



UNIVERSITÀ  
DI SIENA 1240



## XV AISV CONFERENCE

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

February 14-16, 2019

*Department of Education, Human Sciences and  
Intercultural Communication*

Campus del Pionta, Arezzo

## BOOK OF ABSTRACTS



# Indice

ELEONORA TODDE, VALERIA ZEDDA L'Archivio Sonoro Luisa Orrù tra passato e presente: nuove strategie per la valorizzazione	7
SONIA CENCESCHI, ALESSANDRO TRIVILINI, SANDRO DENICOLÀ The scientific disclosure of speech analysis in audio forensics: remarks from a practical application	10
LAURA GAVIOLI, NATACHA NIEMANTS The AIM Corpus: Some notes on the electronic collection of interpreter- mediated interaction	14
FEDERICA CUGNO, FEDERICA CUSAN Digital Toponymy. Il progetto MAPforUS e l'Atlante Toponomastico del Piemonte Montano on line	16
EMANUELA CRESTI, MASSIMO MONEGLIA, ALESSANDRO PANUNZI The LABLITA spoken corpora collection and the Language into Act Theory	19
CLAUDIA ROBERTA COMBEI, GIOVANNA MAROTTA Where do accents come from? Factors affecting the degree of foreign- accented Italian	23
CLAUDIO BARALDI, CHIARA BALLESTRI, VITTORIO IERVESE Combining speech and images: the case of the SHARMED archive	27
GRAZIANO TISATO Acquisizione Digitale dell'Intero AIS	30
ALICE BETTI Archivi orali e socioprosodia: un'analisi dell'interrogativa globale a Ripacandida (PZ) e Siena (SI)	34

CAMILLA BERNARDASCI, DIETER STUDER-JOHO L'Archivio fonografico zurighese tra passato e futuro	39
MICHELE TOSS Conservazione e schedatura delle fonti orali: alcune riflessioni	42
HYEON-YEOL IM Study on the Pronunciation of Korean TTS, conditioned on the presence or absence of a morpheme boundary	45
PIETRO BRAIONE, CARLO CECCHETTO, CARLO GERACI, ROLAND PFAU, JOSEP QUER Hidden languages in a digital world: the case of sign language archives	48
FABIO ARDOLINO Adattare i protocolli IAT allo studio del parlato: criticità e proposte metodologiche	51
MARIA FRANCESCA STAMULI Fonti orali, documenti e archivi: riflessioni e proposte per la nascita di un 'archivio vivo'	55
PHILIPPE MARTIN Prosodic annotation of oral archives (1913-2018)	59
ALESSANDRO RUSSO, NICCOLÒ PRETTO, ANTONIO RODÀ, SERGIO CANAZZA Information engineering methodologies for managing and preserving audio (speech and music) archives	61
MANUELA FRONTERA, ANDREA TARASI, ELVIRA GRAZIANO Le consonanti occlusive sorde aspirate in Calabria: un confronto tra aree dialettali	63
VALENTINA COLONNA, ANTONIO ROMANO VIP: un archivio per le voci della poesia italiana	67
ANNE BREITBARTH, MELISSA FARASYN, ANNE-SOPHIE GHYSELEN, JACQUES VAN KEYMEULEN What linguistic resources reveal about rusted bicycles and hoop making	71

FRANCESCO CANGEMI, JESSICA FRÜNDT, HARRIET HANEKAMP, MARTINE GRICE A semi-automatic workflow for orthographic transcription and syllabic segmentation	73
FRANCA ORLETTI, FRANCESCA SOCRATE Voci del 68: donne e uomini ridono in un corpus di storia orale	77
DANIELA-CARMEN STOICA, ANYLA SARAÇI Oral Histories and the Dialectal Research in Albania	80
ŁUKASZ STOLARSKI Building a Phonetic Corpus from Librivox Audiobooks	84
ANNA-MARIA SICHANI Data documentation, modeling and analysis of a digital oral history archive: “The Connected Histories of the BBC” project as a case-study	86
MANUELA SOLDI L’archivio sonoro di Festivaletteratura di Mantova	88
SONIA CENCESCHI, ALESSANDRO TRIVILINI, LICIA SBATTELLA , ROBERTO TEDESCO Collecting Italian spontaneous social media speech: the WAsp2 project	92
PAOLO MAIRANO, ENRICO ZOVATO, VITO QUINCI Do sentiment analysis scores correlate with acoustic features of emotional speech?	96
RICCARDO ORRICO, RENATA SAVY, MARIAPAOLA D’IMPERIO Individual variability in Salerno Italian intonation: evidence from read and spontaneous speech	100
ROSALBA NODARI, CHIARA MELUZZI A sociophonetic investigation of Roman rhotic degemination	104
CHIARA MELUZZI, NICHOLAS NESE Dental affricates variation in Italian regional varieties: introducing the ZIta corpus	109

VINCENZO GALATÀ, LORENZO SPREAFICO, ALESSANDRO VIETTI Costruzione e manutenzione di un archivio fonarticolatorio	112
LORENZO FILIPPONIO, DAVIDE GARASSINO, DALILA DIPINO Tra fonetica e tipologia: la durata consonantica in due dialetti italo-romanzi settentrionali	115
DANIELE SANTINI, GIOVANNA MAROTTA, GIOVANNA LENOCI, PIER MARCO BERTINETTO Vocali ripetute in italiano. Uno studio esplorativo	120
PHILIPPE BOULA DE MAREÜIL, VALENTINA DE IACOVO, ANTONIO ROMANO, IOANA VASILESCU, FRÉDÉRIC VERNIER Verso un atlante sonoro integrato delle lingue e dei dialetti di Francia e d'Italia	123
VIOLETTA CATALDO, LOREDANA SCETTINO, RENATA SAVY, ISABELLA POGGI, ANTONIO ORIGLIA, ALESSANDRO ANSANI, ISORA SESSA, ALESSANDRA CHIERA Tratti fonetici e funzionali delle pause, e gesti concomitanti, nel parlato delle guide turistiche	125
ROSANNA SORNICOLA, GIOVANNI ABETE, ELISA D'ARGENIO, CESARINA VECCHIA Raccontare un archivio di fonti orali: il progetto <i>Voci, parole e testi della Campania</i>	129
CLAUDIO ZMARICH, SABRINA BONICHINI, SERENA BONIFACIO, VINCENZO GALATÀ, YVAN ROSE Bringing the phonological Development of Italian Children aged 30-47 months into PhonBank	132
PIERO COSSU La metatesi di /r/ nel dialetto di Suelli (sardo campidanese). Una analisi descrittiva dei nessi aberranti	136
EMMA MILANO L'Archivio Sociolinguistico ed Etnografico di Greci nel progetto <i>Voci, parole e testi della Campania</i> . La fonte orale come risorsa per lo studio linguistico e per la valorizzazione e la tutela dei valori linguistico-culturali di minoranza	140

CLAUDIO ZMARICH, ERICA BONATO, ROBERTO BOVO, VINCENZO GALATÀ, ALESSANDRA PINTON A new phonological discrimination test for children aged 48-72 months	144
GIOVANNA LENOCI, IRENE RICCI Curve-to-curve distances to assess lingual articulations in the speech of Italian children who stutter and control peers	148
STEFANO NEGRINELLI, GIULIA DONZELLI Il progetto <i>AIS, reloaded</i> : un archivio sonoro per 36 varietà dialettali della Svizzera meridionale	151
MARIA DI MARO, GIUSY GAUDINO, FRANCESCO CUTUGNO Conversing on Artistic Representation Topics (CART): an Android Audio Guide with Dialogic Skills	154
SILVIA CALAMAI, GRETA BIANCHINI, LUCILLA GIGLI, CATERINA PESCE, MARICA SETARO, PETRA VANGELISTI Ritrovare voci. Il lavoro intorno all'archivio sonoro di Anna Maria Bruzzone	157
MIKI MORI The CoPLACo Project: Creating a Multi-use, Digital Platform of Oral Language for the Comorian Islands	160
SONIA I. D'APOLITO, BARBARA GILI FIVELA Interazione contesto-accuratezza del parlato L2: studio pilota sulle produzioni di geminate in italiano L2 da parte di parlanti francofoni	161
DUCCIO PICCARDI Bilingual Frequency in a Favorable Context (BFFC) in the Italian dialectal area. A protocol for the analysis of geminate lateral retroflexion in Antona (MS)	165
CARLO BIANCHI Interviste orali su musica e storia. Un'ipotesi di archivio consultabile fra problemi scientifici ed etici	169

NORAH KARROUCHE, LILIANA MELGAR, ROELAND ORDELMAN  
Never the twain shall meet? Collaborations between speech technologists  
and oral historians in the Netherlands 173

GIOVANNA MAROTTA, MARTINA TURCONI  
Ritmo linguistico e ritmo musicale. Un'analisi pilota sul ritmo della lingua  
cantata 175

ELEONORA TODDE<sup>1</sup>, VALERIA ZEDDA<sup>1</sup>

<sup>1</sup> Università degli Studi di Cagliari

## **L'Archivio Sonoro Luisa Orrù tra passato e presente: nuove strategie per la valorizzazione**

L'Archivio Storico dell'Università di Cagliari ha recentemente accolto tra i suoi fondi l'importante archivio sonoro della prof.ssa Luisa Orrù, docente di Antropologia Culturale presso l'Ateneo cagliaritano dal 1986 al 1997, anno della sua prematura scomparsa.

