inner weights and biases that tend to reduce the error between predicted and observed data. The training is executed by using the backpropagation algorithm, which is composed of two phases: the forward phase, in which the weights are fixed, and the input is propagated through the network to calculate the error between the predicted and actual output; and the backpropagation phase, where the error is propagated through the network in the backward direction and weights are adjusted to minimize the error.

In our experiment we wanted to predict SA classes from input acoustic features. In other words, our goal was to find a transformation function from the acoustic level to the valence level (text based). We chose a relatively simple architecture, opting for a multi-layer feed-forward network, that is a graph in which all nodes of a layer are fully connected to the nodes of the preceding/following layer, and there are no recurrences. The transformation function can be written as:

(1) $$h(x) = g\left(\sum_i w_i\, h_i(x) + b_i\right)$$

where *g* is the output activation function, *i* is the layer index, are the layer functions, are the weights, and are the biases.

The network was composed of three feed-forward hidden layers and we used Rectified Linear Units (ReLU) as activation functions. On top of them, we stacked a softmax layer, which simply converts the output activation values to class probabilities which sum to 1. The softmax function is:

(2) $$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

The input of the network was composed of the acoustic features described in section 3. The targets were the SA categories, according to the compound sentiment scores classification:
- negative (Neg)
- neutral (Neu)
- positive (Pos)

Our model was tested on a portion of data excluded from training, in order to assess the accuracy of the model and its generalization capability.

Two classifiers were trained, the first with sentence-level features as input (SNT), the second with word-level features, extracted from lexically stressed vowels of content words (WRD).

The features used in WRD models were: DUR, HNR, NHR, F0 min, F0 max, F0 mean, F0 stdev, jitter, shimmer, Do1000, Pe1000, HAMM. For SNT models, AR and SR were also included.

A further distinction was made about the number of target categories: we considered three-class models (Neg, Neu, Pos), as well as binary classification with just Pos and Neg classes.

The network weights were trained on the whole train_clean_100 set of LibriSpeech data. During training, an early stopping criterion was adopted, on the basis of loss values of the validation set (30%), and on 10 consecutive non-improving epochs. Adam optimizer (Kingma, Ba, 2015) was used and the loss function was based on categorical cross-entropy. Input data were normalized to zero mean and unit variance. Data sampling was also applied in order to have a balanced number of occurrences among the three/two SA categories. A subset of the data (10%), which was not used for training, was later used as a test set for evaluation.

4.2 Results

We initially run an experiment in which all the acoustic features were used as input to the network model. We evaluated the models on the test set and results indicated low levels of accuracy levels, barely above chance level. This happens for both two- and three-class models. To some extent, these results confirm the outcomes of Mairano et al. (2018), indicating that correlations between the textual (SA) and acoustic levels are weak, at least in audiobook data. Accuracy, precision and recall values are reported in Table 2.

Table 2 - *Accuracy, precision and recall scores of SA by model*

	Accuracy	**Precision**	**Recall**
WRD-2	0.561	0.555	0.619
WRD-3	0.400	0.402	0.400
SNT-2	**0.638**	0.633	0.690
SNT-3	**0.462**	0.463	0.463

SNT-2 and SNT-3 models provide better accuracy than WRD-2 and WRD-3 models respectively. This indicates that sentiment scores are more predictable in wider contexts, or on global rather than local parameters. This can be also observed in the confusion matrices, that are a useful technique which describes the performance of a classification model. The rows and columns of these matrices represent respectively the predicted and true classes. The diagonal of a confusion matrix indicates the true positives (i.e. cases in which the predicted class matches the true class). As it can be observed in Figures 1, 2, 3 and 4, SENT models have a more "diagonal" pattern, despite the significant classification errors. In these figures the darker the color in the diagonal, the higher is the match between predicted and actual labels. Hence, we consider this model as the best performing (even though the accuracy score might point otherwise).

Figure 1 - WRD-3 *evaluation confusion matrix*

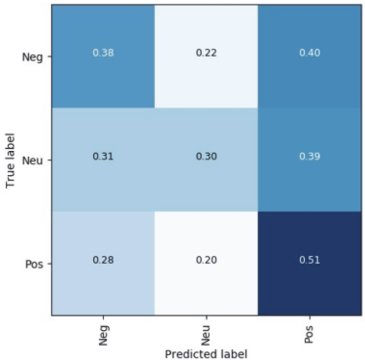


Figure 2 - SNT-3 *evaluation confusion matrix*

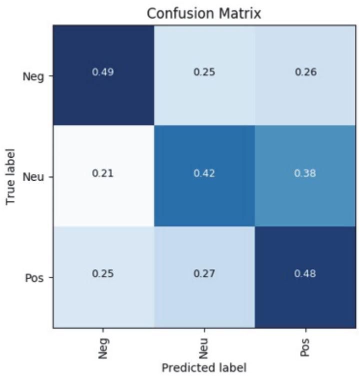


Figure 3 - WRD-2 *evaluation confusion matrix*

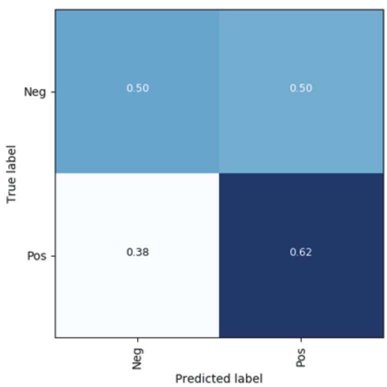
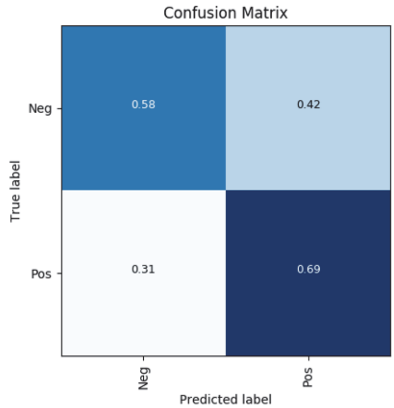


Figure 4 - SNT-2 *evaluation confusion matrix*



A second experiment analyzed different sets of input features applied to the SNT-3 models. In fact, according to the previous experiment, sentence models provided better results. Moreover, the 3-class model, despite its slightly lower accuracy than the 2-class model, (SNT-2 and SNT-3 are 0.138 and 0.129 above chance level, respectively), fulfills a more extended task, in line with more refined data classification. In this context we trained SNT-3 models whose input layer was composed of duration features (*dur*), pitch features (*f0*), duration+pitch features (*pros*) and spectral features (*spec*), as reported in Table 3. The motivation behind this experiment was to analyze the relative contributions of the features to model accuracy.

We tested the models on the same test set as the previous experiment. Results indicate that none of these variants reaches the accuracy of the "full set" model, meaning that all features bring relevant information for the prediction. The feature set that is closer in accuracy to the "full set" is the *pros* one: this result seems to confirm some results observed in linear mixed effects models in Mairano et al. (2018), according to which, prosodic features provided the higher correlation rates. On the other hand, duration and spectral features alone produce the worst results in terms of accuracy.

Table 3 - *Accuracy, precision and recall scores of SA SNT-3 models, trained on subsets of the acoustic features*

SNT-3	Accuracy	Precision	Recall
pros	**0.4397**	0.4407	0.4402
spec	0.3706	0.3711	0.3725
dur	0.3950	0.4288	0.3944
f0	0.4168	0.4167	0.4170

Table 4 - *Feature groups*

dur	DUR, SR, AR, PSR
f0	F0 mean, F0 stdev, F0 range, F0 max, F0 min
spec	Shimmer, Jitter, HAM, Do1000, Pe1000, HNR, NHR

5. *Discussion and Conclusions*

In this study we have analyzed audiobook data trying to find correlations between textual polarity, as provided by SA tools, and acoustic features extracted from the corresponding speech data. The results of our experiments have shown that SA classification from acoustic data is challenging. We have used neural network models which can learn complex transformations between input features and target categories. The underlying reason was to determine whether this modeling technique can provide better accuracy than other techniques used in past experiments (cfr. Mairano et al., 2018). Results, however, confirm the existence of a relation between pitch parameters and polarity text SA scores. Nonetheless, such relation is not strong enough to allow for a reliable prediction of SA categories. The prediction accuracy is merely 0.13 points above chance level in the best case (i.e. three-class sentence models). The correlation between lexical and acoustic cues of emotion seems thus to be limited in audiobooks, possibly because of a reduced involvement on the part of the speaker with respect to the content of the utterance. This seems to suggest that textual and acoustic cues are mostly complementary, rather than cumulative, in this type of data. If so, SA may not be well suited for the selection of emotional speech from audiobooks, or more generally for the study of emotions in speech.

The choice of audiobooks is dictated by the availability of many data resources in which a match between the textual content and an appropriate speaking style should occur. The problem, as we have shown in our experiments, is that either this kind of speech is not too much emotionally characterized or the SA tools we used are not adequate for capturing emotional content in the audiobook texts. Using specific databases of elicited emotional speech produced by actors or professional speakers would shed some light on this aspect. The risk is however that these data are too prototypical or overacted, sounding too distant from natural speech.

In future work we will try to extend the analysis to more data, taking into account audiobooks read by professional speakers. Another aspect to further investigate is the effect due to the sequence of words and their SA scores, as well as the sequence of acoustic features. In this regard, more complex neural network architectures will be considered, such as multi-layer recurrent networks with long-short term memory cells.

Bibliography

Baccianella, S., Esuli, A. & Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of LREC*, La Valetta, 17-23 May 2010, 2200-2204.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). Fitting Linear Mixed-Effects Models Using lme4. In *Journal of Statistical Software*, 67, 1, 1-48.

Boersma P., Weenink D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 3 February 2018 from http://www.praat.org/.

Burkhardt, F., Sendlmeier, W.F. (2000) Verification of acoustical correlates of emotional speech using formant-synthesis. In *SpeechEmotion-2000*, 151-156.

Charfuelan, M., Schröder, M. (2012). Correlation analysis of sentiment analysis scores and acoustic features in audiobook narratives. In *Proceedings of 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals ES3*, Istanbul, 26 May 2012, 99-103.

Drioli, C., Tisato, G., Cosi, P. & Tesser, F. (2003). Emotions and voice quality: experiments with sinusoidal modeling. In *Proceedings of Workshop the Voice Quality: Functions Analysis and Synthesis (VOQUAL)*, Geneva, 27-29 August 2003, 127-132.

Gilbert C.J., Hutto E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, Ann Arbor, MI, 2-4 June, 2014.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities In *Acta Otolaryngologica*, 90, 441-451.

Kingma, D.P., Ba, J. (2015). Adam: A Method for Stochastic Optimization. In *Proceedings of ICLR*, San Diego, CA, 7-9 May 2015.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. & Jackel, L.D. (1990). Handwritten digit recognition with a back-propagation network. Advances. In *Neural Information Processing Systems*, 2, 396-404.

Mairano, P., Zovato, E. & Quinci, V. (2018). La sentiment analysis come strumento di studio del parlato emozionale?. In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it*, Turin, 10-12 December 2018.

Scherer, K.R. (1989). Vocal correlates of emotion. In Manstead, A., Wagner, H. (Eds.), *Handbook of psychophysiology: Emotion and social behavior*. London: Wiley, 165-197.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M. & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings EUROSPEECH 2001 – Seventh European Conference on Speech Communication and Technology*, Aalborg, 3-7 September 2001, 87-90.

PARTE III

TECNOLOGIE DEL PARLATO

DOI: 10.17469/O2106AISV000024

FRANCESCO CANGEMI, JESSICA FRÜNDT, HARRIET HANEKAMP, MARTINE GRICE

A semi-automatic workflow for orthographic transcription and syllabic segmentation

Manual orthographic transcription of spontaneous speech is notoriously time consuming, and segmentation at the level of the syllable requires further large amounts of processing time. Automatic orthographic transcription and automatic syllable segmentation, on the other hand, usually yield unsatisfactory precision, especially when applied to spontaneous speech. In this contribution, we report on a semi-automatic workflow that uses freely available software and combines the speed of automatic processing with the quality of manual transcription and segmentation. We apply the procedure to highly spontaneous speech and show (i) no loss of quality compared with manual output and (ii) an average 75% reduction of manual processing time.

Key words: orthographic transcription, forced alignment, syllabic segmentation, spontaneous speech, automation.

1. *Introduction*

The first step in the phonetic analysis of spontaneous speech is often represented by an orthographic transcription and a syllabic (or phone-level) segmentation of the recordings. Both operations can be performed either manually or automatically. When performed by a trained phonetician, fully manual transcription and segmentation can be highly reliable, but they have a high cost in terms of time and effort. For certain speech types (such as an excited conversation between friends), accurate manual transcription and segmentation can require up to one hour for each minute of speech. On the other hand, fully automatic transcription and segmentation can immensely reduce the workload placed on the analyst's shoulders, but they are also comparatively less reliable in terms of output quality. Whereas read speech can be automatically transcribed and segmented with remarkable results (Vagnini-Holbl, Draxler, 2018), the performance of automatic procedures remains unsatisfying for spontaneous speech, where words can bear little resemblance to the canonical forms used in building speech recognition systems. Similarly, performance of automatic procedures can be satisfying for languages with extensively trained models (such as English), but is comparatively poorer for understudied languages, for regional varieties and for foreign accented speech. As a result, researchers often combine automatic and manual workflows when analysing spontaneous speech, for example by providing a manual transcription first, and then using the transcription as input for the automatic segmentation. In this paper, we detail a semi-automatic workflow in which an orthographic transcription is generated automatically, then manually corrected. The corrected transcription is then used as input for an automatic segmentation tool, whose output is then once more manually verified. This procedure yields an output as precise as a reference manual segmentation, but requires significantly shorter manual processing time (in average, down to 25% of the time required for a fully manual annotation). Crucially, this semi-automatic procedure only uses software which is compatible with all operating systems, freely available, and easy to use.

2. *Method*

2.1 Material

In order to test the semi-automatic transcription and segmentation workflow, we recorded a spontaneous conversation between two speakers. We used two head-mounted microphones (AKG C544L) connected through an audio-interface (Focusrite Scarlett 6i6) to a laptop running a digital audio workstation (*Reaper*, Cockos, 2018), with a sample rate of 44100 Hz and a bit depth of 24 bit. This setup is highly portable and relatively inexpensive, but it allows for professional recording quality and matches the standard practices in phonetics research.

The setting and content of the recording, on the other hand, are comparably less usual. Recordings took place at the first author's home, taking only basic precautions for maximising sound quality: windows were closed and drinking glasses were provided with a straw, but no curtains or rugs were used to reduce reverberation, and there was no physical barrier between the two speakers' microphones. This caused voices bleeding into the other speaker's channel, introduced a certain amount of background noise, and included reverberation into the recordings. Nonetheless, the achieved audio quality was deemed acceptable, especially if considering that the unusual recording setup (at home, having drinks, with no barrier between interlocutors) had a positive effect on the naturalness of the interaction.

To further increase naturalness, the recordings capture two good friends (the first author and another phonetician) as they play a videogame. The game offers an enjoyable task, which is also often performed for mere recreational purposes, and is thus substantially different from most other tasks used in phonetic data collection. Compared to recordings where participants are asked to perform tasks they would normally not perform in their daily lives (such as reading aloud a list of words, complete dialogues based on pictorial context, et cetera), speaker involvement is notably higher in this interaction. As a result, most of the recordings sound spontaneous and relaxed, as attested by the relatively high frequency of colloquialisms and profanities captured by the microphones.

In order to test our semi-automatic workflow on maximally challenging material, we introduced a further obstacle beyond the sub-standard recording conditions and the colloquial speech style: none of the two speakers is a native speaker of the language used in the interaction, which is thus a mixture of Italian-accented and Hebrew-accented English. Moreover, by referencing to in-game characters and actions which bear custom-made names (e.g. *Taurukh* for a centaur-like creature), speakers often used lexical items that are absent from automatic speech recognition dictionaries, and are thus treated as non-words.

In summary, despite the use of relatively standard recording equipment, we voluntarily hindered the performance of the semi-automatic workflow through poor recording conditions, extreme spontaneity of the interactional situation, use of non-native English, and abundance of non-words. On the other hand, this resulted in speech samples that feel remarkably genuine, with both speakers (including the experimenter) reporting having lost track of the scientific purpose of the recordings early on during the gaming session. Nonetheless, in order to avoid confounds, for the following analyses we will not employ speech uttered by the first author, and will only use 8 sound files (of 1 minute each) extracted from the 29 minutes long recording of the Hebrew-accented English speaker.

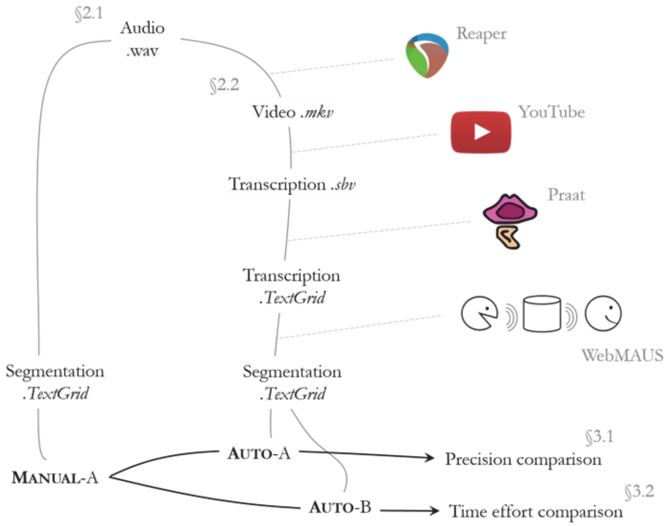
2.2 Method

Each of the 8 audio files was submitted to both a manual and a semi-automatic workflow. In the manual workflow, the third author used *Praat* (Boersma, Weenink, 2018) to create orthographic transcription and manual syllable segmentation following standard best practices (Machač, Skarnitzl, 2013). We will refer to these files as MANUAL-A (manual annotation, annotator A, i.e. the third author).

In the semi-automatic workflow, the second author prepared audio files for upload onto *YouTube*. Since *YouTube* only accepts video files, audio files were combined with an image in *Reaper* (Cockos, 2018). The file was rendered with uncompressed (maximal) audio quality. With the consent of the speaker, the video files were uploaded to a private channel on *YouTube*, and a first pass of the orthographic transcription was obtained using the *YouTube* automatic transcription function. The automatic transcription contains timestamps for suggested beginning and end of each interpausal unit. The transcription was exported in one of the available subtitles formats (sbv), and transformed via script into a *Praat* annotation file (TextGrid). Errors in the automatic transcription (including imprecise timestamps) were manually corrected in *Praat*. Audio and annotation files (wav and TextGrid) were then further processed using *WebMAUS* (Kisler, Reichel & Schiel, 2017), one of the tools available on the webpage of the Bavarian Archive for Speech Signals. *WebMAUS* is a web-based forced alignment service, which takes audio and orthographic transcription files as input, and returns phone-level segmentation as output. We used the so-called *G2P-MAUS-PHO2SYL pipeline*, which provides a phonologisation of the orthographic transcription (*G2P*), a phonetic segmentation (*MAUS*), and the reconstruction of syllables based on the phonetic segmentation (*PHO2SYL*); see Kisler et al. (2017) for details on these three modules. The output TextGrids were further processed by eliminating the tiers generated by *WebMAUS* as intermediate steps, and by only saving the orthographic transcription and the syllable segmentation. Using these simplified TextGrids as starting point, the second author manually corrected the syllabic boundaries automatically placed by

WebMAUS, thus completing the semi-automatic workflow. We will refer to these files as AUTO-B (semi-automatic annotation, annotator B, i.e. the second author). In addition, the simplified TextGrids were also submitted to the third author for manual correction, but only after she had already provided the final TextGrids for the manual workflow. We will refer to these files as AUTO-A (semi-automatic annotation, annotator A, i.e. the third author).

Figure 1 - *Diagram of workflows, including annotators and comparisons*



2.3 Performance comparisons

The use of three segmentations (MANUAL-A, AUTO-A and AUTO-B) aimed at countering annotator bias in the comparison of the two workflows (cf. Figure 1). When focussing on the *time effort* required by the two workflows, we compare MANUAL-A with AUTO-B, since a comparison between MANUAL-A and AUTO-A would introduce order effects (i.e. presenting the same audio files twice to the same annotator would lead to an advantage for any workflow performed last). Similarly, when comparing the *precision* of the two workflows, we compare MANUAL-A with AUTO-A, in order to avoid the unescapable differences between the segmentation habits of the two annotators. This approach has the additional advantage of avoiding the pitfalls of promoting a manual annotation to the rank of reference segmentation (cf. Cosi, Falavigna & Omologo, 1991).

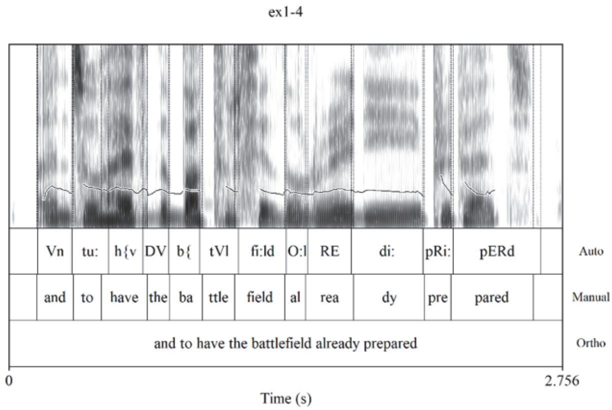
Since the focus of the present paper lies with a first description of the semi-automatic workflow as tested on a limited speech database, we will refrain from using inferential statistics in the comparisons reported below.

3. *Results*

3.1 Precision comparison

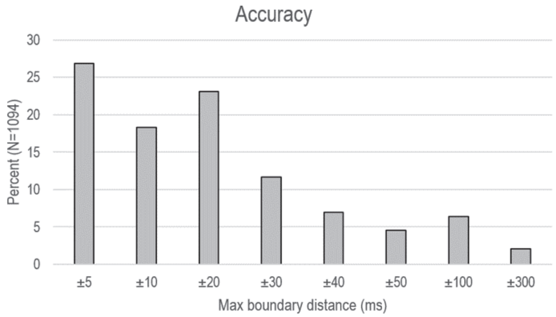
In order to assess the precision of the two workflows, we compared the position of syllabic boundaries in the two segmentations provided by the third author (MANUAL-A vs. AUTO-A). Figure 2 shows spectrogram and pitch track for a small portion of one of the test sound files. In the annotation panel are visible the orthographic tier (Ortho) and syllabic tiers for both manual workflow (Manual) and semi-automatic workflow (Auto, in SAMPA). A visual inspection suggests that the two workflows yield virtually undistinguishable quality.

Figure 2 - *Example of manual and automatic segmentations*



In order to quantify this assessment, we extracted the timestamps of 1094 syllables for each of the two segmentations (Manual and Auto), and calculated distances between matching boundaries across the two segmentations. Figure 3 shows the distribution of maximal temporal distance (in milliseconds) between boundaries in the automated and manual workflows, using negative numbers for late automatic boundaries. The figure shows that 45% of automated-workflow boundaries fall within ±10ms of boundary in the manual workflow; 68% of boundaries fall within ±20ms of boundary in the manual workflow, indicating highly comparable performances. Approximately half of automated-workflow boundaries fall at the maximum level of precision available for automatic (and human) segmentation (±10ms), while less than 9% of boundaries fall in a ±50ms window.

Figure 3 - *Precision comparison*

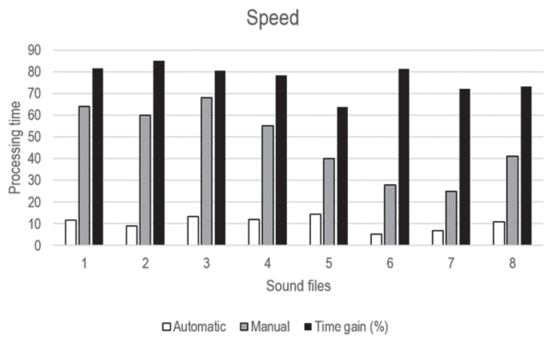


3.2 Time effort comparison

Figure 4 shows the crucial comparison between the processing times required by the two workflows, as performed separately by the two annotators (Manual-a vs. Auto-b). Processing times are expressed in minutes for each of the 8 test files (on the x-axis), separately for the Automatic workflow (white bars) and in the Manual workflow (grey bars). The 8 test files were excerpted from relatively varied moments in the interaction, with some samples featuring 1 minute of excited monologue, and other samples capturing 1 minute of a relaxed dialogue. As such, they can be more or less challenging to the manual and automatic annotator. It is thus unsurprising that the processing times vary greatly from sample to sample. In the Manual workflow, processing times vary from 25 to 68 minutes, in line with the expectations for orthographic syllabic segmentation of highly spontaneous non-native speech. Processing times are equally varied for the Automatic workflow, ranging from 5 to 14 minutes.

Crucially, however, the percent time gain when using the Automatic workflow is less variable, as shown by the black bars. Depending on the files, the advantage can range between 64% and 85%, suggesting that, in average, the automated workflow requires a fourth of the processing time required for manual annotation.

Figure 4 - *Time effort comparison*



4. *Discussion*

The preliminary evidence presented above suggests that, compared to the manual workflow, the semi-automatic workflow provides virtually undistinguishable precision, with a substantial processing time reduction. These results are particularly encouraging, considered that the materials used in this test were maximally challenging due to the spontaneous nature of the interaction, the familiarity between the speakers, the use of non-native English, the sub-standard recording conditions, and the use of a large number of non-words (see §2.1). Applying the semi-automatic procedure on native read speech with lemmatised words would surely require less manual processing time.

Manual correction of the orthographic transcription posed no particular challenge, with the exception of false starts and non-words. Syllable segmentation seemed to be more affected by the spontaneous nature of the interaction. For example, manual correction was required for several cases of long frication noise in a turn-opening 'so', a common phenomenon in conversational speech.

Despite these challenges, the semi-automatic workflow for orthographic transcription and syllabic segmentation presented here reduced processing time by approximately 75%. Importantly, it only required multiplatform and free software, and it was applied to extremely challenging speech, recorded with inexpensive and portable equipment.

Bibliography

Boersma, P., Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Retrieved from www.praat.org.

Cockos (2018). *Reaper* [Computer program]. Retrieved from www.reaper.fm.

Cosi, P., Falavigna, D. & Omologo, M. (1991). A preliminary statistical evaluation of manual and automatic segmentation discrepancies. In *Proceedings of Eurospeech*, Genova, 1991, 693-696.

Kisler, T., Reichel, U. & Schiel, F. (2017). Multilingual processing of speech via web services. In *Computer Speech & Language*, 45, 326-347.

Vagnini-Holbl, C., Draxler, C. (2018). Comparing acoustic measurements from manual and automatic segmentations. Paper presented at *Phonetik und Phonologie im deutschsprachigen Raum*, Vienna, 6-7 September 2018.

Acknowledgments

This work was partly supported by the German Research Foundation through the Collaborative Research Centre "Prominence in Language" (SFB 1252).

DOI: 10.17469/O2106AISV000025

GIUSY GAUDINO, MARIA DI MARO

Conversing on Artistic Representation Topics (CART): An android audio guide with dialogic skills

The present work describes the architecture of an interactive audio guide (CART) specifically conceived for the Neapolitan Charterhouse "San Martino", as part of a project, whose aim is to collect and model multimodal data for the development of spoken technologies for museum fruitions. In addition to the cultural contents that an audio-guide can provide, this application is enriched with a spoken dialogue system, a system interacting with humans by means of spoken language commands and requests. The application was developed using Android Studio. Furthermore, the developed interface for the Museum was endowed with dialogic skills modelled in DialogFlow using its Android Software Development Kit (SDK), to which Text-to-Speech technologies were integrated to record the user query. The results are promising in that they display positive feedback for the introduction of such technologies in museums.

Key words: spoken dialogue systems, oral archive applications, audio guide, cultural heritage, android, dialogflow.

1. *Introduction*

Several authors have described the act of experiencing as a conversation: «it always involves at least two voices» and its «meaning is always changing as a consequence of what comes next» (McCarthy, 2008). The act of experiencing is indeed a dialogue of the self with the other, being it another human being or not. Consequently, museum experience is describable as a dialogic exchange with the artistic product where the conveyed messages are mainly emotional: the artwork, in fact, communicates emotions first, which are differently perceived according to the characteristics and preconception of the experimenting subject. The perceived emotions are then empowered at a deeper stage of analysis with cultural contents once the museum experience became informative and educational.

The presence of tour guides in historical places or in cultural sites traces back to the Greek Empire, when the Periegeta (gr. περιηγέω [*periegeo*] 'I show around') used to show the visitors places new to them. In this scenario, the dialogic aspect of the experience becomes physical, since the influence of the interlocutor on the main speaker can be explicitly communicated via spoken requests or multimodal feedback of understanding (i.e. nodding, smiling, vocalizations, etc.). Besides tour guides, at the end of the XVIII century, the art dealer Tommaso Puccini introduces little tags indicating the name of the author of each artwork in the Uffizi Gallery, representing the first informative 'technological' means provided to visitors (Kaplan, 2014). The technological progress has brought, in some cases, to the substitution of tour guides with audio guides. The first audio guide was introduced in the Milan Cathedral by the electronics technician Giovanni D'Uva in 1959 and was called the "audiophone". This device enabled users to listen for the first time to informative audios without the physical presence of an expert (Kaplan, 2014). Nowadays, the use of audio guides is extremely widespread, resulting in a transformation of the museum experience, which is becoming less dialogic and more monologic, and, therefore, passive. Interestingly, some scholars pointed out how the inclusion of recorded dialogues in audio guide devices can be actually used as an engagement strategy for museum visitors (Elliston, Fitzgerald, 2012). Specifically, creating audio guides using expert-expert spontaneously generated dialogues was shown to better engage listeners with learning contents. Surprisingly, the resulting users' engagement caused visitors to increase the amount of spoken interactions among them. On the other hand, with the introduction of mobile phones in our daily life, different services started to migrate on these devices. The CIMI (Computer Interchange of Museum Information) studied the possible application of mobile computing in museums, ranging from virtual guides, to electronic maps, communication channels, and virtual diaries for collecting visitors' impressions (Gay, Spinazze & Stefanone, 2002). In this scenario, Economou and Meintani (2011) underline the power of the exploitation of mobile apps in museums, for the possibility they create to reach new audiences through personal and familiar devices. Furthermore, those tools would enable life-long learning and edutainment (i.e. game-based apps) as they could be used before, during and even after the visit. The authors' evaluation of the available systems points out that the transition from the previous audio guides to the new multimedia devices has not brought any changes in the design of the museum experience. Conversely, the aim should concern the design of a system, which is considered as an active tool in the content creation process.

For all the aforementioned reasons, we believe that users' engagement could be further increased if the audio guide would become interactive, enabling the user to personalize the visit according to their interests and curiosities. Consequently, in this work we propose to integrate a dialogue system – i.e. a conversational agent interacting with human users – in a museum guide app to enhance the visitors' experience. In the second Section, we are going to focus the attention on the description of the system architecture, specifying the developed interfaces of our app, the modules created for the different provided services, and the requested requirements. The third Section is dedicated to the interaction modelling, where we pay attention to the description of the interaction design principles and the corpus-based content extraction. In Section 4, we describe the test sessions and the collected users' feedback, which are going to be part of the future developments of our app.