Il materiale sonoro consta di 1235 unità sonore prodotte, le audiocassette, e 1400 unità di rilevazione, le interviste. Queste interviste inizialmente erano condotte con il metodo del “dialogo informativo libero senza questionario” per poi passare all'utilizzo di questionari (rigido e semilibero)<sup>1</sup>. L'archivio è strettamente legato all'attività della cattedra ed è caratterizzato da una specificità di argomenti: in particolare il ciclo riproduttivo, la medicina popolare e l'igiene della casa e della persona, quella che oggi può essere definita Antropologia medica; i sogni e le credenze relative alle anime dei morti; la valenza simbolica della casa; la veterinaria tradizionale.

Una parte del materiale sonoro, circa 300 cassette audio, era stato inizialmente duplicato su cassetta e cd, per poi essere digitalizzato in formato MP3 e Wave. Si è inoltre digitalizzato circa 200 cassette di cui si possedeva solamente il duplicato, poiché l'originale è conservato presso i familiari della prof.ssa Orrù.

Assieme a questo materiale, la prof. Orrù aveva predisposto tre registri ordinati per unità sonora, che veniva numerata in ordine progressivo, e altre 168 cartelle di materiale cartaceo prodotto durante la sua attività e quella del gruppo di lavoro da lei formato nel corso dell'attività didattica. Questi fascicoli contengono prevalentemente schede (predisposte per rilevatore, unità sonora e informatore), relazioni e trascrizioni.

Dopo la scomparsa della Orrù, un nuovo gruppo di lavoro coordinato dalla prof.ssa Fulvia Putzolu<sup>2</sup>, che aveva ereditato il patrimonio archivistico, decise in un primo momento di informatizzare i dati dei registri utilizzando un file Excel per poi passare a un database Access organizzato con tabelle correlate,

---

<sup>1</sup> ORRÙ, L. (1984-86). Stato della documentazione e prospettive di ricerca sul ciclo riproduttivo in Sardegna. In *BRADS*, 12-13, 17-37; ORRÙ, L. (1994). Produzione e archiviazione di documenti orali sul ciclo riproduttivo in Sardegna. In ORRÙ, L. & PUTZOLU, F. (Eds.), *Il parto e la nascita in Sardegna. Tradizione, medicalizzazione, ospedalizzazione*. Cagliari: Cuec, 260-290.

<sup>2</sup> Grazia Loi e Carmina Orrù hanno curato il lavoro di duplicazione; Fulvia Putzolu e Valeria Flore hanno realizzato rispettivamente il file Excel e il database Access; sempre la Putzolu ha curato la digitalizzazione dei file audio; Salvatore Capra ha digitalizzato il patrimonio cartaceo.

predisponendo dei collegamenti ipertestuali tra i dati e i file audio. In un secondo momento, grazie alla collaborazione del Dipartimento di Scienze della Vita e dell'Ambiente dell'Università degli Studi di Cagliari, è stato digitalizzato anche il restante materiale cartaceo.

Il lavoro che si intende portare avanti è preliminarmente conservare i materiali presenti nell'archivio cercando di evitare qualsiasi danneggiamento che possa portare alla perdita di dati: si dovrà curare la conservazione a lungo termine con riversamenti diretti e sostitutivi<sup>1</sup>. Una volta completata la digitalizzazione bisognerà pensare alla descrizione: devono essere previsti tutti gli elementi che rispettino gli standard di descrizione archivistica ISAD (G) (tipologia, titolo, datazione, contenuto, stato di conservazione, segnatura) e i metadati dell'oggetto digitale che garantiscano la migliore diffusione dei dati.

Una volta terminato il lavoro sui materiali audio si valuteranno le attuali possibilità offerte dalla tecnologia e dall'umanistica digitale per rendere fruibile al pubblico questo patrimonio documentario. Attualmente è in fase di progettazione il portale web dell'Archivio Storico dell'Università di Cagliari che ha come obiettivo la realizzazione di uno strumento innovativo che possa garantire l'accesso alle fonti che l'Archivio conserva. Nel portale uno spazio sarà dedicato all'Archivio Sonoro Demoantropologico Luisa Orrù che presenterà agli utenti informazioni generali sulla storia dell'archivio e su come sono stati trattati i materiali; inoltre, consentirà di consultare l'inventario digitale che si sta predisponendo e sarà corredato da una nota introduttiva in cui verranno spiegati l'attuale struttura del fondo e la tipologia dei materiali conservati. Ci saranno anche altri strumenti di ricerca come, ad esempio, il database delle interviste. Questo permetterà di svolgere una ricerca semplice e veloce: l'utente potrà selezionare le interviste in base all'argomento che maggiormente gli interessa oppure potrà inserire nei campi di ricerca che verranno messi a sua disposizione le informazioni che conosce. Il risultato della ricerca fornirà la scheda dell'intervista con le informazioni principali come la data di registrazione, la lingua in cui si è svolta, il contenuto e la durata e inoltre l'utente avrà la possibilità di ascoltarla. L'intento è di fornire uno strumento accessibile sia ad un pubblico specializzato che ad una utenza "inesperta": quest'ultima avrà a disposizione dei percorsi tematici e una ricerca semplice Google-like, con indicizzazione per parola; mentre i primi si potranno avvalere di una ricerca avanzata con la possibilità di scelta tra tutti i campi della scheda descrittiva.

Il presente contributo vuole portare alla conoscenza della comunità scientifica il patrimonio dell'Archivio Sonoro DemoAntropologico Luisa Orrù, raccontarne le vicende che hanno portato alla sua costituzione e analizzare le strategie che sono state poste in essere per trasformare l'archivio "analogico" in "digitale", mettendo in evidenza le criticità riscontrate. Infine si valuteranno le attuali

---

<sup>1</sup> Associazione Internazionale degli Archivi Sonori e Audiovisivi, *LASA-TC 03 La Salvaguardia del Patrimonio Sonoro: Etica, Principi e Strategie di Preservazione*, Versione 3, dicembre 2005.

possibilità offerte dalla tecnologia e dall'umanistica digitale per rendere fruibile al pubblico questo materiale documentario.

SONIA CENCESCHI<sup>1</sup>, ALESSANDRO TRIVILINI<sup>2</sup>, SANDRO DENICOLÀ<sup>2</sup>

<sup>1</sup> Politecnico di Milano, <sup>2</sup> University of Applied Sciences and Arts of Southern Switzerland (SUPSI)

## **The scientific disclosure of speech analysis in audio forensics: remarks from a practical application**

This work describes a university teaching experience of main speech analysis methodologies to non-expert professionals active in the legal and security areas. Results and remarks are being used to develop a final protocol for introductory courses in forensic linguistics. The final purpose is to transmit the potential of the scientific analysis in order to stem the multiple problems caused by misinformation (Carrada, 2005) and enhance the dissemination of scientific contents.

The diffusion of scientific (even if constantly evolving) methods is an urgent need, specially in national unregulated audio-forensics contexts. Moreover, the same word “expert”, generically defined as audio forensics expert, has countless facets: better speaking therefore about a universe of different specializations which should collaborate, and understand each other, to satisfy everchanging questions. Since the situation is very complex, this work is narrowed just on one aspect: the optimal simplification of specialized concepts in order to increase the defence capabilities of the whole society from superficial, irresponsible and misleading experts reports and conclusions.

The focus is not on performing researches in speech analysis (users already have their specialties), but on making the person independent and prepared to what he/she must expect from alleged technical experts.