2. *Experimental Work*

2.1 CART App

CART is an interactive audio guide that allows visitors to interact with the audio guide itself with the aim of asking questions about artworks and history of a cultural site. The integration of a conversational agent is the true innovation of this product that adds to the static information provided from traditional audio guides the possibility for the users to deepen some topics of interest, just as they would do with a human guide. For the implementation of the contents of the audio guide, the transcriptions of experimental guided tours within the San Martino Charterhouse in Naples were used. The guided tours were recorded for the data collection projected for the Italian national project CHROME (Cultural Heritage Resources Orienting Multimodal Experiences), whose goal is to define a methodology to collect, to analyse, and to model multimodal data in designing virtual agents serving in museums[1] (Cutugno, Dell'Orletta, Poggi, Savy & Sorgente, 2018). The cultural sites chosen for the CHROME project are the Charterhouses in Campania. A trivia on the name of our application is indeed that it takes back the official logo of the *Chartusian order* that had four crossed letters, a C, an A, an R and a T (from the initials of Carthusia – Latin for Chartreuse, the place near Grenoble where the first Charterhouse was born). Therefore, the app's logo is inspired to the Chartusian one.

The front-end architecture of the application is divided into four different interfaces:

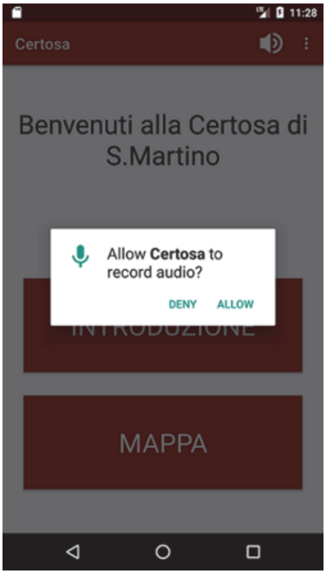
a. Introduction;
b. Map;
c. Description;
d. Dialogue interface.

Furthermore, the application uses a module for operating the Text-to-Speech service and another control module (cfr. Section 2.7) to manage data required by the application.

The architectural choice was designed to maximize the application's usability and the software modularity at the same time. The reduced number of interfaces has the effects to allow the user to have a good learning curve and to facilitate browsing the application. Furthermore, the ordered display of the interfaces is designed to guide the user to a progressive visit of the tour stages without, however, restricting its freedom and autonomy to explore the place as they prefer. The application was designed for newbies, for this reason it follows learning and memorability principles (Polillo, 2010).

[1] http://www.chrome.unina.it.

Figure 1 - *Introduction screen asking for "record audio"*



2.2 Introduction Interface

The introduction interface (Figure 1) is the main interface of the application. Deliberately plain, it drives the user to discover the app and the place of the visit. The two provided buttons (Introduzione, and Mappa) lead respectively to the first descriptive screen (first step of the tour) and to the map screen where an overall view of the tour is provided. In Figure 2 is shown a pop-up asking for "recording audio" permission (cfr. Section 2.8), later required to interact with the vocal assistant. In view of future developments, this screen could provide options for selecting the language to be used for visitors of different nationalities.

2.3 Map Interface

The map screen (Figure 2) shows to the user a general distribution of the points of interest (POIs) described within the app. This is useful to the visitors to get oriented and to verify their position according to the description being provided. From this screen, the descriptive screen of all tour's stages can be accessed.

Figure 2 - *Map screen*

2.4 Description Interface

The description interface (Figure 3) takes up the classic functionality of the audio guide, giving information on the single step of the visit. There is an image that can help the visitor to recognize the place in which they should be while watching a specific description interface, or rather the described place. The description is both textual and vocally synthesized. Everywhere in the application the audio option can be enabled and disabled and the preamble interface, the map or the other steps of the tour can be reached. Through this interface, one can easily pass from the description of the current stage to the next or previous one with the side arrows (respectively) ">" and "<". There is a rating-bar on the screen, with a score from 1 to 5 stars that allows users to leave a feedback about the quality of the description related to that place. This function can be used to collect the user satisfaction within the app fruition, in order to improve the different sections. On the bottom right, the icon of a microphone, is strikethrough if the user doesn't grant microphone access permission, enabled otherwise. This icon, if clicked, give access to the dialogue interface, where users can ask questions to the virtual assistant about the current section they are seeing.

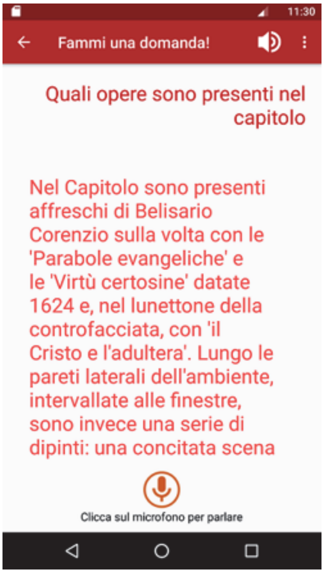
Figure 3 - *Description interface*

2.5 Dialogue Interface

The dialogue interface (Figure 4) handles the true innovation of this digital audio guide that supplies the service offered from traditional audio guides with querying services, peculiar of human guides. The screen recalls the model of the classic vocal assistants commonly integrated in smartphones, such as Google Assistant. By clicking on the microphone icon an audio feedback is provided to let users know that the application is starting to listen, and their input can therefore be processed. A different feedback audio informs the user, that their question has been received. The online service DialogFlow[2] (cfr. Section 3.3) is consequently questioned, and the question-answer pair appears on the screen well distinguished by different colours and different alignments (respectively the question on the right and the answer on the left). If the user doesn't disable the audio service, the answer will be acoustically synthetized. If an Internet connection is not available, the query cannot be submitted to the DialogFlow service. This information is given to the user through a pop-up message.

[2] https://dialogflow.com/.

Figure 4 - *Dialogue interface*



2.6 Text-to-Speech Manager

This module is responsible for correct loading and setting of the voice synthesizer used in the application. The vocal synthesis is useful for two reasons: i) to reproduce the POI's description (as for the traditional audio guide) ii) to vocally synthesize the answer of the assistant (Hoy, 2018). The acoustic description of the POI would for sure sound more natural with a recorded file audio rather than with a speech synthesiser. Recognizing this, the choice fell on speech synthesis because of the scalability and the generalization of the application's content. In fact, despite the changes concerning the application's contents, vocal synthesis grants an acoustic reproduction of information of interest, making the service much more versatile.

2.7 Information Module

"Information Module" is the focal point for the modularity of the application. It lets the application to scale in size and it is the point where all the information showed to the user are collected. It is the only section where the developer can edit the theme of the application. By editing the information contained and saving the aesthetic design of the application, the same design can be fitted to any other museum, creating an interactive audio guide ready to be used. For each step of the tour, the stored information comprises: i) chronological order of the point of interest (POI) one is visiting (as part of the general tour); ii) name and description; iii) an image of the place one is visiting; iv) a link to the vocal assistant trained on this particular argument (step of the tour).

2.8 Required Permissions

An access permission to the microphone is required in order to interact with the vocal assistant. The request is carried out at the first access to the application and, if denied, it is requested again when the user tries to access to the dialogue interface, as this screen is not available when the permission is denied. Furthermore, an internet access is required in order to interact with the vocal assistants integrated in the app and for the Speech-to-Text service whose results are used to query the vocal assistant.

2.9 Used Services

The app uses DialogFlow service that allows the building of conversational interfaces. To include the DialogFlow service into the Android application the API dialogflow-android-client "api.ai"[3] has been used. The API includes Speech-to-Text services that turn the user's vocal input into textual information and submit the information achieved in the DialogFlow service through an HTTP request. The answer obtained is shown to the user in a textual way and worked out from a Text-to-Speech service, specifically the one offered by Google and integrated for Android.

3. *Dialogue Management*

3.1 Conversation Design

In this section, particular attention is drawn to the development of the interaction module integrated in the application. Before adding contents to the app, design principles were defined in order to build an interaction, which could be easily carried out by human users, with no need to teach them how. The first rule chosen for the conversation design is *Identify your user*. This rule is essential when designing interfaces of any kind, as it can be used to circumscribe what the interface is actually going to do, considering technological constraints as well. Knowing what kind of users is going to make use of the interface can help the designer to predict which contents have to be included in it. For the definition of the user and its needs we adopted the questionnaire results collected during the collection phase of the CHROME corpus (cfr. Section 3.2). For the CHROME project, audio-visual recordings of gatekeepers during tour guides were collected. Immediately after the recorded visits, the recruited visitors were asked to compile a questionnaire composed of 23 items including both Likert scale evaluations and open answer questions. The items were designed to collect anagraphic data, self-evaluations of artistic competence, evaluations of personal satisfaction, and evaluations of the gatekeeper's performance (Origlia, Savy, Poggi, Cutugno, Alfano, D'Errico, Vincze & Cataldo, 2018). Therefore, the questionnaire was used to get an idea of the possible visitors' profiles and, since open questions

[3] https://github.com/dialogflow/dialogflow-android-client (last access 03/10/2019). The v1 version was used, since v2 version does not support Android.

about what they liked the most or what they would like to hear from a guide were included, even topic preferences could be modelled.

The second rule that was taken into account is *Give a reason why*. With this principle, we wanted the agent to guide the interaction by stimulating the interlocutor to ask questions. In order to trigger questions in the users we followed two different strategies. On the one hand, the strategy of the unsaid (Todd, 2013) was exploited to stir curiosity in the users: instead of giving all the information to them in the different description screens (cfr. Section 2.2.4), some details were left pending with rhetorical or provocative questions, and highlighting pieces were introduced without giving explanations, as in (1).

(1) In uno degli affreschi (?) del Parlatorio è ritratto lo stesso Sant'Ugo (?) mentre dorme su un letto a baldacchino. Su di lui aleggiano sette stelle (?).

In one of the frescos of the Parlor (?), St. Ugo is represented while sleeping on a canopy bed (?). Seven stars hover on him (?).

In (1), the interrogation points represent the missing details that can be asked to the system, and specifically Which other frescos are situated in the Parlor? Who is St. Ugo? What's the name of the fresco representing St. Ugo? What does it mean? What do the seven start represent? On the other hand, curiosity was guided through direct questions, options, and class of topics to further be investigated (2).

(2) Vuoi saperne di più sul Parlatorio? Vuoi conoscere il significato del ritratto di Sant'Ugo? Vuoi passare al Coro?

Do you want to learn more about the Parlor? Do you want to know the meaning of the painting representing St. Ugo? Do you want to move on to the Choir?

There is another designed rule connected to the second strategy, named *Provide help*. This means that, when the user does not know what to ask, they can rely on options given by the agent.

The fourth regulation principle for the conversation design is *Be adaptable*. With this rule, we refer to the necessity to adapt to different sentence structures that can be uttered by the users. Therefore, the first important step, after the definition of the semantic category to develop, was to define the possible questions a user could pose to ask for information for each specific category (cfr. Section 3.3).

As far as the interaction style was concerned, other important rules were adopted. The first one was *Be relevant*, that is do not give more information than the one needed. When a question could be answered with more details, an in-depth option was used (cfr. Section 3.3). On the other hand, it was also found to be important to *Be conversational*. In fact, in order to sound more natural, no formal words were used, and second-person references were adopted. Furthermore, chitchat intents were included to build this conversational perception (cfr. Section 3.3). Especially as far as the chitchat intents were concerned, we also followed the *Use variations* rule, randomizing answers when appropriate. Moreover, with the aim of not only entertaining but also educating the user, the *Be clear* principle was adopted, as ambiguity or obscure words were avoided. In fact, one important goal to reach when designing *infotainment* devices is to be clear and be sure that the contents are well received by the users.

Particularly important was also to react promptly to scenarios where the agent could not understand the user input. *Error handling* is indeed another principle adopted. With the use of DialogFlow (cfr. Section 3.3), this aspect was automatically taken into account, as rephrasing requests were activated by the agent (i.e. *I'm not sure I understand your question. Could you repeat?*), therefore avoiding assigning a wrong category to the misunderstood user input.

In the next two sections we are going to explain in detail the contents of the CHROME Corpus used as a reference corpus for the application development and how the contents were modelled using DialogFlow. The rules here displayed are then going to be enriched with the feedback collected from the testing phase (cfr. Section 4).

3.2 CHROME Corpus

In order to achieve the main aim of the CHROME project[4] (cfr. Section 2.1), data need to be collected, analysed and annotated. In fact, the first step taken was to record audio-visual material comprising real gatekeepers giving tours to four groups of four visitors in the Neapolitan Charterhouse "San Martino". The visit is divided into six points of interest (POIs), selected as the most relevant parts of the Charterhouse from an architectural and artistic point of view, specifically *Pronaos*, *Great Cloister*, *Parlor*, *Chapter Hall*, *Wooden Choir*, *Treasure Hall*, corresponding to the description screens selected for the app development (*Pronaos* was left out). The collected data are going to be analysed and annotated multimodally by using the ELAN software (Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006). In fact, gestures and speech are going to be annotated: on the one hand, gestures are analysed as far as their textual or interactional goals are concerned; on the other, linguistics features are pointed out, as far as orthographic, phonetic, syllabic, disfluency, macro-syntactic, syntactic, and intra-syntactic levels (Origlia et al., 2018). For the development of the app, these multimodal annotation levels were not taken into account, as only contents and user preferences were exploited (cfr. Section 3.1).

3.3 DialogFlow Web Services

Following the principles explained in Section 3.1, the interaction was built within the DialogFlow framework. DialogFlow is a Google platform that enables designer to exploit Google's Machine Learning algorithms and Natural Language Processing techniques to easily create conversational agents. DialogFlow provides a powerful natural language understanding (NLU) engine to process and understand natural language inputs. "DialogFlow allows to create *agents* that can understand the vast and varied nuances of human language and translate them to standard

[4] http://www.chrome.unina.it.

and structured meaning that a service can understand". As different users can express the same meaning differently, as far as syntax and vocabulary are concerned, DialogFlow parse the requests in a way that each information is mapped to the pertinent data in order to find the appropriate response[5].

The semantic and conceptual category corresponding to the classes of questions that the agent is capable of answering to are named *intents*. Each intent is made up of training sentences given to the system to learn to recognize the intention of the user which can be differently expressed (cfr. *Be adaptable* in Section 3.1), and of textual, visual or multimodal answers (i.e. text, cards, links, pictures). Training sentences are annotated with *entities*, representing the semantic classes to which words belong. For instance, given the entity *artist*, different lexical items can be used to express it, such as *painter*, *sculptor*, *architect*, *author*, *creator*, *designer*, *inventor*, *master*. The use of such annotations is advantageous to improve disambiguation processes and to correctly classify an intent that can be uttered with different terms according to the user. Follow-up intents can also be added to ask for details concerning the previous intent or for yes/no confirmations.

The conversation was designed for the five of the six POIs reported in Section 3.2. For each of the selected POIs specific intents (41 in total) were developed, as reported in Appendix A.

4. *Testing and Results*

CART functionality was tested by 10 users. The test phases are divided in two different steps. At first, users were provided with a task accomplishment assignment, for which they were asked to compile information requested in a questionnaire. In details, the questionnaire was made up of four tasks: i) collect information about the fresco in the Treasure Hall Cupola (i.e. name of the fresco, name of the artist, characters represented, meaning of the fresco), ii) collect missing information about the Carthusian in the Great Cloister and in the Parlor screens, iii) collect information about Wooden Choir frescos, iv) converse freely with the agent.

The second part of the test was a usability questionnaire given to users to measure the Quality of Service (QoS) and the Quality of Experience (QoE), where QoS is used to evaluate the effectiveness of the system in giving the right information and QoE to capture the user perception of interaction quality (Fiedler, Hossfeld & Tran-Gia, 2010). The evaluation was divided in 5 sections: i) utility, ii) ease of use, iii) ease of learning, iv) satisfaction, v) comments. The evaluation was computed through a five value Likert Scale (Never, Rarely, Sometimes, Often, Always). The results, as shown in the figures 5-8, are generally positive in all users.

[5] https://dialogflow.com/docs/intro.