Italy and Cantone Ticino share the same official language but also similar needs: the main difference lies in usage frequency of speech recordings in trials and investigations. While Italy has been doing a decades-long and huge use of audio recordings, in Italian Switzerland they are, to date, sporadically introduced. Anyhow, the interest is growing up, so that the proposal applies to both countries, whether it be to remedy or to lay foundations for dialog and education.

The protocol is limited to needs and suggestions collected among judges, lawyers and law enforcement officers during a job lasted years and later used to create a training course. Ideas have been so applied during the course “Biometrics and emerging technologies” included in the Certificate of Advanced Studies (CAS) in Advanced Digital Forensics, organized by the Digital Forensics Service of the University of Applied Sciences of Southern Switzerland (SUPSI).

Participants possess an average informatics preparation, but no requirements concerning audio analysis and human voice: they were a homogeneous group, good to test the difficulty of concepts which have been carried out in about 15 hours. Even if it was the first attempt, collected feedbacks confirmed the strong interest and the acquisition of critical sense and correct basic knowledge.

The following list reports most common technical forensics requests limited to the human voice and related examples of wrong procedures or conclusions. They are just few examples, but complex enough to extrapolate a common developing basis and show the gap between reality and utopia.

TECHNICAL REQUEST	EXAMPLES OF WRONG CONCLUSIONS/RESULTS
Increase of intelligibility	The audio has been cleaned too much or transcripts are made listen to clips in loop so that speech perception is altered.
Disambiguation of linguistic contents	The phonetics analysis is not done and conclusions rely only on subjective perception.
Voice comparison	Conclusions are carried out without controlling SNR value requirements, or doing them perceptively, or without underlining integrity/conservation issues and providing answers without uncertainty (Y/N).
Geographical/Dialect detection	Linguists and dialectologists are not consulted and the analysis is conducted on erroneous perceptive bases.

The protocol application started showing typical wrong methodological answers to diffuse requests, and applying a reverse process, listing procedural and technical errors. Once introduced realistic answers to forensics queries, the human speech complexity and phonetics fundamentals have been addressed in order to underline discrepancies. Communicate these concepts to nonexperts, without speech and sound analysis knowledge, is the most crucial point. The experience showed that spectrogram concept can be well explained without going into detail. Furthermore, concepts related to noise and harmonicity, vowels and consonants are also easier to be understood if exploiting parallels with music (Wolfe, 2002). This work has produced excellent results on the critical sense of listeners, showing it is possible also with a few hours available.

A superficial introduction to complex topics has been enough to allow the analysis of practical cases, where the signal has not been correctly processed. Proceeding then for practical examples, it has been showed conclusions can never be "yes" or "no", but a range of variable nuances according to the starting situation conditions, with a brief excursus about Likelihood Ratio (Martire et al., 2014; Becker et al, 2008).

The danger in communicating doubts necessity, and explaining casual, systematic and inevitable process errors, resides in transmit the wrong perception that, then, speech recordings are useless. The found remedy has been then to underscore the potentiality of uncertainties, specially respect to social security

and potential judicial errors. Absolute certainty is not a proof, but a false solution emptied of meaning.

After completing this work of “dismounting and reassembling with cognition” process, two main topics have been addressed: voice comparison and speech perception. The desire was to provide a comprehensive overview being there more in vogue on the media compared to many others. The synthesis consisted of automatic, semi-automatic and perceptive analysis for speakers’ comparisons, with explanations regarding main problems and obstacles (Romito & Galatà, 2004). Perception issues have been listed instead (Romito, 2005; Fraser, 2003), to emphasize transcripts are not so trivial, especially to emphasize that the transcripts are not so trivial, even more with background noise.

The last theme dealt with is the need for a division of tasks, describing, with practical examples, the needed knowledge to address different issues. If the constant dialogue between different professions is utopia, it is also true that the situation can always be improved and it can be done thanks to dissemination and involvement. Dialogue must not increase just between police forces, judges, lawyers and the scientific world, but also among experts within the latter.

The complete protocol deeps and orders the themes here summarized, but all the project is a starting point based on one personal experience, consequently representing a partial view and needing deepen discussion and practical applications. Moreover, this topic involves other essential and serious practical issues can just be cited here, such as the standardization and regulation of specialized courses, procedures, and remuneration [Romito & Galatà, 2007]. The vastness of the topic scares, but it is certainly better to be frightened than not, since judgments or investigations directly affect safety, costs and life of citizens.

### *Riferimenti bibliografici*

BECKER, T., JESSEN, M., & GRIGORAS, C. (2008). Forensic speaker verification using formant features and Gaussian mixture models. In *Ninth Annual Conference of the International Speech Communication Association*.

CARRADA, G. (2005). *Comunicare la scienza: kit di sopravvivenza per ricercatori* (Vol. 12). Alpha Test.

FRASER, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. In *Forensic Linguistics*, 10, 203-226.

MARTIRE, K. A., KEMP, R. I., SAYLE, M., & NEWELL, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. In *Forensic science international*, 240, 61-68.

ROMITO, L. (2005). La competenza linguistica nelle trascrizioni Forensi: l'intelligibilità, l'oggettività e il rapporto segnale/rumore. In *Detective And Crime*.

ROMITO, L., & GALATÀ, V. (2004). Towards a protocol in speaker recognition analysis. In *Forensic science international*, 146, S107-S111.

ROMITO, L., & GALATÀ, V. (2007). Speaker recognition: stato dell'arte in Italia valutazione dei corpora, dei metodi e delle professionalità coinvolte. In *Scienze vocali e del linguaggio-metodologie di valutazione e risorse linguistiche*, 3, 223-242.

WOLFE, J. (2002). Speech and music, acoustics and coding, and what music might be 'for'. In *Proc. 7th International Conference on Music Perception and Cognition*, 10-13.

LAURA GAVIOLI<sup>1</sup>, NATACHA NIEMANTS<sup>1 2</sup>

<sup>1</sup> Università di Modena e Reggio Emilia, <sup>2</sup> Università di Bologna

## **The AIM Corpus: Some notes on the electronic collection of interpreter-mediated interaction**

While corpora of spoken language have increased enormously in the last years and their availability is now large, methods of analyzing such corpora electronically do not seem to have followed accordingly.

Corpora have traditionally grown in the realm of phraseology (Sinclair, 1991), and, within this realm, attempts to apply concordance analysis to speech are quite an exception. Aston (2015) for instance is one of the few studies focusing on recurrent speech prosody of phraseological items. The corpus discussed by Aston is a small set of TED talks, which he has experimented in teaching (and learning) conference interpreting.

In the area of Conversation Analysis, talk has been collected since the first works of Harvey Sacks in the 1970s, but reasoning about how to implement such collections of talk for electronic analysis is very recent and experimental (Stivers, 2015; Steensig & Heinemann, 2015).

For the last 15 years, we have collected spoken data for research in a relatively novel area of studies that has become known as Dialogue Interpreting (DI, Mason, 1999). Dialogue interpreting consists in interactions involving speakers of different languages and bilingual participants who translate to allow their interlocutors to communicate with each other. Our corpus of data has been collected specifically in healthcare settings located in highly industrialised migration areas in North Italy. The conversations in our corpus normally involve a health worker (doctor, nurse or midwife), a migrant patient and a so called language-cultural mediator, a bilingual speaker with both an interpreting and a migration experience.

The following are, in short, the data we have gathered at the moment. The corpus is called AIM (Analysis of Interaction and Mediation) after the name of the national group of researchers who have collaborated to its implementation and analysis (<http://www.aim.unimore.it/site/home.html>). The AIM corpus (2004-2018).

Language couples	No. encounters	Recording time	No. mediators
Italian-English	262	2587'	4
Italian-Arabic	163	2265'	5
Italian-Chinese	81	1195'	2
Italian-French	22	266'	5
<b>Tot.</b>	<b>548</b>	<b>6977' (over 100 h)</b>	<b>17</b>

The corpus is a large one in its category and for some years now we have tried to reorganize it in a way as to make it more easily available and searchable (see Niemants, forthcoming). We have used the annotation tools EXMARaLDA and ELAN to start recreating our transcripts as audio annotations.

In this paper, we report on this experience by showing our preliminary analysis of a sub-set of 65 encounters, whose transcript is now aligned with audio, for a total amount of 1.124 minutes (almost 19 hours) interaction. Our work has gone in two directions: 1. Exploring which lexical items may constitute interesting starting points to retrieve relevant structures of interaction; 2. Finding ways of extracting audioaligned transcripts to use in e.g. specific interpreting training or learning activities. In our presentation, we will show and discuss one example of each type of search.