Figure 5 - *Utility Results*

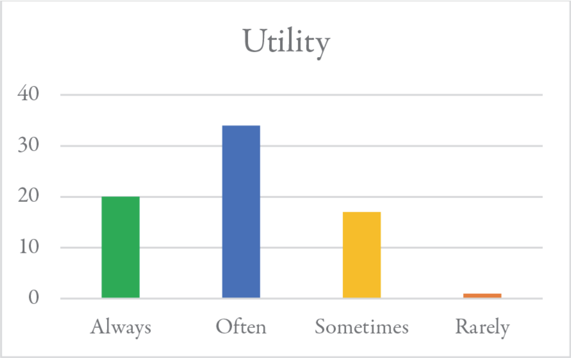


Figure 6 - *Ease of use Results*

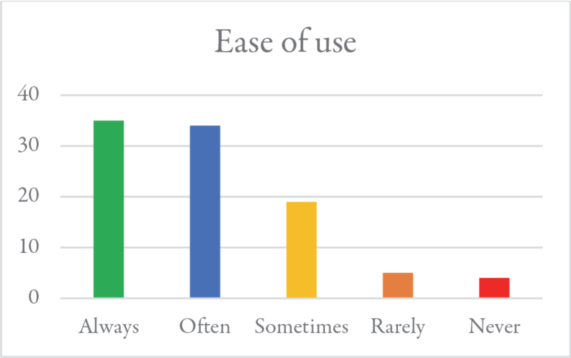


Figure 7 - *Ease of learning Results*

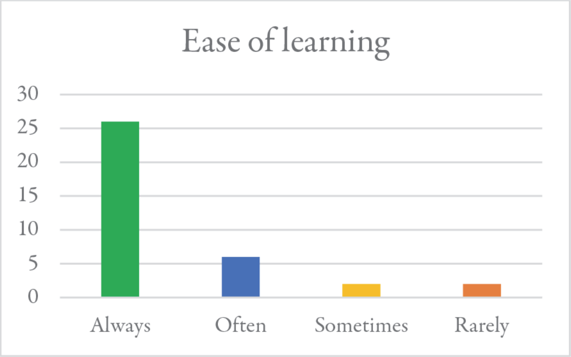


Figure 8 - *Satisfaction Result*

Nevertheless, the application needs further development as suggested in the comments collected from the open question of the evaluation questionnaire. In fact, user's feedbacks can help to improve the definition of conversational design principles that may be used for next implementations of CART. One important need highlighted by the users may be referred to one of the Grice's Maxims (Grice, Cole & Morgan, 1975) that was not entirely followed in the conversation design of CART, namely the maxim of quantity, or in other words *Be efficient*. Users signalled the necessity to get vivid and significant contents, which are not too long-winded. Indeed, the attention of users tends to decrease when listening at someone talking without having that someone in the same room (as for a conversational agent which is not embodied). The necessity to provide relevant and detailed information to users who are to be educated on a particular subject collides with the necessity of expressing a concept using few and simple words. This contradiction can be solved by using shorter descriptions for which a *read/hear more* option can be added. Among the other feedback we mention the need to model more intents, as some information asked were missing. This means that the maxims of *Identify your user* and *Be adaptable* should be further improved. This necessity is going to be easily fulfilled also thanks to the conversation which were collected during the testing phase. In fact, DialogFlow permits to save the user inputs, enabling the improvement of the intents design. Moreover, it has interestingly been noticed that users find multimodality as an important expected need in such interfaces. As a matter of fact, they suggested that links to events organised in the descripted site, or links to further details, or more images or even videos should be added within the interaction interface. This leads to another important principle that is *Conversation is multimodal.* In fact, no matter with whom or what one is interacting, because the demand for the multimodal conveyance and/or reception of meanings will be always on the spotlight.

5. *Conclusions and Future Work*

The present paper reported on the development of an Android interactive audio guide thought for the museum fruition. The interactivity of the application is concerned with the possibility for users to ask further information besides the descriptions provided within different sections of the guide. For the development of the interaction module DialogFlow was used. The dialogue was modelled following specific design principles in order to ensure the naturalness and the success of the interaction. The testing phase pointed out that further principles need to be considered for the improvement of the audio guide, specifically as far as efficiency, adaptability, and multimodality are concerned. In details, we aim at enriching the intents collection of our agent, starting from user inputs collected during the testing phase. Furthermore, multimodal functionality, such as pictures, links, buttons, and videos, need to be added in our framework. In fact, some information needs visual support to be better received by users. Furthermore, in order to give users a more fluent interaction with the app and allow them to pause and play the audio description of the place (vocally synthesized), like a traditional audio file, the description text could be chopped into more sentences which could be sequentially synthesized. Indexing the sentences and tracking the last played one, the impossibility of "pause" a vocally synthesized text could be therefore overcome and the interaction could increase in naturalness.

Acknowledgment

This work is funded by the Italian ongoing PRIN project CHROME - *Cultural Heritage Resources Orienting Multimodal Experience* (Cutugno et al. 2018) #B52F15000450001. We also thank Valentina Lipari who is working on the improvement of CART.

Bibliography

Cutugno, F., Dell'Orletta, F., Poggi, I., Savy, R. & Sorgente, A. (2018). The CHROME Manifesto: integrating multimodal data into Cultural Heritage Resources. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLIC-IT 2018)*, Turin, 10-12 December 2018.

Economou, M., Meintani, E. (2011). Promising beginning? Evaluating museum mobile phone apps. Rethinking Technology. Paper presented at *Museums 2011: Emerging experiences*, University of Limerick, 26-27 May 2011.

Elliston, B., FitzGerald, E. (2012). Encouraging museum visitor engagement using spontaneous talk-in-interaction audio guides. Paper presented at the *4th International Conference on Computer Supported Education (CSEDU 2012)*, Porto, 16-18 April 2012.

Fiedler, M., Hossfeld, T. & Tran-Gia, P. (2010). A generic quantitative relationship between quality of experience and quality of service. In *IEEE Network*, 24, 2, 36-41.

GAY, G., SPINAZZE, A. & STEFANONE, M. (2002), Handscape: Exploring potential use scenarios for mobile computing in museums. In *Cultivate Interactive*, 8, 15.

GRICE, H.P., COLE, P. & MORGAN, J.L. (1975). *Logic and conversation*. In COLE, P., MORGAN, J.L. (Eds.), *Syntax and Semantics vol. 3. Speech Acts*. New York: Academic Press, 41-58.

HOY, M.B. (2018). Alexa, siri, cortana, and more: An introduction to voice assistants. In *Medical reference services quarterly*, 37, 1, 81-88.

KAPLAN, F. (2014). Fantasmagories au musée. In *Alliage*, 73, 132-140.

MCCARTHY, J., CIOLFI, L. (2008). Place as Dialogue: Understanding and Supporting the Museum Experience. In *International Journal of Heritage Studies*, 14, 3, 247-267.

ORIGLIA, A., SAVY, R., POGGI, I., CUTUGNO, F., ALFANO, I., D'ERRICO, F., VINCZE, L. & CATALDO, V. (2018). An Audiovisual Corpus of Guided Tours in Cultural Sites: Data Collection protocols in the CHROME Project. In *Proceedings of the AVI-CH 2018 Workshop on Advanced Visual Interfaces for Cultural Heritage*, Castiglione della Pescaia, 29 May 2018.

POLILLO, R. (2010). *Facile da usare. Una moderna introduzione all'ingegneria della usabilità*. Milan: Apogeo.

TODD, S. (2013). *Learning desire: Perspectives on pedagogy, culture, and the unsaid*. London: Routledge.

WITTENBURG, P., BRUGMAN, H., RUSSEL, A., KLASSMANN, A. & SLOETJES, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Genoa, 24-26 May 2006, 1556-1559.

Appendix A

In this appendix, the list of intents developed for each POI is displayed.

a. *Great Cloister*: Cemetery, Fanzago_info, Cloister_function, Carthusian_day, Carthusian_meal, Carthusian_crest, Small_talk.
b. *Parlor*: Bruno_Calabria, Bruno_Italia, Bruno_born_dead, Bruno_studies, StUgo_Painting_meaning, StUgo_Painting_name, Parlor_artwoks, Parlor_info, Silence, Heremitage, Small_talk.
c. *Chapter Hall*: Chapter_info, Chapter_artworks, Chapter_Name_meaning, Lunette_artworks, LayBrother, Prior, Preator, Small_talk.
d. *Wooden Choir*: Choir_info, Choir_artworks, Carthusian_day, Carthusian_crest, Women_ban, Small_talk.
e. *Treasure Hall*: Ribera_painting, Ribera_info, Fapresto_meaning, Giordano_info, Giuditta_info, Fresco_name, Fresco_meaning, Treasure_info, Small_talk.

DOI: 10.17469/O2106AISV000026

PHILIPPE MARTIN

Prosodic Annotation of oral archives

Prosodic analysis of oral archives is often hampered by adverse recording conditions, making popular pitch tracking algorithm fail in many cases, so that gathered data may be useless. To address these limitations and allow prosodic research to be conducted on almost any kind of ancient or recent recordings showing less than optimal acoustic conditions, several dedicated functions have been integrated in the software program WinPitch. A first set of these functions allow the user to apply alternate pitch tracking algorithms among 7 available on selected signal segments, as some algorithms may perform better than others. If this process appears still unsatisfactory, a set of easy to use graphical commands are available to directly annotate pitch curves graphically, relying on an underlying narrow-band spectrographic display, whose frequency scale is automatically matched with the fundamental frequency curve scale. Piecewise linear graphic lines can be adjusted to melodic movements of any complexity as displayed on the spectrogram. Furthermore, the phonetic or phonological categories of these user placed annotations can be automatically labelled according to predefined classes, such as those available in ToBI notation system, or as melodic contours indicating dependency relations between stress groups.

Key words: Oral archives, prosodic annotation, sentence intonation, fundamental frequency.

1. *Introduction*

Oral archives found in old or not so old newsreels, radio and TV recordings are often characterized by less than optimal recording quality, making them unsuitable for prosodic research. In particular, the reduced speech bandwidth due to the then available technology for sound capture and conservation (carbon microphone, vinyl records, optical and magnetic tape sound recording, etc.) makes intonation analysis difficult and unreliable for some pitch tracking algorithms, such as the ones found in the popular software program Praat. Most analysis problems are linked to reduced signal intensity, lack of speech fundamental frequency in the spectrum, presence of echo and overlapping voices or sound sources, etc.

To address these limitations and pursue prosodic research on valuable speech archives, two sets of dedicated functions have been implemented in the software program WinPitch. Both sets rely on visual inspection and comparison of the pitch curve against the simultaneous displayed narrow-band spectrogram first (or second) harmonic, whose frequency scale has been aligned on the pitch curve scale. Using one of the seven tracking algorithms available in WinPitch (autocorrelation, AMDF, spectral comb, spectral brush, Cepstrum…, see Martin, 1981), the user is able to visually check the validity of a given pitch track section against the corresponding spectrogram first harmonic segment, select graphically the F0 curve section of interest and apply an alternate algorithm on the selected section. As some algorithms perform better than others depending on the specific conditions of recorded segments, a more satisfactory F0 curve can be obtained, as possibly assessed by the underlying spectrogram first harmonic. Usually, switching between spectral comb and autocorrelation gives satisfactory results, the ancient AMDF method being surprisingly efficient when the other two algorithms fail.

When the situation is desperate, i.e. when no available pitch tracking method appears reliable enough, another set of user-friendly commands using only mouse controls allow to position graphic piecewise linear segments approximating any shape of the fundamental frequency curve. These graphic annotations can be user-labelled in 14 programmable categories (type, color, segment thickness, etc.) and once positioned on screen have their characteristics in time and frequency transferred to any spreadsheet program such as Excel in one single mouse click.

The user-defined classes can be adapted to virtually any phonetic or phonological model of sentence intonation, using for instance either the ToBI notation system or melodic contours notation.

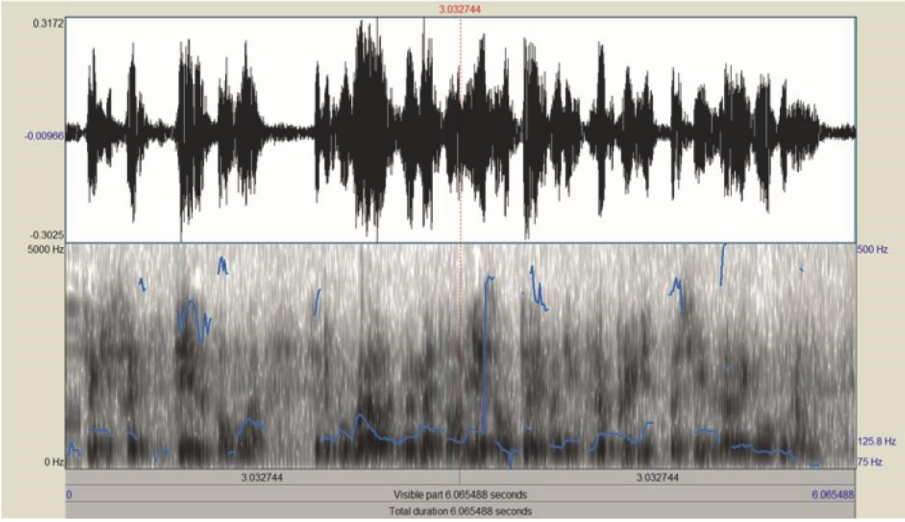
2. *Underlying spectrogram*

Although not frequently used, one of the best ways to validate the fundamental frequency curve of a given speech segment consists to compare the melodic curve with the first harmonic of the corresponding narrow-band spectrogram. Alternatively, the second or even the third harmonic can be used if the first is not visible, has been filtered or is obscured by another sound source spectral component. This is achieved in WinPitch with a single mouse click, aligning both spectrogram and fundamental frequency curve scales.

Except in rare case of a very fast changing fundamental frequency (as in in the transition from a voiced stop to a high intensity vowel), the spectrum first harmonic is totally reliable, although not allowing a precise measurement of the frequency value due to the frequency resolution of the spectrogram (in the order of 20 to 30 Hz).

In this domain of research, many users may not be aware that commonly available pitch tracking algorithms can display erroneous results and trust blindness the displayed fundamental frequency curves. Actually, a simple comparison with an underlying narrow-band spectrogram would give them some hints pertaining to the reliability of the acoustic analysis. Unfortunately, in many examples using Praat for instance (Fig. 1), when a spectrogram is displayed together with a pitch curve, it is by default a wide-band where harmonics and the fundamental component cannot be discriminated.

Figure 1 - *An example of pitch analysis of a noisy recording with Praat. Ce qui serait utile pour moi c'est de pouvoir me mettre en liaison téléphonique le plus vite possible avec mon gouvernement. "What would be useful for me is to be able to get in touch with my government as quickly as possible"*



3. *Automatic segmentation*

WinPitch integrates an automatic text to API segmentation and alignment system. Contrary to other realizations (e.g. EasyAlign, Maus…), the alignment algorithm proceeds by alignment with a reference generated by an embedded TTS system for each text segment. This approach integrates specific handling for liaisons and vowel linking specific for French. It allows the Viterbi based alignment to operate for more than 42 different languages (German, Italian, Spanish, European Portuguese, Swedish, Danish, Polish, Russian, Japanese, Korean, Mandarin…) based on the available TTS system in Microsoft Windows, so that there is no need to create new Gaussian models for each language as in EasyAlign. This is especially critical in the case of old speech recordings, where statistical approximation of phone properties is rarely valid. The automatic segmentation implemented in WinPitch generates automatically a segmentation layer words and API labelled phone units (Fig. 2).

Figures 3 to 5 illustrate the process applied to a recording made secretly on June 22nd 1940 by the German Wehrmacht in Rethondes. The French general Huntziger is negotiating the armistice while the German invasion of France is progressing. Fig. 3 shows the pitch segments manually placed along the spectrogram narrow-band first harmonic, and automatically labelled according to predefined criteria (terminal, major and minor continuation, neutralized contour).

Figure 3 displays the prosodic structure as indicated by the melodic contours of Fig. 2, and Figure 4 gives an example of the spectrogram output corresponding to the example.

Figure 2 - *An example of prosodic annotation of a noisy recording. Segments of the melodic contours are graphically annotated to fit both the pitch curve and the corresponding spectrogram first harmonic. Ce qui serait utile pour moi c'est de pouvoir me mettre en liaison téléphonique le plus vite possible avec mon gouvernement. "What would be useful for me is to be able to get in touch with my government as quickly as possible"*

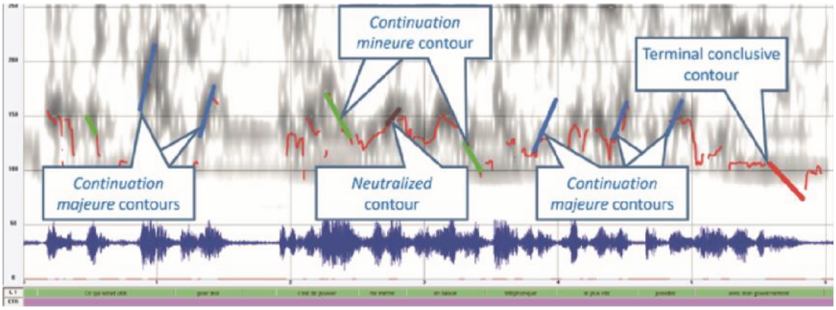


Figure 3 - *Automatic generation of the prosodic structure for the example of Fig. 3, illustrating the contrast of melodic slope characterizing French. Each contour indicated a dependency relation towards a contour situated later in the sentence and at a higher level in the Cneu-> Cfal-> Cris-> Cdec hierarchy*

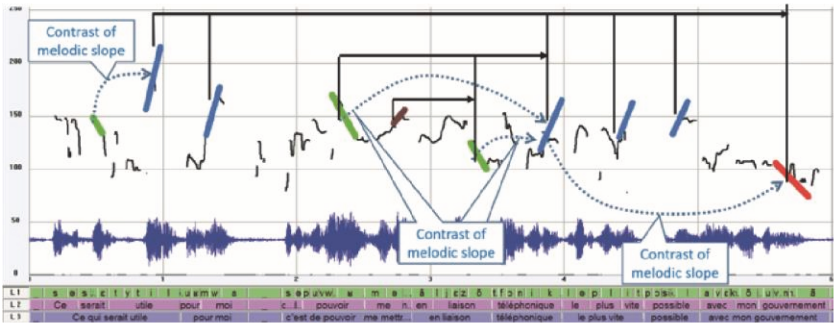


Figure 4 - *Output of the data on a spreadsheet giving the actual values the fundamental frequency, intensity and duration values of stress group segments values*

1	File	Contour	Nb	Time [s]	F0 [Hz]	Int [dB]	Dur [ms]	Fmin [Hz]	Fmax [Hz]	Diff [Hz]	L 1
2											
3											
4	Rothondes 1940	C2	1	0.475	149	34 S	69	134	149	15	Ce qui serait utile
5				0.544	134	23 S					
6											
7	Rothondes 1940	C1	2	0.873	156	23	109	156	215	59	Ce qui serait utile
8				0.982	215	35					
9											
10	Rothondes 1940	C1	3	1.322	132	35	103	132	177	45	pour moi
11				1.426	177	35					
12											
13	Rothondes 1940	C2	4	2.265	170	35 S	175	131	170	39	c'est de pouvoir
14				2.441	131	27 S					
15											
16	Rothondes 1940	Cn	5	2.723	143	27 S	86	143	156	13	me mettre
17				2.810	156	27 S					
18											
19	Rothondes 1940	C2	6	3.300	125	27 S	118	100	125	25	en liaison
20				3.418	100	28 S					

4. *Prosodic annotation*

Prosodic annotation is performed directly on screen, by matching the displayed curve segments with a piecewise linear curve placed on screen by the user with the mouse and automatically encoded in color and classified along the user predefined classes. These user-defined categories can be adapted to virtually any phonological model of sentence intonation, using for instance either the ToBI or melodic contours notation.

In the incremental prosodic structure model for example (Martin, 2018), sentence intonation uses melodic contours categories *Cdec*↓ (terminal falling and low for declarative sentences), *Cris*$\boxed{\nearrow}$ (rising above the glissando threshold), *Cfal*$\boxed{\searrow}$ (falling above the glissando threshold) and Cneu→ (rising or falling below the glissando threshold. The corresponding phonological category is automatically assigned by the software to the segment placed by the user, together with the color coding eventually assigned beforehand. The glissando values (Rossi, 1971) are evaluated with the formula St2-St1/t2 – t1, with the threshold 0.32 (St2-St1) / (t2 – t1)^2, St1 and St2 are the beginning and end values of the contour in semitone units, and t2-t1 is the duration of the melodic contour.

Using the ToBI notation symbols such as H*H%, L*L-, L+H* for FToBI, it is equally easy to predefine as actual acoustic melodic movements are then approximated in terms of tone targets (Delais-Roussarie, Post, Avanzi, Buthke, Di Cristo, Feldhausen, Jun, Martin, Meisenburg, Rialland, Sichel-Bazin & Yoo, 2015). Targets can be acoustically predefined, with values of fundamental frequency jumps (in semitones) for example, and assigned automatically or manually.

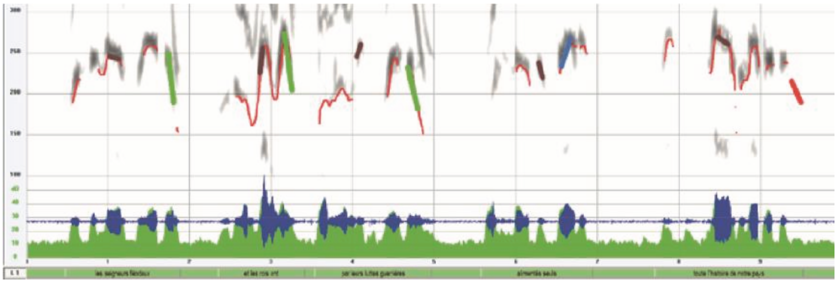
5. *Some oral archives*

The described integrated system has been applied to the analysis of some number of speech archives, both old and recent. Examples pertaining to the changes occurred in the realization of melodic contours in French speech archives (early XX century period) are given below, showing the capability of the system in degraded recording conditions to illustrate the evolution of prosodic style in political and news speeches of these periods.

Available speech archives recorded before 1918 are essentially characterized by the use of tremolo segments by the speakers, but apparently only used in very formal circumstances. Such tremolos realizations are not found in the same period in non-formal situations, where the contrasts of melodic slope, found as well in contemporary recordings, was already implemented in the phonological system of French.

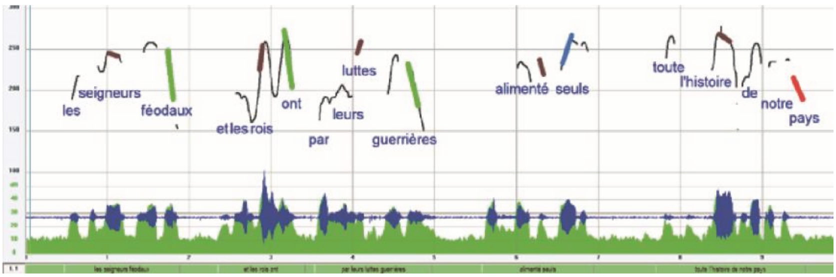
A first example pertains to Henry Le Châtelier (1850-1936), recorded in 1910 (Fig 5 and 6). *Les seigneurs féodaux et les rois ont par leurs luttes guerrières alimenté seuls toute l'histoire de notre pays* "The feudal lords and kings have by their war fights fueled alone the entire history of our country".

Figure 5 - *Recording of H. Le Châtelier in 1910: Les seigneurs féodaux et les rois ont par leurs luttes guerrières alimenté seuls toute l'histoire de notre pays. "The feudal lords and kings have by their war fights fueled alone the entire history of our country"*



Since the fundamental frequency is almost totally absent from the narrow-band spectrogram, it is the second harmonic that makes it possible to validate the plot of the melodic curve obtained by the spectral comb method.

Figure 6 - *The recording of Fig. 5 aligned with text, showing the contrast of melodic slope and the absence of tremolo found in a more formal speech style*



The annotated melodic contours show the incremental merging of successive stress groups forming the overall sentence prosodic structure: [[*Les seigneurs Cneu→ féodaux Cfal*$\boxed{\searrow}$] [*et les rois Cn→ ont Cfal*$\boxed{\searrow}$] [*par leurs luttes Cn→ guerrières Cfal*$\boxed{\searrow}$] [*alimenté Cneu→ seuls Cris*$\boxed{\nearrow}$]] [*toute l'histoire Cneu→ de notre pays Cdec*↓]

The melodic slope contrasts are constantly realized by the speaker, with an unusual segmentation in the group *et les rois ont*, and there is no trace of tremolo as found in speech formal recordings of the same period as shown below.

By contrast, the analysis of a speech given by Fernand Brunot in 1911 shows a large use of pitch tremolo on stressed vowels, where the phonological system of French would expect a *continuation majeure*, instantiated by a rising melodic contour above the glissando threshold (Fig. 7 and Fig. 8). Ferdinand Brunot was professor of linguistics at the Sorbonne University in Paris from 1900- to 1934. He became very interested in the available speech recording techniques of his time, and especially to speech pronounced by ordinary people. The segments analyzed in Fig. 8 and Fig. 9 come from his inauguration speech given on June 3rd, 1911 for the inauguration of the *Archives de la parole* he created with Émile Pathé.

Figure 7 - *Ferdinand Brunot in 1911... de tant de beauté de gloire et d'espérance de tant d'accord si doux d'un instrument divin... "so much beauty of glory and hope of so much sweet agreement of a divine instrument..."*

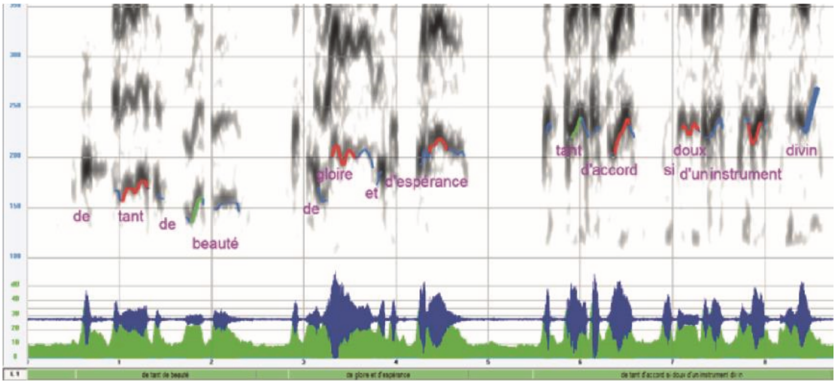
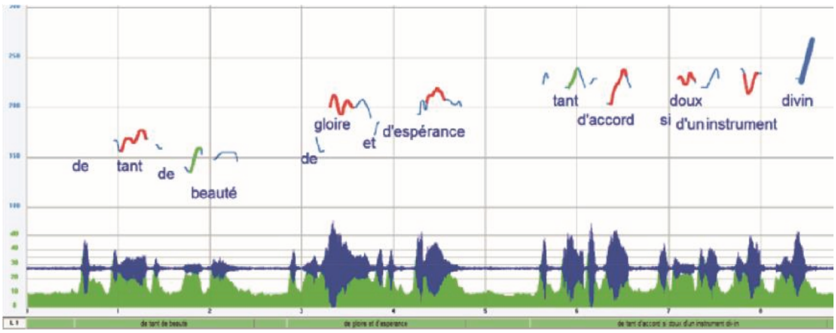


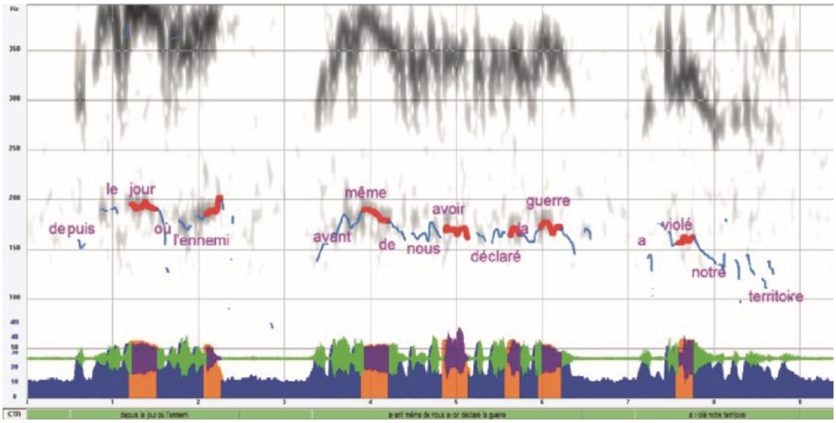
Figure 8 - *Ferdinand Brunot in 1911... de tant de beauté de gloire et d'espérance de tant d'accord si doux d'un instrument divin... showing the large use of tremolos on stressed vowels*



Rising emphatic stress on the first syllable of *beauté* as well as the first syllable of the group *tant d'accord*. Despite the presence of many syllables realized with tremolos, Brunot sill produces a first *Cris* contour on the last syllable of the sentence.

The following example of formal speech (Fig. 9) also illustrate the use of tremolos. The speaker is Paul Deschanel then president of the *Assemblée nationale* at the time of the recording, just before the beginning of the first World War.

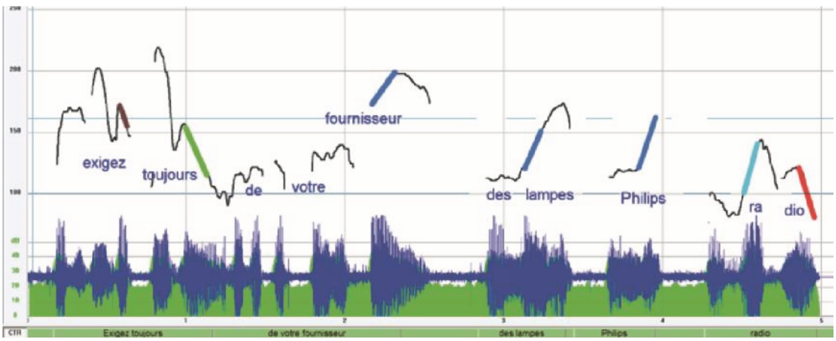
Figure 9 - *Paul Deschanel 1914. Depuis le jour où l'ennemi avant même de nous avoir déclaré la guerre a violé notre territoire. "Since the day when the enemy even before we declared war has violated our territory"*



Without tremolos, the realization with melodic contours would have been: *Depuis le jour Cfal* ↘ *où l'ennemi Cris* ↗ *avant même Cfal* ↘ *de nous avoir Cneu→ déclaré Cneu→ la guerre Cris* ↗ *a violé Cneu→ notre territoire Cdec*↓

Fig. 10 gives an example of radio advertising recorded in 1928, where the contrast of melodic lope is clearly illustrated.

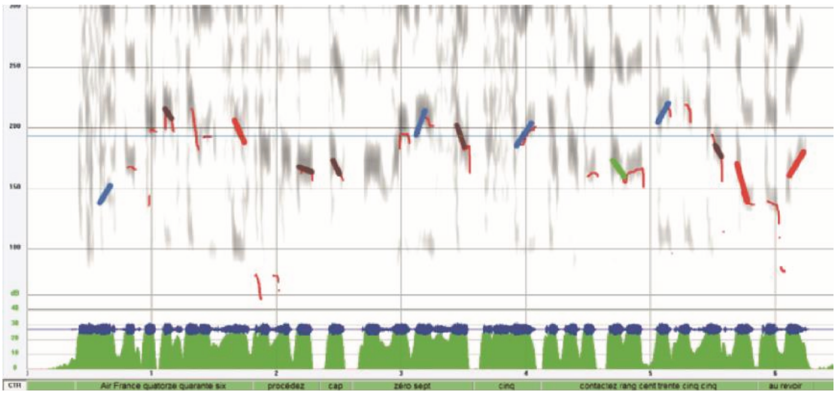
Figure 10 - *Radio advertising in 1918. Exigez toujours de votre fournisseur des lampes Philips radio. "Always ask your supplier for Philips radio lamps"*



The tone of voice is clearly non-formal in this example. The sequence of melodic contours on stressed vowels shows a contrast of melodic slope on *Exigez toujours Cfal* ↘ *de votre fournisseur Cris* ↗, whereas the first syllable of radio bears an emphatic stress (always rising) on the first syllable.