The problems we are coping with, and would like to discuss in our presentation, have to do with a potential change of perspective following the new possibilities we have to search the corpus. While interaction analysis is based on the observation and qualitative analysis of talk sequences, finding “interesting” sequences by corpus search (rather than human eye) opens a series of questions involving data annotation and encoding.

### *Riferimenti bibliografici*

- ASTON, G. (2015). Learning phraseology from speech corpora. In LENKO-SZIMANSKA, A., BOULTON, A. (Ed.) *Multiple affordances of language corpora for data-driven learning*. Amsterdam: Benjamins, 6384.
- MASON, I. (Ed.) (1999). *Dialogue Interpreting. The Translator*, 5/2. Special Issue.
- NIEMANTS, N. (forthcoming) Des enregistrements aux corpus: transcription et extraction de données en milieu médical. In *Meta*, 63(3). Special Issue.
- SINCLAIR, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- STEENSIG, J., HEINEMANN, T. (2015). Opening up codings? In *Research on Language and Social Interaction*, 48(1), 20-25.
- STIVERS, T. (2015). Coding social interaction: A heretical approach in Conversation Analysis? In *Research on Language and Social Interaction*, 48(1), 1-19.

FEDERICA CUGNO<sup>1</sup>, FEDERICA CUSAN<sup>1</sup>

<sup>1</sup> Università di Torino

## Digital Toponymy. Il progetto MAPforUS e l'Atlante Toponomastico del Piemonte Montano on line

La tecnologia informatica applicata agli studi umanistici ha registrato negli ultimi anni un avanzamento esponenziale tanto da determinare i canoni di una progettualità che, nei diversi ambiti, è sollecitata a indirizzare la propria ricerca verso il modello digitale per quanto riguarda sia l'archiviazione dei dati sia la loro consultazione. Il digitale e il multimediale nella progettualità umanistica non solo riorientano le pratiche di ricerca, ma, in linea con i dettami del *web* e dei *social networks*, restituiscono al ruolo di prodotti di primo piano materiali prima considerati secondari o il cui valore era al più meramente strumentale: è questo, per esempio, il caso dei documenti sonori (registrazioni audio e video) e fotografici realizzati in un contesto di indagine linguistica sul campo.

Rispetto a questi elementi i due progetti che intendiamo presentare, l'*Atlante Toponomastico del Piemonte Montano* (ATPM) e *MAPforUS: Mapping Alpine Place-names for Upward Sociality*, sono esemplari. Il primo è un trentennale progetto di ricerca del Dipartimento di Studi Umanistici dell'Università di Torino, ancora in corso d'opera, il cui obiettivo è la registrazione dei toponimi di tradizione orale di tutti i comuni piemontesi appartenenti alle Unioni montane: un patrimonio linguistico e culturale attualmente documentato da 77 mila denominazioni di luogo elicitate mediante inchieste sul campo svolte da raccoglitori locali con un numero variabile di informatori (da qualche decina a più di un centinaio) appartenenti alle comunità indagate. I dati così raccolti, schedati e stoccati in un database in locale, sono successivamente pubblicati in monografie, con tiratura limitata. Il secondo è un progetto di Ateneo – cofinanziato da un istituto bancario per il triennio 2017-2019 – che eredita dall'ATPM il quadro teorico, ma che da esso si distingue per il carattere inclusivo della sua prassi di ricerca, per la presenza di un gruppo di lavoro con competenze assai differenziate (vi collaborano linguisti, antropologi, geografi, web developers), e, non da ultimo, per l'importanza della componente multimediale che corredata il dato linguistico. L'intento di MAPforUS è l'allestimento di mappe interattive dedicate a una ventina di borgate piemontesi – ma il modello di ricerca è potenzialmente esportabile ed espandibile – scelte sulla base di criteri geografici (carattere insediativo, qualità delle pertinenze), demografici (vitalità demografica, dinamismo valutato in termini di sostituzione generazionale) e linguistici (le

località coinvolte rientrano nell'area di diffusione delle minoranze linguistiche storiche tutelate a livello regionale e nazionale); sulle mappe sono rappresentati e interpretati una serie di dati georeferiti sia linguistici (i toponimi) sia extralinguistici (dati geografici, storici, antropologici, demografici) reperiti nel corso dell'indagine sul campo. Le mappe sono via via allocate nel sito web del Progetto e proposte alla libera consultazione e all'uso da parte degli utenti. Tale fruizione ha imposto un ripensamento riguardo all'impiego del materiale sonoro che si è risolto in una sua decisiva valorizzazione, con ricadute anche in termini di ampliamento delle possibilità di analisi: se nelle inchieste promosse dall'ATPM la registrazione del toponimo è richiesta come documento di verifica per la sua trascrizione ortografica e fonetica, in MAPforUS la modalità del sonoro costituisce una delle possibilità offerte all'utente di consultare l'archivio e muoversi sulla mappa e restituisce al dato toponimico un'immediatezza e una replicabilità che difficilmente una trascrizione, per quanto accurata e impressionistica, potrebbe superare, soprattutto per un pubblico di non specialisti. In MAPforUS ciascun toponimo aggancia fino a tre documenti audio (e altrettanti audio-video) che possono riguardare: 1) la sua diversa pronuncia da parte degli informatori – il Progetto, del resto, si svolge su un ritaglio territoriale caratterizzato da una spiccata micro variazione linguistica che è un aspetto assolutamente interessante da rilevare, finché l'azione livellante dei codici dominanti (il piemontese e l'italiano sulle lingue di minoranza e l'italiano sul piemontese) lo consente – 2) la sua pronuncia in un contesto di frase, per verificare l'attivarsi di determinati fenomeni paralinguistici e morfosintattici, aspetto del tutto trascurato dagli studi linguistici correnti; 3) la sua ricorrenza in una *performance* narrativa: un toponimo è pienamente interpretabile se è analizzato nell'interazione discorsiva, in cui è richiamato dagli informatori.

### *Riferimenti bibliografici*

- BERRY, D. M. & FAGERJORD A. (2017). *Digital Humanities: Knowledge and Critique in a Digital Age*. Cambridge: Polity Press.
- BURDICK A. & DRUCKER J., *et al.* (2012). *Digital Humanities*. Cambridge (Mass.)-London: Massachusetts Institute of Technology Press; trad. it. (2014) *Umanistica digitale* (Ed. M. BITTANTI). Milano: Mondadori.
- CINI M. (Ed.) (2017). *Humanities e le altre scienze. Superare la disciplinarietà*. Roma: Carocci editore.
- CUGNO F. & CUSAN F. (2018). MapforUS: toponimia tradizionale e nuovi modelli di socialità. In MARCATO G. (Ed.). *Dialetto e società*. Padova: Cleup, 49-56.
- GENRE A. & JALLÀ D. (1990). L'Atlante Toponomastico del Piemonte Montano. In *Nowelle Revue d'Onomastique*, 15-16, 169-179.

LAPICCIARELLA ZINGARI V. (2012). I paesaggi delle voci in un percorso di frontiera. In *Lares*, 78(1-2): «Fonti orali. Approcci e dialoghi tra antropologia e storia orale (Gennaio-Agosto 2012)». Firenze: Casa Editrice Leo S. Olschki, 525-568.

MALFATTI F. (2011). Sviluppo sostenibile, etnografia, GIS e fonti orali: la sfida dei saperi tradizionali. In KRIPPENDORF K. & LA ROCCA G. (Eds.). *Ricerca qualitativa e giovani studiosi. Atti del Convegno Internazionale RiQGios 2011*, Università di Enna. Palermo: Social Book, 221-229.

PENNACINI C. (2005). *Filmare le culture. Un'introduzione all'antropologia visiva*. Roma: Carocci editore.

EMANUELA CRESTI<sup>1</sup>, MASSIMO MONEGLIA<sup>1</sup>, ALESSANDRO PANUNZI<sup>1</sup>

<sup>1</sup> Università di Firenze

## The LABLITA spoken corpora collection and the Language into Act Theory

The Language into Act Theory (L-AcT) has been developed in Italy since the eighties and aims at providing a framework suitable for the corpus-based study of spontaneous speech (Cresti, 2000, 2018; Cresti & Moneglia, 2018; Cresti *et al.*, 2018b).

L-AcT was applied to the Italian LABLITA corpus (1.135.000 transcribed words, 107.000 reference units text/sound aligned) (Cresti *et al.*, 2018a) and has been heavily tested in the collection and annotation of Romance corpora: C-ORAL-ROM (Cresti & Moneglia, 2005), C-ORAL-BRASIL (Raso & Mello, 2012), Cor-DiAL (Nicolas Martinez, 2013). Transcripts complies with the CHAT LABLITA Format (Cresti & Moneglia, 2005), text / sound alignment and acoustic analysis have been achieved through the software WINPITCH (Martin, 2003, 2015). Both the LABLITA corpus and the C-ORAL-ROM Italian collection, for the qualities of their corpus design, constitute reference corpora for Italian.

This framework was also used for grounding the cross-linguistic comparison of Information Structure in spontaneous speech (Moneglia & Raso, 2014). To this end the IPIC Data Base was created and applied to Italian and Brazilian Portuguese tagged corpora (Panunzi & Mittmann, 2014). The extension of IPIC to Spanish was recently achieved (Nicolas Martinez & Lombán, forthcoming). American English has been also tagged according to the same methodology on the basis of a selection of S. Barbara corpus (Du Bois *et al.*, 2000; Cavalcante & Ramos, 2016).

Within the Austinian tradition, L-AcT assumes that the utterance is the counterpart of a speech act and constitutes the primary reference unit for the analysis of speech. Its main novelty with respect to Austin is to consider that the spoken activity manifests through prosodic devices, specifically for what regards the core levels of Illocutionary force and Information structure (IS). Therefore, the processing of prosody is assumed as a mandatory step for the identification in the flow of speech of both Utterances and their Information Units.

L-AcT foresees the systematic correspondence between stretches of speech ending with a *terminal prosodic break* and the accomplishment of an utterance, and, within the utterance, between chunks segmented by a *non-terminal break* and information functions (Cresti & Moneglia, 2005; Izre'el & Mettouchi, 2015).

The presentation will sketch the spoken corpora previously mentioned and will focus on the methodology for the detection of prosodic breaks and its validation (Danieli *et al.*, 2004; Moneglia *et al.*, 2010; Raso & Mittmann, 2009).

The relevance of prosodic breaks will be highlighted tracing back to the IPO tradition that assumes that intentionally performed prosodic cues are significant to perception ('t Hart *et al.*, 1990; Firenzuoli, 2003). Current trends in the L-Act framework concerning both perceptual and automatic detection of breaks will be also referred (Barbosa & Raso, 2018).

Sequences ending with a terminal break strictly correspond to speech Reference Units (Izre'el *et al.*, forthcoming) and may correspond to *utterances* matching with one speech act (90% of cases in the above corpora) or to *stanzas*, corresponding to the expression of a flow of thought (Chafe, 1970). Utterances and stanzas are the reference entities suitable for the identification of syntactic and semantic relations in speech.

The added value of the annotation of terminal breaks for the use of spontaneous speech data in linguistic research will be considered. Their detection in the acoustic source determines the alignment unit and specifies the higher level of linguistic annotation for parsing the speech flow into information units and syntactic chunks. In other words, the annotation of terminal breaks in the acoustic source determines the reference units and specifies the higher level of linguistic annotation.

### *Riferimenti bibliografici*

- BARBOSA, P. & RASO, T. (2018). Spontaneous Speech Segmentation: Functional and Prosodic Aspects with Applications for Automatic Segmentation A. *Revista de Estudos da Linguagem*, 26(4), 1361-1396.
- CAVALCANTE, F. & RAMOS, A. (2016). The American English spontaneous speech minicorpus. Architecture and comparability. *CHIMERA*, 3(2).
- CHAFE, W. (1970). *Meaning and the Structure of Language*. Chicago: CUP.
- CRESTI, E. (2000). *Corpus di italiano parlato*. Firenze: Accademia della Crusca.
- CRESTI, E. & MONEGLIA, M. (Eds.) (2005). *C-ORAL-ROM. Integrated reference corpora for spoken Romance languages*. Amsterdam: Benjamins.
- CRESTI, E. & MONEGLIA, M. (2018). The illocutionary basis of Information Structure. Language into Act Theory (L-Act). In ADAMOU, E., HAUDE, K. & VANHOVE, M. (Eds.). *Information structure in lesserdescribed languages: Studies in prosody and syntax*. Amsterdam: Benjamins, 359-401.
- CRESTI, E., MONEGLIA, M. & PANUNZI, A. (2018a). The LABLITA Corpus & the Language into Act Theory: analysis of Viterbo excerpts. In DE DOMINICIS, A. (Ed.). *Speech audio archives: preservation, restoration, annotation, aimed at supporting the linguistic analysis*. Contributi del Centro Linceo "Beniamino Segre", 137. Roma: Bardi Edizioni, 47-63.

- CRESTI, E., GREGORI, L., MONEGLIA, M. & PANUNZI, A. (2018b). The Language into Act Theory: A Pragmatic Approach to Speech in Real-Life. In KOISO, H. & PAGGIO, P. (Eds.). *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), LB-ILR2018 and MMC2018 Joint Workshop: Language and Body in Real Life Multimodal Corpora & Multimodal Data in the Online World*. Paris: ELRA, 20-25.
- DANIELI *et al.* (2004). Evaluation of Consensus on the Annotation of Prosodic Breaks in the Romance Corpus of Spontaneous Speech C-ORAL-ROM. In DRAXLER, C., VAN DEN HEUVEL, H. & SCHIEL, F. (Eds.). *Proceedings of LREC 2004*. Paris: ELRA, 1513–1516.
- DU BOIS, J., CHAFE, W., MEYER, Ch. & THOMPSON, S. (2000). *Santa Barbara Corpus of Spoken American English Part 1*, LDC2000S85. Philadelphia: Linguistic Data Consortium.
- FIRENZUOLI, V. (2003). *Le Forme Intonative di Valore Illocutivo dell'Italiano Parlato: Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA)*. PhD Thesis, Università di Firenze.
- 'T HART, J., COLLIER, R. & COHEN, A. (1990). *A Perceptual Study on Intonation. An Experimental Approach to Speech Melody*. Cambridge: Cambridge University Press.
- IZRE'EL, S., MELLO, H., PANUNZI, A. & RASO, T. (Eds.) (forthcoming). *In search for a reference unit of spoken language: a corpus driven approach*. Amsterdam: Benjamins.
- IZRE'EL, S. & METTOUCHI, A. (2015). Representation of speech in CorpAfroAs. Transcriptional strategies and prosodic units. In METTOUCHI, A., VANHOVE, M. & CAUBET, D. (Eds.). *Corpus-based Studies of Lesserdescribed Languages: The CorpAfroAs Corpus of Spoken Afro-Asiatic Languages*. Amsterdam: Benjamins: 13–41.
- MARTIN, Ph. (2003). Winpitch corpus, a software tool for alignment and analysis of large corpora. Paris: University Paris 7 Denis Diderot. <https://www.researchgate.net/publication/228702364>
- MARTIN, Ph. (2015). *The structure of spoken language. Intonation in romance*. Cambridge: CUP.
- MONEGLIA, M., RASO, T., MITTMANN, M. & MELLO, H. (2010). Challenging the Perceptual Relevance of Prosodic Breaks in Multilingual Spontaneous Speech Corpora: C-ORAL-BRASIL / C-ORAL-ROM. In *Speech Prosody Satellite Workshop: Prosodic Prominence Perceptual and Automatic Identification*. Chicago. Université de Neuchâtel.
- MONEGLIA, M. & RASO, T. (2014). Notes on the Language into Act Theory. In RASO, T. & MELLO, H. (Eds.). *Spoken corpora and linguistics studies*. Amsterdam: Benjamins, 468-494.
- NICOLAS MARTINEZ, C. (2012). *Cor-DiAL*. Madrid: Liceus.
- NICOLAS MARTINEZ, C. & LOMBÁN, M. (2018). The Spanish spontaneous speech minicorpus. Architecture and comparability. *CHIMERA*, 5(2).
- PANUNZI, A. & MITTMANN, M. (2014). The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In RASO, T. & MELLO, H. (Eds.). *Spoken corpora and linguistics studies*. Amsterdam: Benjamins, 129-151.
- RASO, T. & MELLO, H. (Eds.) (2012). *C-ORAL-BRASIL I: Corpus de referência de português brasileiro falado informal*. Belo Horizonte: UFMG Press.