Figure 11 provides a last example, an air traffic control recorded in 1996, while more recent, shows adverse recording conditions for prosodic research. In this example, the fundamental frequency is almost totally absent in the signal, due to a severe filtering of components below 300 Hz (as in old telephone land lines).

Figure 11 - *An air traffic control recorded in 1996. Air France Cris*$\boxed{\nearrow}$ *14 Cneu →46 Cdec↓ procédez Cneu→ cap zéro Cris*$\boxed{\nearrow}$ *sept cinq Cris* $\boxed{\nearrow}$ *contactez rang cent trente cinq cinq Cdec↓ au revoir* $\boxed{\nearrow}$. *"Air France 1446 proceed cap zero seven five contact rank one hundred thirty five five goodbye"*



6. *Conclusions*

The oldest speech recordings available in French indicate that, if some phonetic realization of melodic contours did change in some 100 years (e.g. no more penultimate stress, ex.: on suffixes *-ation*, rare tremolos after 1918 in formal speech), the phonological contrast of melodic slope, a feature specific to French, did not change and was already effective at this time.

The prosodic annotation function implemented in WinPitch allow for the analysis of desperate cases, where no fundamental frequency tracking algorithm seems to be suitable.

Although it requires some basic expertise from the annotator, the semiautomatic annotation constitute a valuable approach for the study of old and degraded speech archives. WinPitch is available at www.winpitch.com.

Bibliography

Delais-Roussaire, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S-A., Martin, Ph., Meisenburg, T., Rialland, A., Sichel-Bazin, A. & Yoo, H-Y. (2015). Intonational phonology of French: Developing a ToBI system for French. In Frota, S., Prieto, P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 63-100.

EASYALIGN. http://latlntic.unige.ch/phonetique/easyalign.php.

Martin, Ph. (1981). Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne. In *Proc. 12e Journées d'Etude sur la Parole, Montréal, 1981*, 221-232.

Martin, Ph. (2018). *Intonation, structure prosodique et ondes cérébrales.* London: ISTE.

MAUS. https://www.bas.uni-muenchen.de/Bas/BasMAUS.html.

PRAAT (2019). Computer program, www.praat.org.

Rossi, M. (1971). *Le seuil de glissando ou seuil de perception des variations tonales pour la parole.* In *Phonetica*, 23, 1-33.

WINPITCH (2019). Computer program, www.winpitch.com.

DOI: 10.17469/O2106AISV000027

ALESSANDRO RUSSO, NICCOLÒ PRETTO, ANTONIO RODÀ,
SERGIO CANAZZA

L'informatica per la gestione e la conservazione di informazioni acustiche (musica e voce)

The lack of a standard preservation methodology is an important issue that digital speech archives have to face nowadays. Information that needs to be stored requires the implementation of a methodological framework in order to prevent cultural losses. The Centro di Sonologia Computazionale (CSC) of the University of Padova, cooperating with several digital archives (both speech and music) implemented a scientific methodology for the active preservation of audio documents. The framework is still a work in progress and the methodology is continuously improved with new additions, such as neural networks and machine learning-based techniques for detecting the right equalization curves or recognizing discontinuities on the tape. This contribution presents the abovementioned methodology, as well as several informatics tools for the study, the access and the preservation of audio documents.

Key words: Audio documents, Digitization, Preservation, Oral archives.

1. *Introduzione*

Negli ultimi anni è emersa in maniera sempre più evidente la necessità di individuare una metodologia condivisa per la digitalizzazione e la conservazione dei documenti sonori. Numerosi archivi, infatti, affrontano il problema adattando i metodi tradizionali utilizzati per la gestione del materiale cartaceo, i quali, seppure ormai ampiamente consolidati e idonei nel caso di beni librari, mostrano i loro limiti e la loro inadeguatezza nel caso di documenti complessi e di difficile lettura quali i documenti sonori. Infatti, oltre al segnale audio che costituisce l'informazione primaria, un documento sonoro è costituito da un insieme di numerosi elementi, che vanno dal supporto fisico a tutta una serie di informazioni aggiuntive (scatola, note dell'autore, segni sul nastro, ecc.) di grande importanza e che, in quanto tali, devono essere preservate per ottenere una copia digitale che rispecchi fedelmente l'originale.

Gli ultimi 20 anni sono stati interessati da una vera e propria "corsa alla digitalizzazione", in quanto, grazie al progresso tecnologico e all'introduzione di nuovi formati digitali, per la prima volta è stato possibile vedere nelle tecniche di salvaguardia digitali una soluzione efficace per la conservazione di svariate tipologie di beni culturali (IFLA/UNESCO, 2000). Questo era precedentemente impossibile a causa delle limitazioni offerte dalla tecnologia esistente, che comportavano un'inevitabile e non trascurabile perdita di qualità nel passaggio da analogico a digitale, ragione per la quale la digitalizzazione era utilizzata solo ai fini dell'accesso, ma non della conservazione a lungo termine dei materiali (Smith, 1999).

Questo processo è nato dall'esigenza di mettere rapidamente in sicurezza e salvaguardare un patrimonio ad alto rischio, soprattutto a causa di problematiche inerenti al degrado dei supporti analogici, all'obsolescenza dei macchinari utilizzati per riprodurne i contenuti e a un'inevitabile perdita di conoscenze e competenze relative al loro utilizzo, sempre più ristrette a una piccola cerchia di appassionati e tecnici del settore (Davies, 2007).

In questo processo, qualora non si tenga conto dell'importanza del documento sonoro nella sua interezza, si rischia di andare incontro a una enorme perdita di informazioni. I laboratori del Centro di Sonologia Computazionale (CSC) del Dipartimento di Ingegneria dell'Informazione (DEI) dell'Università di Padova si occupano da decenni di salvaguardia e digitalizzazione di documenti sonori di interesse storico, artistico e culturale, principalmente musica elettronica ed elettro-acustica, ma anche registrazioni di parlato ed etnomusicologia. Nella messa a punto di una metodologia per la conservazione dei documenti sonori, una grande attenzione è stata data alla raccolta dei metadati e ai supporti analogici, i quali, specialmente nel caso della musica elettronica, devono essere considerati parte integrante dell'opera stessa e come tali devono essere tutelati e conservati al pari dell'informazione audio registrata al loro interno. I documenti sonori si differenziano dai beni culturali realizzati attraverso altre forme di espressione artistica, quali per esempio la pittura e la scultura, dove l'oggetto di interesse non può essere separato dalla sua espressione fisica (Brandi, 1988). Questo processo, invece, può essere effettuato nel caso dei documenti sonori, estraendo l'informazione audio e riversandola su un altro supporto. Per questo motivo, spesso gli interventi di digitalizzazione non tengono conto del ruolo di grande importanza che rivestono i supporti fisici che, da un punto di vista strettamente musicologico, oltre a costituire un importante testimonianza tecnologica dell'epoca, possono fornire informazioni importantissime per capire la genesi e il processo creativo che ha portato alla realizzazione delle diverse composizioni, soprattutto attraverso lo studio delle annotazioni che il compositore può aver preso direttamente sul supporto (frequente nel settore della musica elettronica) e sulla sua custodia. Anche quando si ha a che fare con registrazioni che contengono parlato, sul supporto o sulla custodia in cui è conservato il documento originale è spesso possibile trovare indicazioni guida inerenti alla configurazione dei canali e alla corretta velocità di lettura e curva di equalizzazione e, di conseguenza, rivestono un ruolo di fondamentale importanza. In numerosi casi, inoltre, i nastri magnetici sono copie uniche o possono contenere i materiali preparatori utilizzati per la composizione di una determinata opera ed è quindi possibile, attraverso il loro studio, estrarre informazioni di grande interesse musicologico, analizzando per esempio i diversi segmenti di nastro giuntati insieme per realizzare l'opera finale.

Generalmente, durante la digitalizzazione di documenti sonori analogici contenenti registrazioni di parlato e/o musica tradizionale ed etnografica, ci si trova a fronteggiare problematiche inerenti a una minor qualità dei materiali utilizzati e delle registrazioni audio. Al contrario di quanto avveniva nel settore della musica elettro-acustica, infatti, spesso l'informazione sonora è registrata da non professionisti, con una strumentazione che deve necessariamente scendere a compromessi sulla qualità in favore della portabilità. Le registrazioni sul campo, inoltre, a prescindere che si tratti di cassette o di bobine aperte, non essendo effettuate in ambienti trattati acusticamente, sono spesso di qualità inferiore rispetto a quelle realizzate in studio di registrazione. I magnetofoni portatili, quali per esempio i popolari modelli a transistor G257 prodotti da Geloso, registravano a velocità più basse rispetto ai modelli da studio più costosi e questo, se da un lato permetteva una maggior durata del nastro, dall'altro comportava una rilevante perdita di qualità generale.

Questo patrimonio, preziosissimo per gli studi linguistici e dialettali, riveste una grande valenza storico-culturale e negli ultimi anni ha visto un graduale interessamento da parte delle biblioteche digitali internazionali. Data la loro complessità e la loro eterogeneità, gli archivi sonori richiedono la messa a punto di strategie mirate e metodologie di conservazione specifiche. In questo settore, lo sviluppo e l'utilizzo di *tool* informatici per la conservazione e l'accesso ai documenti sonori possono essere di grande aiuto per agevolare lo studioso nelle diverse fasi del processo di analisi del materiale digitalizzato.

In questo contributo viene presentata la metodologia di conservazione messa a punto dal CSC in tutte le sue fasi (*Sezione 2)* e alcuni strumenti informatici per la conservazione, l'accesso e l'analisi dei documenti sonori (*Sezione 3*).

2. *Metodologia di conservazione*

Sebbene negli ultimi anni la comunità archivistica abbia aperto un grande dibattito sulle modalità di intervento più efficaci da adottare durante la digitalizzazione e la conservazione dei documenti sonori, la qualità dell'audio digitalizzato resta l'aspetto più importante e costituisce un punto fermo e universalmente condiviso. Dalla qualità della conversione A/D, infatti, dipende la continuità tra il documento originale e le copie digitali, che devono essere in grado di ricrearlo il più fedelmente possibile. In questa sezione, gli autori introdurranno la metodologia messa a punto all'interno del CSC sulla base della decennale esperienza di collaborazione con le realtà archivistiche italiane e internazionali, con particolare attenzione sull'importanza che rivestono i metadati e i supporti analogici nella digitalizzazione dei documenti sonori contenenti registrazioni di parlato e musica popolare.

2.1 Preparazione del supporto

La prima fase della metodologia proposta è incentrata sullo studio preliminare del documento sonoro e l'ottimizzazione del supporto, passi fondamentali per poter procedere con l'acquisizione del segnale audio. In questa fase è necessario procedere con una dettagliata documentazione fotografica, utile a testimoniare le condizioni del supporto, l'eventuale presenza di sintomi di degrado chimico-fisico in corso e a tenere una traccia digitale delle note dell'autore e delle informazioni aggiuntive presenti sulla scatola del documento. Eventuali documenti cartacei trovati all'interno delle scatole vengono sottoposti a scansione e a successiva archiviazione. Una prima ispezione visiva del supporto, a volte, può essere sufficiente per identificare l'insorgere di particolari sindromi di degrado, ma qualora la situazione fosse particolarmente di difficile lettura o fossero necessarie informazioni aggiuntive sui materiali e sulle loro condizioni, è possibile procedere con un'ispezione approfondita tramite tecniche di analisi chimica. Il supporto viene poi ottimizzato con appositi trattamenti, che vanno dalla "cottura" in forno nel caso dell'individuazione della *Sticky Shed-Soft Binder Syndrome SSS-SBS* (Bressan, Bertani, Furlan, Simionato & Canazza, 2016) al rifacimento di giunte in cattive condizioni con adesivi specifici. Prima di procedere con la seconda fase, il nastro viene svolto più volte a bassa velocità per ovviare ai problemi derivanti da un cattivo avvolgimento [Figura 1] o a leggeri fenomeni di magnetizzazione locale.

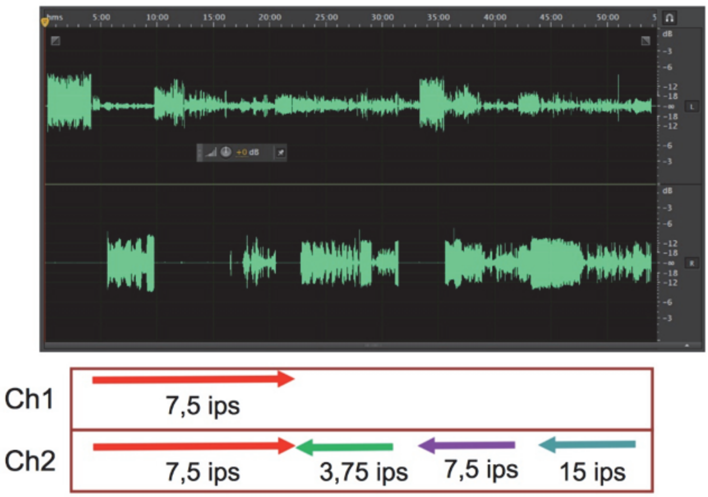
Figura 1 - *Esempio di bobina soggetta a cattivo avvolgimento*

2.2 Acquisizione del segnale

Una volta ottimizzato il supporto analogico è possibile procedere con l'acquisizione del segnale audio. La strumentazione, prima di proseguire con le operazioni di digitalizzazione, deve essere accuratamente calibrata e i parametri impostati correttamente, affinché non si introducano errori in fase di digitalizzazione. Il segnale audio viene convertito in digitale tramite l'utilizzo di convertitori A/D e software professionali per l'acquisizione e l'elaborazione audio. Eventuali dispositivi secondari, quali per esempio sistemi per la riduzione del rumore, possono essere introdotti in questo punto della catena tra il dispositivo di lettura e i convertitori A/D. La componente sonora viene digitalizzata secondo gli standard proposti dalla International Association of Sound and Audiovisual Archives, IASA (Bradley, 2009), così da ottenere copie digitali ad alta fedeltà. Durante l'intera acquisizione, il segnale audio viene monitorato attraverso un sistema di diffusori acustici per permettere all'operatore di annotare eventuali eventi significativi (rumori, disturbi, cambi di velocità, ecc.) che possono interessare la registrazione audio, facendo particolare attenzione a discernere tra quelli propri del documento originale e quelli che potrebbero essere accidentalmente introdotti dalla catena di digitalizzazione (rumori digitali e/o analogici) o da errori dell'operatore. In particolare, i principali disturbi rilevabili durante la riproduzione di un supporto magnetico sono: (a) rumori locali quali *click*, *pop* e attenuazioni del segnale dovute a giunte in cattive condizioni o a porzioni di nastro degradate, (b) rumori estesi quali fenomeni di *hum*, rumori di fondo o distorsioni, (c) rumori e alterazioni del segnale prodotti da dispositivi elettronici durante la registrazione del nastro, quali distorsioni microfoniche, rumori elettrici e feedback e (d) degrado del segnale audio dovuto al malfunzionamento del sistema di registrazione (Brock-Nannestad, 2001). Nel caso di registrazioni di parlato, spesso effettuate da operatori non professionisti del settore, quali linguisti, antropologi ecc., sono molto frequenti i nastri che presentano più velocità di lettura, in quanto rallentare la velocità di registrazione permetteva, a scapito della qualità generale, di ottenere tempi di acquisizione più lunghi, spesso necessari nel caso di interviste, conferenze ecc., o nel caso in cui lo studioso non disponesse di nastri magnetici di riserva a portata di mano.