RASO, T. & MITTIMANN, M. (2009). Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. *Revista de Estudos da Linguagem*, 17, 73-91.

CLAUDIA ROBERTA COMBEI<sup>1</sup>, GIOVANNA MAROTTA<sup>2</sup>

<sup>1</sup>Università di Bologna, <sup>2</sup>Università di Pisa

## Where do accents come from? Factors affecting the degree of foreign-accented Italian

Foreign accents reflect our identities and origins, but at the same time they convey information about our sociolinguistic background. In addition, the degree of foreign accent may reflect how open we are to embrace the language and the culture of the foster country, or it could simply be a predictor of the quality and the quantity of the input we have had for the specific language we are learning (Flege, 2009). In recent years, non-native speech has been a topic of continuous research interest in theoretical literature, applied linguistics, and speech technology. Many studies in second-language acquisition (SLA) have covered production and perception of non-native speech, often targeting English (e.g. Piske, MacKay, & Flege, 2001; Moyer, 2013; etc.). Nonetheless, in the last decades, various studies have also addressed foreign-accented Italian (e.g. Marotta, 2008; Marotta & Boula de Mareuil, 2010; Bianchi & Calamai, 2012; Pellegrino, Caruso, & De Meo, 2015; etc.).

Differently from previous studies on spoken L2 Italian, we investigate whether speaker- and context-related factors affect the degree of the perceived foreign accent. Such an approach may be advantageous in explaining and understanding the production and perception of non-native Italian speech from a sociolinguistic perspective. At the same time, the results of this study may be beneficial in the field of second language acquisition, especially in teaching Italian pronunciation to foreigners.

In this respect, and in order for the stimuli to meet balancedness and representation criteria, when selecting the speakers from the *CorAIt* corpus (Combei, 2017), we controlled for the following variables: the type of speech sample (i.e. read and spontaneous), the speaker's mother-tongue (L1), the occupation (all speakers were enrolled as regular or exchange students in degree, Ph.D., or specialization programmes in Bologna), the age (ages range from 20 to 30 years old), the gender, the self-assessed proficiency level in Italian (ranging from B1 to C2 levels, based on *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment-CEFR1*), 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), 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 French, Romanian, Spanish, English, German, and

Russian, compared to native varieties of Italian. Ten-to-twelve-seconds-samples of read and spontaneous speech (*cf.* Combei, 2017, for a complete description of the speech samples), produced by a gender-balanced set of 42 speakers – 6 for each L1 –, were included in a perceptual experiment delivered to 288 native speakers of Italian; the stimuli were previously validated by five expert listeners who assessed the degree of foreign accent (inter-rater reliability: ICC (A, 1) = 0.762; p-value 0; 95% – confidence interval), they were automatically randomized for each experimental session, and they could be played twice.

Keeping in mind that also a wide range of factors regarding listeners could affect the outcomes of this study, we built a computer-based experiment that was conducted over the Internet on a *WordPress*<sup>1</sup> platform. This allowed us to reach a large and varied pool of listeners. The experiment was password-protected, and the IP verification blocked multiple submissions. The first section of the experiment was compulsory and qualifiable for further rounds; thus, the questions were designed to collect background information on the listeners. We sampled for the region of origin, the gender, the L1 (only monolingual<sup>2</sup> 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. French, Romanian, Spanish, English, German, Russian), the degree of familiarity with the foreign accents investigated (i.e. French, Romanian, Spanish, English, German, Russian), and experience in the field of linguistics (i.e. having attended B.A. and/or M.A. courses in Linguistics). An adequate number of participants for each of the levels of the afore-mentioned variables was collected. In the accent rating experiment, listeners were asked to assess the degree of foreign accent using a 6-gradients Likert scale (i.e. from 0 = “no accent/native speaker of Italian” to 5 = “very strong accent”). Before analysing the results, an inter-rater reliability test was performed on the listeners’ judgements, yielding a positive outcome: ICC (A,1) = 0.68<sup>3</sup>; p-value: 0; 95% – confidence interval. After a treatment coding of the categorical variables, we verified the strength of the effect of listener-related factors on the score listeners 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

---

<sup>1</sup> A description of this platform and its functions is available online at this webpage: <https://it.wordpress.com> (accessed on the 12th of September 2018).

<sup>2</sup> 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).

<sup>3</sup> According to Cicchetti’s (1994) guidelines for the interpretation of inter-rater agreement measures, an ICC value of 0.68 is considered ‘good’.

variable, namely, the higher the listener's education level, the more severely they will judge the speakers' degree of accentedness (p-value: 0.03).

Following Piske, MacKay, & Flege (2001), we focus on assessing which factors contributed to the degree of foreign accent. For such purpose, we fitted Linear regression models, where we tested the effect of all the variables under control on the accentedness score (i.e. the dependent variable). The weakest factors in predicting the degree of foreign accent were gender and the age of onset and they were excluded from the final model. 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 reading errors that might have arisen due to conflicting orthographic norms between the speakers' L1 and Italian. At the same time, in spontaneous speech the speaker will tend to use familiar and less problematic constructions.

Next, all other grouping variables and levels being comparable, Romanian speakers outstripped the other participants, since, most of the times they were rated with "no accent" or "mild accent" scores. On the other hand, 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 (e.g. Flege, 1995). In fact, 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); this, could advocate for shifting the view towards the role of input in SLA (Flege, 2009). However, having learned Italian in a specific learning setting is not often associated to a better pronunciation score (p-value: 0.06). Additionally, 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. 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.

The outcomes of the final model – consisting in the following independent variables: the type of speech sample, the speaker's L1, the Italian learning method, the presence or absence of pronunciation training, the length of stay in Italy, and the proficiency level in Italian – may be summarised as follows:  $F(13,22) = 11.34$ ,  $p = 0.00$ ;  $R^2 = 0.87$ ;  $Adj. R^2 = 0.79$ , Residual standard error: 0.59 on 22 degrees of freedom. According to our data, the degree of foreignaccented 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, and the type of oral production. However, in

order to confirm the afore-mentioned associations at a larger scale, in future research, other accents should be investigated as well.

### *Riferimenti bibliografici*

- BIANCHI, F. & CALAMAI, S. (2012). Voci italiane e straniere a confronto. Indagine sugli stereotipi associati agli accenti stranieri. In *Cambio*, 2(4), December 2012, 149-164.
- 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). CorAIIt – A non-native speech database for Italian. In BASILI, R., NISSIM, M. & SATTA, G. (Eds.), *Proceedings of the Fourth Italian Conference on Computational Linguistics, 11-13 December 2017, Rome*, 113-118, Torino: Accademia University Press.
- 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.
- GROSJEAN, F. (2010). *Bilingual: Life and Reality*. Cambridge, Massachusetts; London, England: Harvard University Press.
- FLEGE, J. (2009). Give Input a Chance! In PISKE & YOUNG-SCHOLTEN (Eds.). *Input Matters in SLA*. Bristol: Multilingual Matters, 175-190.
- MAROTTA, G. (2008). Sulla percezione dell'accento straniero. In LAZZERONI *et al.* (Eds.). *Diachronica et synchronica. Studi in onore di Anna Giacalone Ramat*. Pisa: ETS, 327-347.
- MAROTTA, G. & BOULA DE MAREUIL, P. (2010). Persistenza dell'accento straniero. Uno studio percettivo dell'italiano L2. In SCHMID *et al.* (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.
- PELLEGRINO, E., CARUSO & DE MEO. (2015). Discriminating between pathological and foreign- accented speech: an acoustic-perceptual study on Italian. In VAYRA, M. *et al.* (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*. Milano: Officinaventuno, 271-287
- PISKE, T., MACKAY, I.R.A. & FLEGE, J. (2001). Factors affecting degree of foreign accent in an L2: a review. In *Journal of Phonetics*, 29(2), 191-215.

CLAUDIO BARALDI<sup>1</sup>, CHIARA BALLESTRI<sup>1</sup>, VITTORIO IERVESE<sup>1</sup>

<sup>1</sup>Università degli Studi di Modena e Reggio Emilia

## **Combining speech and images: the case of the SHARMED archive**

This presentation regards the archive created as output of the European Erasmus + project SHARMED (Shared Memories and Dialogues) ([www.sharmed.eu](http://www.sharmed.eu)). This project was realised in 48 classrooms in Italy, Germany and United Kingdom and aimed at promoting innovative experiences of teaching and learning in the classroom, with specific consideration for respect of differences and promotion of dialogue, and empowering children's contributions to classroom interaction. The SHARMED activities in the classroom enhanced: (1) the production of children's stories about their personal memories, and (2) the facilitation of comparison and sharing of these stories through dialogic communication. The production and negotiation of children's stories was based on the use of private photographs, representing children's memories.