Prima di procedere con l'acquisizione del segnale audio, è necessario individuare la velocità di lettura e la curva di equalizzazione corretta. A questo proposito, mentre la maggior parte dei magnetofoni da studio danno la possibilità di registrare alle differenti velocità di 9,5, 19 e 38 cm/s (in alcuni casi addirittura fino a 76 cm/s), i magnetofoni portatili spesso presentano velocità molto più basse, che scendono dai 4,75 cm/s ai 2,38 cm/s, spesso alternate nel corso dello stesso nastro. In molti di questi casi, anche la configurazione dei canali audio (*mono, dual mono*, stereo, quadrifonia, doppio stereo, ecc.) veniva sfruttata per registrare sullo stesso nastro il maggior numero di informazioni possibili e non è raro trovare documenti sonori nei quali il segnale audio è stato registrato con configurazioni, velocità e direzioni variabili all'interno dello stesso nastro e/o canale [Figura 2].

Figura 2 - *Esempio di segnale audio registrato a velocità e direzione variabili*



Monitorare l'acquisizione del segnale audio durante l'intero svolgimento del nastro è, pertanto, fondamentale per creare copie conservative che rispecchino fedelmente il documento di origine. L'acquisizione del segnale e il monitoraggio audio possono essere accompagnati da una documentazione video che interessa l'intero svolgimento del nastro, così da permettere all'operatore di analizzare gli eventi significativi isolando i singoli fotogrammi e confrontando la componente sonora con quella visiva. Dopo ogni fase, i dati acquisiti vengono controllati e validati, prima di procedere con l'archiviazione degli stessi e del supporto originale.

2.3 Analisi dei dati, copie conservative e copie di accesso

L'ultima fase della metodologia è centrata sull'analisi dei metadati estratti e raccolti durante l'intero processo, fondamentali per completare il quadro generale del documento sonoro insieme alle informazioni contestuali, quali foto, video ecc., utili per minimizzare la perdita di informazioni nel passaggio da analogico a digitale e ottenere copie conservative che rispecchino il più fedelmente possibile quello che è il documento sonoro d'origine. A questo proposito è fondamentale distinguere il concetto di copia conservativa da quello di copia d'accesso.

Le copie destinate all'archiviazione e alla conservazione contengono i file audio ad alta qualità digitalizzati a 96 kHz, 24 bit in formato non compresso (.wav) e senza alcuna elaborazione secondaria. Come definito da IASA (Bradley, 2009), la copia conservativa di un documento sonoro è "l'artefatto creato per essere archiviato e conservato". Da questa definizione si evince come l'utilizzo e l'accesso alla stessa si limiti a pochi casi eccezionali, essendo principalmente destinata alla conservazione. Il documento sonoro deve essere ricreato digitalmente nella sua interezza e, per questo motivo, le copie conservative devono contenere anche le informazioni contestuali, quali documentazione fotografica, documentazione video ecc. Per quanto riguarda i metadati, la metodologia messa a punto dal CSC prevede l'inclusione nelle copie conservative di schede riassuntive (.pdf) compilate grazie all'utilizzo di un software realizzato appositamente per l'acquisizione e la gestione dei metadati inerenti al documento di origine, sia dal punto di vista del supporto fisico (tipologia, marca, sintomi di degrado chimico/fisico in corso ecc.), sia da quello dei file digitali ottenuti nelle fasi dell'intero processo, quali audio, video e documentazione fotografica (formato di acquisizione, frequenza di campionamento, durata, dimensione del file, ecc.). Inoltre, per quanto riguarda i sintomi di degrado eventualmente riscontrati sul supporto di origine, le schede prevedono una valutazione da 0 a 5 delle condizioni dello stesso. Qualora fosse necessario procedere con elaborazioni quali processi di restauro digitale, filtraggi, enfasi di alcune frequenze ai fini estetici, o semplicemente fosse richiesto di accedere velocemente e facilmente ai contenuti, è possibile creare, partendo dalle copie conservative (che vengono archiviate come tali e non vengono pertanto modificate), corrispondenti copie destinate all'accesso. Queste copie sono generalmente realizzate utilizzando formati compressi (per es. mp3) di qualità e dimensioni inferiori e il loro utilizzo agevola la gestione e l'accesso ai dati da parte dello studioso. Qualora le finalità di utilizzo non richiedano la compressione dei file, possono essere realizzate delle copie d'accesso ad alta qualità da utilizzare, per esempio, per processare i segnali audio con algoritmi di restauro digitale finalizzati a migliorare l'intelligibilità dei fonemi nelle registrazioni di parlato.

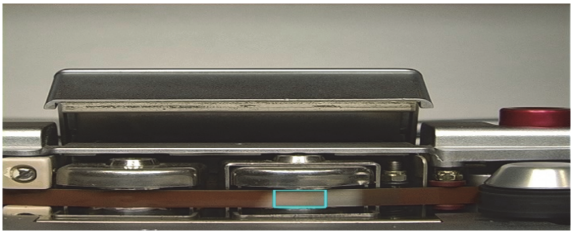
3. *Strumenti informatici*

L'ingegneria dell'informazione può essere di grande aiuto sia agli archivi digitali, sia allo studioso, attraverso l'utilizzo di strumenti informatici sviluppati *ad hoc* per l'accesso, la conservazione e l'analisi dei contenuti digitalizzati. Anche per quanto riguarda gli studi sulle registrazioni di parlato esistono diversi software molto utili per facilitare l'analisi del contenuto dei documenti sonori. In particolare, gli studiosi possono avvalersi di strumenti dedicati alla trascrizione automatica del parlato, tra cui, per esempio, Nuance Dragon, alla trascrizione fonetica quali Kaldi (Povey, Ghoshal, Boulianne, Burget, Glembek, Goel, Hannemann, Motlíček, Qian, Schwarz, Silovský, Stemmer & Vesely, 2011) e i diversi *tool* disponibili all'interno del progetto CLARIN, ecc. o, semplicemente, alla gestione e l'inserimento dei dati con possibilità di visualizzare le forme d'onda delle registrazioni sonore, come ELAN (Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006).

Oltre a questi tradizionali ambiti di analisi, relativi al contenuto sonoro, le nuove metodologie e tecnologie informatiche possono essere applicate anche allo studio del supporto. Quest'ultimo coinvolge una serie di analisi, i cui risultati sono preziosi per supportare i ricercatori nel loro lavoro. In questa sezione vengono presentati alcuni *tool* per la conservazione e l'analisi dei documenti sonori, attualmente in fase di sviluppo presso i laboratori del CSC, che coinvolgono sia il segnale audio, sia le informazioni contestuali.

Il video del nastro che scorre sulle testine del magnetofono, incluso nelle copie conservative (vedi Sezione 2.3), è un'importante fonte di informazioni, che richiede però un'attenta analisi che necessita di molto tempo e attenzione. Gli eventi di interesse possono essere di varia natura: corruttele, segni sul nastro, giunte, ecc. Per ridurre i tempi di analisi e il carico di lavoro degli studiosi, è in fase di sviluppo uno strumento per il riconoscimento automatico di punti di interesse sul video e della loro classificazione basato rispettivamente su tecniche di *computer vision* e reti neurali (Pretto, Fantozzi, Micheloni, Burini & Canazza, 2019). L'obiettivo ultimo è l'integrazione di questi algoritmi all'interno di un *tool* per l'analisi, nella quale i punti di interesse estratti automaticamente siano facilmente accessibili e integrabili con l'analisi dei contenuti [Figura 3].

Figura 3 - *Punto di discontinuità (giunta) sulla superficie del nastro riconosciuto e classificato automaticamente grazie alle tecniche di computer vision e reti neurali*



Tecniche di *machine learning* possono essere utilizzate per il riconoscimento automatico delle curve di equalizzazione (Micheloni, Pretto & Canazza, 2017). Durante la registrazione del nastro magnetico, infatti, viene applicata una curva di pre-enfasi che deve essere compensata in fase di lettura applicando una curva di post-enfasi inversa per ottenere l'audio originale come era stato registrato. Riversare un nastro applicando una curva di equalizzazione non corretta comporta un'inevitabile alterazione del segnale, compromettendo l'integrità e la validità della copia conservativa. Sebbene nel corso degli anni siano state introdotte numerose curve di equalizzazione differenti, le più utilizzate sono la IEC/CCIR (Europa) e la NAB (America). Per questo motivo, sono stati condotti degli studi preliminari per il riconoscimento automatico di queste due curve di equalizzazione (Micheloni et al., 2017). Il *dataset* creato artificialmente era composto da campioni di silenzio e rumore bianco prima registrati in un nastro magnetico e poi digitalizzati. La lettura è stata effettuata sia utilizzando uno standard di equalizzazione corretto, sia quello sbagliata. Utilizzando algoritmi di *clustering* e di classificazione, lo studio preliminare ha evidenziato la possibilità di discernere le curve di pre- e post-enfasi utilizzate. I risultati ottenuti sono stati incoraggianti e quindi ulteriori test su campioni reali sono stati effettuati (Pretto et al., 2019). Una volta realizzati i *dataset*, affinché i risultati potessero essere confrontabili con quelli ottenuti dall'analisi dei campioni di rumore bianco dei test preliminari, sono stati utilizzati i profili estratti da sezioni di silenzio all'interno del nastro. Sono stati così identificati tre principali profili di rumore: quello estratto da sezioni di silenzio nel mezzo di una registrazione (es. oratore che fa una pausa tra una parola e l'altra) (a), rumore dovuto alla testina di registrazione senza l'introduzione di uno specifico input sonoro (b) e profilo di rumore estratto da sezioni di nastro vergine (c). Dai risultati si è visto che i campioni appartenenti alle tipologie (a) e (b) possono essere utilizzate per identificare la curva di pre-enfasi. Grazie ai campioni appartenenti alla tipologia (c) è stato possibile identificare la curva di post-enfasi.

Gli autori, inoltre, hanno sviluppato algoritmi innovativi per il restauro audio (Canazza, 2012) utilizzando la tecnologia VST proprietaria di Steinberg, tra i quali CREAK (Canazza REstoration Audio Extended Kalman filter) (Canazza, De Poli & Mian, 2010) per il restauro di segnali audio estratti dai dischi in gommalacca con cui venivano realizzati i primi 78 giri, CMSR (Canazza-Mian Suppression Rule), algoritmo dedicato al restauro dei segnali audio registrati su cilindri di cera e PAR

(Perceptual Audio Restoration), un sistema di *de-hisser* che combina modelli psicoacustici con algoritmi funzionanti nel dominio della frequenza per il restauro di registrazioni su bobina e audiocassetta. Esistono, inoltre, numerosi software commerciali finalizzati al restauro audio digitale (Izotope RX, Adobe Audition, ecc.). Generalmente il restauro viene effettuato per ragioni estetiche, quali per esempio la rimozione di rumori indesiderati (pop, hum, hiss ecc.), attenuazione di frequenze risonanti ecc., per esempio al fine di realizzare una copia destinata all'uso commerciale del documento sonoro. Tuttavia, il restauro digitale può essere molto utile anche nel settore degli studi sul parlato, in particolare qualora si abbia a che fare con registrazioni che presentano problematiche relative alla difficile intelligibilità della voce registrata. In questi casi, infatti, le stesse tecniche di riconoscimento automatico dei fonemi possono mostrare i loro limiti e risultare inefficaci. In (Rodà, Russo, Canazza, & Cosi, 2018) gli autori hanno evidenziato come, nel caso di sorgenti sonore altamente compromesse e di difficile lettura, per esempio per l'eccessiva presenza di rumore di fondo, gli algoritmi di restauro audio possano essere utilizzati per migliorarne l'intelligibilità. In particolare, sono stati messi a confronto i risultati ottenuti filtrando campioni audio di difficile intelligibilità della durata di 8s con strumenti di restauro sviluppati nel settore della ricerca, tra i quali CREAK (Canazza REstoration Audio-extended Kalman Filter), CMSR (Canazza-Mian Suppression Rule) e PAR (Perceptual Audio Restoration) e alcuni algoritmi di de-noise e de-hum inclusi nel software commerciale iZotope RX. Analizzando i campioni audio con lo strumento per il riconoscimento automatico dei fonemi Kaldi, prima e dopo il restauro digitale, è stato possibile osservare in quasi tutti i casi una diminuzione del PER (Phoneme Error Rate), con una riduzione massima del 9% nel caso dei file processati con i filtri percettivi PAR. In Tabella 1 è riportato un esempio delle variazioni del PER in un campione audio processato con i diversi algoritmi di restauro. Come è possibile osservare, fanno eccezione i risultati ottenuti applicando il filtro CREAK, per i quali non si sono osservate differenze sostanziali, in quanto questo algoritmo è maggiormente indicato nel caso di rumori impulsivi puntuali.

Tabella 1 - *Esempio di variazione del PER in un campione audio prima e dopo il restauro digitale (DDD = Denoise+Dehum+Denoise, DD = Denoise+Dehum, EKF = ExtendedKalmanFilter, P = Perceptual, H = High, M = Medium, S = Small)*

	PER%
B	68,42
BDDD	64,47
BDD	61,84
BEKF	68,42
BPH	64,47
BPM	65,79
BPS	65,79

Per quanto riguarda l'accesso, al giorno d'oggi è necessario che le copie digitali siano accessibili da qualsiasi luogo e in qualsiasi momento. Tuttavia, data la complessità del caso, l'accesso ai documenti sonori può risultare incompleto, soprattutto nel caso in cui vengano adoperati i comuni lettori per la riproduzione dei contenuti multimediali. Con il loro utilizzo infatti non è possibile effettuare operazioni quali cambi di velocità, configurazione dei canali, ecc. proprie della strumentazione analogica, né accedere a tutto l'insieme di metadati e contenuti aggiuntivi che costituiscono il documento sonoro nella sua interezza. Il CSC ha realizzato delle interfacce di accesso utilizzabili da postazione fissa (Pretto, Canazza, 2014) (Fantozzi, Bressan, Pretto & Canazza, 2017) e da dispositivo mobile, che virtualizzano le apparecchiature analogiche per la riproduzione dei documenti sonori (Canazza, Fantozzi & Pretto, 2015).

L'utilità delle interfacce è stata valutata da professionisti del settore musicale di differente età, esperienza e competenza. Le interfacce sono state accolte positivamente dai professionisti intervistati, a prescindere dalla loro età e livello di esperienza. Tramite interfacce scheumorfiche, le applicazioni ne ricreano fedelmente l'aspetto e le funzionalità, così da preservarne l'esperienza di utilizzo. Per ogni documento sonoro, oltre al contenuto audio, è possibile accedere alle informazioni contestuali, quali la documentazione fotografica e i contenuti video e alla scheda riassuntiva su cui sono riportati i relativi metadati. Futuri sviluppi prevedono di includere nelle interfacce di accesso gli strumenti per l'analisi automatica della superficie del nastro, al fine di consentire all'utente di avere accesso alla visualizzazione istantanea degli eventi che emergono durante la riproduzione del documento sonoro.

4. *Conclusioni*

La grande quantità di registrazioni di parlato e musica popolare costituisce un bene di grande importanza storico culturale e, in quanto tale, deve essere salvaguardato e preservato. In molti casi, quando si ha a che fare con questa tipologia di archivi, i documenti sonori sono conservati in piccole collezioni private che spesso non seguono le linee guida internazionali per la salvaguardia dei documenti sonori (AES, 1997; Bradley, 2009). Infatti, parametri quali temperatura e umidità, qualora non controllati, possono fungere da catalizzatori per l'insorgere di numerosi sintomi di degrado chimico/fisico e ridurre le aspettative di vita dei documenti archiviati. In questo scenario, la digitalizzazione costituisce una strategia efficace per salvaguardare l'enorme quantità di informazioni a rischio a causa delle problematiche relative all'obsolescenza dei supporti analogici. La messa a punto di una metodologia efficace e condivisa è un passo necessario per poter effettuare interventi di salvaguardia mirati e specifici, minimizzando la perdita di informazioni. La metodologia proposta dal CSC risponde efficacemente alla necessità di creare copie conservative digitali fedeli ai documenti sonori nella loro interezza, dove l'informazione audio costituisce il principale centro di importanza, ma grande attenzione è data anche ai supporti fisici e alla loro ottimizzazione, alla raccolta dei metadati e alle informazioni contestuali, quali documentazione fotografica e video. Le metodologie informatiche possono essere estremamente utili, sia per la digitalizzazione e la conservazione degli archivi sonori, sia per gli studiosi del settore, velocizzando e automatizzando processi altrimenti molto dispendiosi dal punto di vista dei costi e dei tempi. Tecniche basate su *machine learning* per il riconoscimento automatico dei punti di discontinuità sul nastro e delle curve di equalizzazione, lo sviluppo di nuove interfacce di accesso e l'utilizzo di algoritmi di restauro per migliorare l'intelligibilità automatica dei fonemi, costituiscono, infatti, solo alcuni esempi di come le metodologie informatiche trovino applicazione nel settore dei documenti sonori.

Bibliografia

AES22-1997 (s2012). *AES recommended practice for audio preservation and restoration -- Storage and handling -- Storage of polyester-base magnetic tape*. http://www.aes.org/publications/standards/search.cfm?docID=25.

Bradley, K. (2009). *IASA TC-04 Guidelines in the Production and Preservation of Digital Audio Objects: standards, recommended practices, and strategies: 2nd edition.* International Association of Sound and Audio Visual Archives.

Brandi, C. (1988). Carta del Restauro 1972. In Brandi, C., *Teoría de la restauración*. Madrid: Alianza Forma, 129-149.

Bressan, F., Bertani, R., Furlan, C., Simionato, F. & Canazza, S. (2016). An ATR-FTIR and ESEM study on magnetic tapes for the assessment of the degradation of historical audio recordings. In *Journal of Cultural Heritage*, 18, 313-320. https://doi.org/10.1016/j.culher.2015.09.004.

Brock-Nannestad, G. (2001). What are the sources of the noises we remove? In *AES 20th International Conference, Archiving, Restoration, and New Methods of Recording*, 175-182.

Canazza, S. (2012). The digital curation of ethnic music audio archives: From preservation to restoration: Preserving a multicultural society. In *International Journal on Digital Libraries*, 12(2-3), 121-135. https://doi.org/10.1007/s00799-012-0088-x.

Canazza, S., De Poli, G. & Mian, G.A. (2010). Restoration of audio documents by means of extended Kalman filter. In *IEEE Transactions on Audio, Speech and Language Processing*, 18(6), 1107 - 1115. https://doi.org/10.1109/TASL.2009.2030005.

Canazza, S., Fantozzi, C. & Pretto, N. (2015). Accessing Tape Music Documents on Mobile Devices. In *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(1), 20. https://doi.org/10.1145/2808200.

Davies, S.W. (2007). Knowledge: The Missing Element in Archiving and Restoration? In *Audio Engineering Society Convention 122*, 7008.

Fantozzi, C., Bressan, F., Pretto, N. & Canazza, S. (2017). Tape music archives: from preservation to access. In *International Journal on Digital Libraries*, 18(3), 233-249. https://doi.org/10.1007/s00799-017-0208-8.

IFLA/UNESCO (2000). *Safeguarding our Documentary Heritage / Conservation préventive du patrimoine documentaire*. Boston, Massachusetts, USA: UNESCO "Memory of the World" Programme, French Ministry of Culture and Communication.

Micheloni, E., Pretto, N. & Canazza, S. (2017). A step toward AI tools for quality control and musicological analysis of digitized analogue recordings: Recognition of audio tape equalizations. In *CEUR Workshop Proceedings*, 17-24.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlíček, P., Qian, Y., Schwarz, P., Silovský, J., Stemmer, G. & Vesely, K. (2011). The kaldi speech recognition toolkit. In *IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU*.

Pretto, N., Canazza, S. (2014). Rewind: simulazione di un'esperienza d'ascolto storicamente fedele di dischi fonografici digitalizzati. In *Proceedings of XX CIM*, 29-34.

Pretto, N., Fantozzi, C., Micheloni, E., Burini, V. & Canazza, S. (2019). Computing Methodologies Supporting the Preservation of Electroacoustic Music from Analog Magnetic Tape. In *Computer Music Journal*, 42(4), 59-74. doi: https://doi.org/10.1162/comj_a_00487.

Rodà, A., Russo, A., Canazza, S. & Cosi, P. (2018). Audio documents restoration as a documentary source in the linguistic research comparison of instruments. In De Dominicis, A. (Ed.), *Speech audio archives: preservation, restoration, annotation, aimed at supporting the linguistic analysis.* Roma: Accademia Nazionale dei Lincei, 117-138.

Smith, A. (1999). *Why digitize? / by Abby Smith*. Washington, D.C.: Council on Library and Information Resources.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings Language Resources and Evaluation*, 1556-1559.

Autori

Giovanni Abete – Dipartimento di Studi Umanistici, Università degli Studi di Napoli "Federico II"
giovanni.abete@unina.it

Alessandro Ansani – Department of Psychology, Sapienza University of Rome – Department of Philosophy, Communication and Performing Arts, Roma Tre University
alessandro.ansani@uniroma1.it

Fabio Ardolino – Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Pisa – Laboratoire Parole et Langage, Aix-Marseille Université, Aix-en-Provence, Francia
fabio.ardolino@fileli.unipi.it

Pier Marco Bertinetto – Scuola Normale Superiore, Pisa
piermarco.bertinetto@sns.it

Greta Bianchini – Università degli Studi di Siena
greta.bianc@gmail.com

Erica Bonato – Dipartimento di Neuroscienze, Corso di laurea in Logopedia, Università degli Studi di Padova
ericabonato@live.it

Roberto Bovo – Dipartimento di Neuroscienze, Università degli Studi di Padova
roberto.bovo@unipd.it

Pietro Braione – Department of Computer Science, University of Milano-Bicocca
pietro.braione@unimib.it

Simonetta Buttò – Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche-ICCU
simonetta.butto@beniculturali.it

Silvia Calamai – Università degli Studi di Siena
silvia.calamai@unisi.it

Sergio Canazza – Dipartimento di Ingegneria dell'Informazione (DEI), Università di Padova
canazza@dei.unipd.it

Francesco Cangemi – IfL Phonetics, University of Cologne, Germany
fcangemi@uni-koeln.de

Violetta Cataldo – Department of Humanities Studies, University of Salerno
vcataldo@unisa.it

Piero Cavallari – Istituto Centrale per i Beni Sonori ed Audiovisivi, Roma
piero.cavallari@beniculturali.it

Carlo Cecchetto – Department of Psychology, University of Milano-Bicocca
carlo.cecchetto@unimib.it

Alessandra Chiera – Department of Philosophy, Communication and Performing Arts, Roma Tre University
ale.chiera@gmail.com

Valentina Colonna – Dottorato in Digital Humanities, Università di Genova-Università di Torino
valentina.colonna@unito.it

Claudia Roberta Combei – Department of Modern Languages, Literatures, and Cultures, University of Bologna
claudia.combei2@unibo.it

Piero Cossu – Department of Philology, Literature and Linguistic, University of Pisa
piero.cossu@fileli.unipi.it

Elisa D'Argenio – Dipartimento di Studi Umanistici, Università degli Studi di Napoli "Federico II"
elisa.dargenio@unina.it

Mariapaola D'Imperio – Rutgers University, Department of Linguistics and The Center for Cognitive Science (RuCCS), New Brunswick (NJ), USA
mdimperio2@gmail.com

Valentina De Iacovo – Dip. di Lingue e Letterature Straniere e Culture Moderne, Università degli Studi di Torino
valentina.deiacovo@unito.it

Maria Di Maro – Università degli studi di Napoli "Federico II"
mariadimaro2@unina.it

Dalila Dipino – Romanisches Seminar, University of Zurich, Switzerland
dalila.dipino@uzh.ch

Lorenzo Filipponio – Institut für Romanistik, Humboldt-Universität zu Berlin, Germany
lorenzo.filipponio@hu-berlin.de

Manuela Frontera – Dipartimento di Culture, Educazione e Società, Università della Calabria
manuela.frontera@unical.it

Jessica Fründt – IfL Phonetics, University of Cologne, Germany

Vincenzo Galatà – CNR - Istituto di Scienze e Tecnologie della Cognizione (ISTC)
vincenzo.galata@pd.istc.cnr.it

Davide Garassino – Romanisches Seminar, University of Zurich, Switzerland
davide.garassino@uzh.ch

Giusy Gaudino – Università degli studi di Napoli "Federico II"
giusy.gaudino@studenti.unina.it

Carlo Geraci – Institut Jean-Nicod & Department of Cognitive Science, CNRS, ENS & PSL, Paris, France
carlo.geraci76@gmail.com

Lucilla Gigli – Università degli Studi di Siena
lucilla.gigli@unisi.it

Elvira Graziano – Dipartimento di Culture, Educazione e Società, Università della Calabria
elvira.graziano@unical.it

Martine Grice – IfL Phonetics, University of Cologne, Germany
martine.grice@uni-koeln.de

Harriet Hanekamp – IfL Phonetics, University of Cologne, Germany
harriet.hanekamp@uni-koeln.de

Giovanna Lenoci – Scuola Normale Superiore, Pisa
giovanna.lenoci@sns.it

Sabina Magrini – Istituto Centrale per i Beni Sonori ed Audiovisivi, Roma
sabina.magrini@beniculturali.it

Paolo Mairano – Departement des Etudes Anglophones, Université de Lille, France
paolo.mairano@univ-lille.fr

Giovanna Marotta – Department of Philology, Literature and Linguistics, University of Pisa
giovanna.marotta@unipi.it

Philippe Martin – Département d'Études Françaises, Université de Toronto
philippe.martin@utoronto.ca

Antonio Origlia – URBANECO Research Center, University of Naples "Federico II"
antonio.origlia@unina.it

Riccardo Orrico – University of Salerno & Aix Marseille University, CNRS, LPL UMR 7309, Aix en Provence, France
rorrico@unisa.it

Caterina Pesce – Università degli Studi di Padova
caterina.pesce1989@gmail.com

Roland Pfau – Department of Linguistics, University of Amsterdam, The Netherlands
r.pfau@uva.nl

Duccio Piccardi – Dipartimento di Filologia, letteratura e linguistica, Università di Pisa
duccio.piccardi@fileli.unipi.it

Alessandra Pinton – Dipartimento di Neuroscienze, Corso di laurea in Logopedia, Università degli Studi di Padova
alessandra.pinton@unipd.it

Isabella Poggi – Department of Philosophy, Communication and Performing Arts, Roma Tre University
isabella.poggi@uniroma3.it

Niccolò Pretto – Dipartimento di Ingegneria dell'Informazione (DEI), Università di Padova
prettoni@dei.unipd.it

Josep Quer – ICREA-Department of Translation and Language Sciences, Pompeu Fabra University, Barcelona, Spain
josep.quer@upf.edu

Vito Quinci – Dipartimento di Informatica, Università di Torino
vito.quinci540@edu.unito.it

Antonio Rodà – Dipartimento di Ingegneria dell'Informazione (DEI), Università di Padova
roda@dei.unipd.it

Antonio Romano – Dip. di Lingue e Letterature Straniere e Culture Moderne, Università degli Studi di Torino
antonio.romano@unito.it

Alessandro Russo – Dipartimento di Ingegneria dell'Informazione (DEI), Università di Padova
russoale@dei.unipd.it

Daniele Santini – Università di Pisa
danyfive@msn.com

Anyla Saraçi – Department of Albanian Language and Literature, "Fan S. Noli" University of Korça, Albania
animaxhe@yahoo.com

Renata Savy – Department of Humanities Studies, University of Salerno
rsavy@unisa.it

Loredana Schettino – Department of Humanities Studies, University of Salerno
lschettino@unisa.it

Isora Sessa – Department of Psychology, Sapienza University of Rome – Department of Philosophy, Communication and Performing Arts, Roma Tre University
isora.sessa@uniroma1.it

Marica Setaro – Università degli Studi di Siena
marica.setaro@unisi.it

Manuela Soldi, Festivaletteratura – Dipartimento di Culture del progetto, Università IUAV di Venezia
manuela.soldi@virgilio.it

Rosanna Sornicola – Dipartimento di Studi Umanistici, Università degli Studi di Napoli "Federico II"
sornicol@unina.it

Maria Francesca Stamuli – Soprintendenza Archivistica e Bibliografica della Toscana
mariafrancesca.stamuli@beniculturali.it

Daniela-Carmen Stoica – Department of Foreign Languages, "Fan S. Noli" University of Korça, Albania
dstoica76@gmail.com

Andrea Tarasi – Dipartimento di Culture, Educazione e Società, Università della Calabria
andreatarasi17@gmail.com

Graziano Tisato – Istituto di Scienze e Tecnologie della Cognizione (ISTC), Consiglio Nazionale delle Ricerche (CNR), Via Martiri della Libertà, 2 – 35127 Padova
graziano@tisato.it

Martina Turconi – Department of Philology, Literature and Linguistic, University of Pisa
martina.turconi@phd.unipi.it

Petra Vangelisti – Università degli Studi di Siena
petravangelisti@hotmail.it

Cesarina Vecchia – Dipartimento di Studi Umanistici, Università degli Studi di Napoli "Federico II"
cesarina.vecchia@unina.it

Claudio Zmarich – CNR - Istituto di Scienze e Tecnologie della Cognizione (ISTC) – Dipartimento di Neuroscienze, Corso di laurea in Logopedia, Università degli Studi di Padova
claudio.zmarich@cnr.it

Enrico Zovato – Laboratorio di Fonetica Sperimentale "Arturo Genre", Università di Torino
ezovato@gmail.com

Studi AISV è una collana di volumi collettanei e monografie dedicati alla dimensione sonora del linguaggio e alle diverse interfacce con le altre componenti della grammatica e col discorso. La collana, programmaticamente interdisciplinare, è aperta a molteplici punti di vista e argomenti sul linguaggio: dall'attenzione per la struttura sonora alla variazione sociofonetica e al mutamento storico, dai disturbi della parola alle basi cognitive e neurobiologiche delle rappresentazione fonologiche alle applicazioni tecnologiche. I testi sono selezionati attraverso un processo di revisione anonima fra pari e vengono pubblicati nel sito dell'Associazione Italiana di Scienze della Voce con accesso libero a tutti gli interessati.

Duccio Piccardi è dottorando di ricerca in Linguistica presso Università di Pisa e Universität Zürich. Si occupa di sociofonetica e di dialettologia italiana. Ha lavorato sulla reinterpretazione sociosemiotica di tratti dialettali ed ha condotto esperimenti di psicolinguistica sugli effetti della frequenza lessicale e sul fonosimbolismo.

Fabio Ardolino è dottorando di ricerca in Linguistica presso Università di Pisa e Aix-Marseille Université. Si occupa principalmente degli aspetti identitari, relazionali e psicologico-sociali legati alla variazione sociofonetica, con particolare attenzione alle loro implicazioni acquisizionali nei contesti migratori.

Silvia Calamai insegna Linguistica Generale e Sociolinguistica all'Università di Siena, nella sede di Arezzo. Fa parte del gruppo di ricerca Oral History and Technology di CLARIN, è membro del consiglio direttivo dell'Associazione Italiana di Scienze della Voce e del Comitato di redazione di *Sonorités Bulletin de l'AFAS*.

AISV - Associazione Italiana Scienze della Voce
sito: www.aisv.it
email: aisv@aisv.it| redazione@aisv.it
ISBN: 978-88-97657-38-5

Edizione realizzata da
Officinaventuno
info@officinaventuno.com | sito: www.officinaventuno.com
via F.lli Bazzaro, 18 - 20128 Milano - Italy