The SHARMED archive includes 300 children's photographs, 424 children's short written texts about these photographs, 100 videos of children presenting photographs and 39 videos recorded during the classroom activities. It has been provided for both innovative activities in the classroom and research. The SHARMED archive combines speech with images, in the double form of photographs and videorecordings: this presentation regards this combination of speech and images.

Before analysing this combination, it is important to understand the limitations and problems of

“completeness” of the SHARMED archive, which depend on three factors: ethical issues, structural complexity and problems of translation. Ethical problems are particularly relevant when speech and images of children are combined, as this combination requires a complex procedure of information and request for consent, which limits the possibility of using the images, in particular the videorecorded images. Structural complexity regards the quantity and quality of the combinations between video-recordings of activities and photographs. The high quantity of video-recordings and their quality requirements, together with the ethical issue, determine the necessity to select them and therefore to elaborate criteria for this selection. Finally, the subtitling of Italian and German videorecordings in English was made necessary by the European level of dissemination of the archive. While subtitling may be effective in rendering speech in another language, it is not only a long and time-consuming process,

but it may also be a factor of discouragement for many teachers and educators with a limited competence in understanding English.

This presentation highlights the advantages of facing these relevant problems in order to combine speech and images, also explaining the way in which this type of archive works. The combination of speech and images is relevant for three reasons.

**Speech is enhanced through the use of images** (photographs). In social research, the use of images to stimulate speech is primarily based on photo-elicitation techniques that include photographs into research interviews (e.g., Harper, 2002). In these interviews, the images are expected to generate insights that do not necessarily or exclusively correspond to those obtained through a verbal inquiry. Photo-elicitation is expected to enhance new attribution of meanings, expression of emotions, and elicitation of relevant information. While these aspects are also included in the use of photographs during the SHARMED activities, this use differs significantly from photoelicitation for three aspects (Baraldi & Iervese, 2017). First, only children have access to the knowledge of photographs, therefore they can talk about them without external elicitation. Children are the responsible and competent agents for talking about the photographs. Second, photographs are used as a medium to convey children's stories of personal memories, rather than an interview. Third, photographs are used as a medium to facilitate the interactional production of stories, rather than individual meanings or feelings. Images are not only used to elicit children's attribution of meanings, but also to stimulate dialogue between children, including questions and comments about the photographs and the stories about them, and the interlacement of different stories.

**The analysis of speech is improved through the use of images** (video-recording). Videorecordings add meaning to speech for three aspects. First, in the classroom, the production of speech in the interaction involves a great number of participants, video-recordings can support the identification of the speakers, thus enhancing the possibility of analysing the interactional production of stories. Second, video-recordings make possible the identification of non-verbal actions produced in the interaction, which can either support the explanation of turns of speech, or complete the analysis of the organized sequence of turns by analysing the combination of turns of talk and non-verbal turns. Third, video-recordings can allow the analysis of the environment surrounding speech, i.e. the way in which the room is filled with people and things, including the spatial disposition of people. Thus, video-recordings can support the explanation of the distribution of the opportunities of speaking.

**Speech can explain images** (photographs and video-recordings), thus allowing the linguistic and sociological analysis of the interaction.

Speech gives meaning to photographs in narrative formats (Norrick, 2012).  
**3.1.1** Narrative formats provide the description of photographs, giving meaning to what they represent. Thus, the medium "photograph" takes a form: the loose

connection between unknown internal elements becomes a tight connection based on meaning (Luhmann, 2012). **3.1.2.** Narrative formats produce knowledge about what is behind the frame of the photograph, such as its story, the situation in which it was taken, the wider context of its production. This production of knowledge can regard either facts, data or events (declarative narratives), or general knowledge of ideas, meanings and concepts (semantic narratives) (Berntsen & Rubin, 2012). It also provides the opportunity for comments and different stories.

Speech gives meanings to video-recordings of interactions. **3.2.1** Types and modes of narrative production in interactions: (a) range of themes and children's preferences in narrating them; (b) children's competence in active negotiation of their identity (Bamberg, 2011); (c) children's ways of negotiating the rights to tell a story (Baraldi & Iervese, 2017). **3.2.2** Facilitation of narrative production in interactions: (a) invitations to tell and questions that support children's narratives; (b) types of feedback that support and recognise children's stories; (c) personal contributions that enrich narratives; (d) ways of dealing with children's unpredictable initiatives (Baraldi, 2014).

In the SHARMED archive, the combination between production of speech and use of images has two relevant effects. First, it provides an integrated analysis: speech explains what images cannot show, while images show what the words cannot say. Second, it expands the possibility of exploiting both of them: the use of images can enhance different stories, while the telling of stories can refer to different images.

### *Riferimenti bibliografici*

- BAMBERG, M. (2011). Narrative practice and identity navigation. In HOLSTEIN, J.A. & GUBRIUM, J.F. (Eds.). *Varieties of Narrative Analysis*. London: Sage, pp. 99-124
- BARALDI, C. (2014). Children's participation in communication systems: A theoretical perspective to shape research. In WAREHIME, N. (Ed.), *Soul of Society: A Focus on the Life of Children and Youth*. Bingley: Emerald Group Publishing, 63-92.
- BARALDI, C. & IERVESE, V. (2017). Narratives of memories and dialogue in multicultural classrooms. In *Narrative Inquiry*, 27(2), 398-417.
- BERNTSEN, D. & RUBIN, D.C. (2012). Understanding autobiographical memory: An ecological theory. In BERNTSEN, D. & RUBIN, D.C. (Eds.). *Understanding autobiographical memory: Theories and approaches*. Cambridge: Cambridge University Press, 290-310.
- HARPER, D. (2002). Talking about pictures: a case for photo elicitation. In *Visual Studies*, 17(1), 13-26.
- LUHMANN, N. (2012). *Theory of Society (vol. 1)*. Stanford: Stanford University Press.
- NORRICK, N. (2012). Remembering for narration and autobiographical memory. In *Language and Dialogue*, 2(2), 193-214.

GRAZIANO TISATO<sup>1</sup>

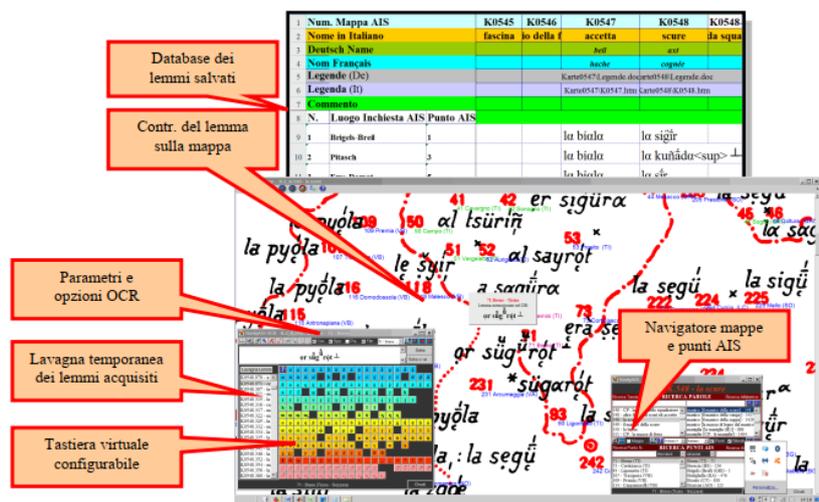
<sup>1</sup> Istituto di Scienze e Tecnologie della Cognizione (ISTC-CNR)

## Acquisizione Digitale dell'Intero AIS

Si presenta in questo articolo la terza fase del progetto NavigAIS, nato nel 2009, con lo scopo di acquisire digitalmente il testo dell'intero AIS ([www3.pd.istc.cnr.it/navigais](http://www3.pd.istc.cnr.it/navigais), [www3.pd.istc.cnr.it/navigais-web](http://www3.pd.istc.cnr.it/navigais-web)), l'Atlante ItaloSvizzero dei linguisti svizzeri K. Jaberg e J. Jud (*Sprach- und Sachatlas Italiens und der Südschweiz*, 1928-1940), e di mettere a disposizione della comunità il relativo *database*.

Il lavoro si è svolto nell'ambito del progetto *AIS Reloaded* (Romanisches Seminar, Università di Zurigo, [www.rose.uzh.ch/de/forschung/AIS-reloaded.html](http://www.rose.uzh.ch/de/forschung/AIS-reloaded.html)), finanziato da SNSF (*Swiss National Science Foundation*), che oltre alla digitalizzazione dell'AIS intende analizzare l'evoluzione diacronica dei dialetti italoromanzi avvenuta in cent'anni nel Canton Ticino e nei Grigioni, con una campagna di registrazioni nelle medesime località, oggetto dell'inchiesta AIS, e sullo stesso questionario (intermedio) usato a quel tempo.

Fig. 1 – NavigAIS: versione con OCR. Sullo sfondo, la tavola AIS 548 ([NavigAIS-web 548.71](http://NavigAIS-web/548.71)). A sinistra, in basso, la tastiera virtuale (configurabile) per la correzione degli eventuali errori del riconoscitore. In alto, il database in cui le sequenze riconosciute sono automaticamente salvate nella locazione opportuna.



L'articolo affronta le problematiche del simbolismo AIS, l'architettura e le funzionalità dell'OCR (*Optical Character Recognition*) realizzato *ad hoc* fra il 2016 e il

2017, la metodologia usata per la validazione dei risultati ed il controllo dei lemmi del database, la valutazione obiettiva delle *performance* del riconoscitore (considerando tutti i caratteri errati, omessi e spuri), ed i tempi di acquisizione dei lemmi.

Si presenta anche, come primo risultato del lavoro, l'inventario fonetico effettivo dell'AIS ricavato dai dati di circa il 25% delle mappe AIS, funzionale per il miglioramento delle performance dell'OCR, in quanto consente di individuare sia i caratteri errati che le combinazioni di caratteri e diacritici non permessi nella sequenza restituita dall'OCR, e di accorciare i tempi di acquisizione del 10% circa.

Il progetto NavigAIS si è sviluppato in tre fasi distinte:

La prima fase è stata completata nel 2009 con la realizzazione della versione digitale e navigabile dell'AIS (scaricabile a [www3.pd.istc.cnr.it/navigais](http://www3.pd.istc.cnr.it/navigais)), resasi indispensabile per portare a termine l'Atlante Multimediale dei Dialetti Veneti (Tisato *et al.*, 2013, [www.pd.istc.cnr.it/amdy](http://www.pd.istc.cnr.it/amdy)). Le mappe sono state preliminarmente separate su due livelli sovrapponibili, l'uno contenente il solo testo, e l'altro i confini geografici e l'orografia, in maniera tale che non interferissero poi con il riconoscimento da parte dell'OCR.

La seconda fase (2014) ha riguardato l'implementazione della versione *online* dell'AIS ([www3.pd.istc.cnr.it/navigaisweb](http://www3.pd.istc.cnr.it/navigaisweb)), per permetterne la consultazione diretta *web*, e l'accesso facilitato ad una mappa e ad un punto AIS, dando un opportuno indirizzo in una qualsiasi pagina HTML (ad es. [mappa 1327 punto 177](#)).

La terza fase è stata portata a termine fra 2016 e 2017, con la creazione di un opportuno OCR (Fig. 1-2), per permettere la digitalizzazione del testo dell'intero atlante AIS (metà delle 1705 mappe saranno completate entro la fine del 2019).

In generale, le sequenze fonetiche dell'AIS, che devono trascrivere lingue di regioni diverse, sono caratterizzate da un numero molto alto di caratteri a bassa frequenza, con un andamento che può essere descritto dalla legge di Zipf (*Human Behaviour and the Principle of Least-Effort*, 1949). Conseguenza inevitabile di questo andamento è il peggioramento dei risultati di un qualsiasi OCR basato su un motore probabilistico.

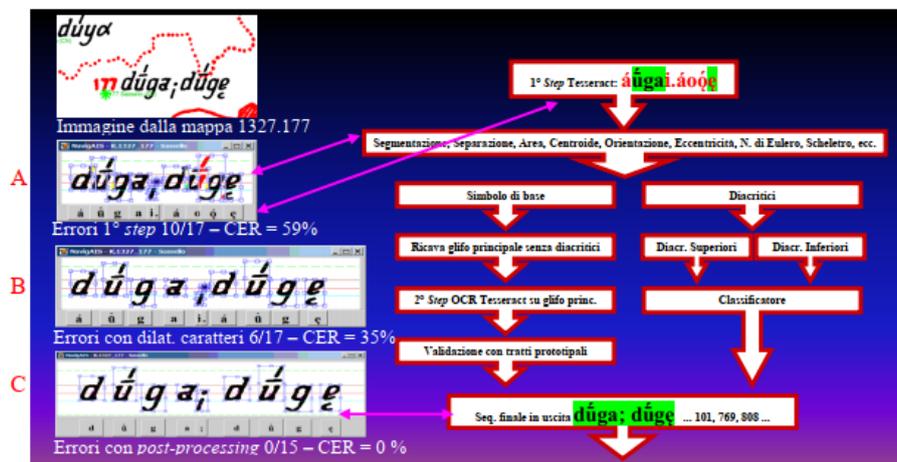
Il numero di simboli diversi da riconoscere (circa 1400), la scrittura manuale, il corsivo (circa 70° di inclinazione) dei caratteri, la struttura dei diacritici, articolata su 7 livelli, e l'incertezza di assegnazione di un lemma ad un punto AIS rendono inutilizzabili gli OCR (si veda [S. Lücke, 2018](#)).

Manca d'altra parte per le trascrizioni AIS un modello del linguaggio, che possa predire le probabilità di successione di un carattere (e bigrammi, trigrammi, ecc.), così come manca un dizionario delle parole esistenti nelle lingue in questione, che permetta di individuare la parola corretta fra le possibili sequenze fornite dall'OCR.

L'approccio innovativo adottato per il progetto AIS è stato la riduzione della complessità del riconoscimento, con un'elaborazione che separa automaticamente il glifo di base dei caratteri dai relativi diacritici sovrastanti e

sottostanti, per cui i simboli distinti da riconoscere calano a circa 70 simboli vocalici e 134 consonantici.

Fig. 2 – Sulla destra lo schema dell'OCR implementato. Sulla sinistra un es. di errori (contando sia i glifi di base che i diacritici restituiti dall'OCR) in uscita dal 1° step da Tesseract (A), e il miglioramento che avviene con la dilatazione dei caratteri (B), e con le correzioni automatiche nelle fasi finali di *post-processing* (C).



Il glifo di base può essere in questo modo trattato con un riconoscitore tradizionale (in questo caso è stato scelto Tesseract, perché è di libero uso e perché mette a disposizione il codice).

Tesseract ha subito un addestramento preliminare su un campione limitato di frasi etichettate. A quel punto inizia il processo di riconoscimento che automaticamente percorre i 407 punti dell'inchiesta AIS per ognuna delle 1705 mappe dell'atlante. Il primo passaggio serve principalmente a segmentare la catena fonetica nei singoli caratteri e relativi diacritici (Fig. 2A). Segue una fase di *pre-processing* con un software Matlab, per spaziare opportunamente i caratteri e migliorare il tasso di riconoscimento (Fig. 2B), e per l'estrazione di varie caratteristiche geometriche e topologiche, e di misure utili ad alimentare lo *step* intermedio, ed infine il processo si completa con la fase finale di *post-processing* (Fig. 2C).

Nella fase centrale i glifi senza diacritici sono assoggettati nuovamente all'OCR Tesseract, mentre i diacritici seguono un percorso apposito e sono classificati in base alle proprietà geometriche e topologiche elementari.

Nel passo finale, caratteri e diacritici sono validati e corretti con una serie di 350 regole implementate con Matlab, ed infine la sequenza è salvata in un *database*. In quest'ultima fase, un supervisore può intervenire per correggere eventuali errori con la tastiera (configurabile dall'utente in base alle esigenze) di Fig. 1.

Per quanto riguarda i risultati (che saranno esposti in dettaglio nell'articolo), si riassume qui solo la percentuale di errore CER (*Character Error Rate*), che si è

ottenuta su un campione rappresentativo di 14 mappe AIS complete per circa 6000 parole complessive (100000 caratteri), e che varia da 1.35% a 5.87% per mappa con valore medio del 3.65% (molto inferiore a quanto inizialmente ipotizzabile per questa tipologia di caratteri), e con tempi di acquisizione (ed eventuali correzioni) inferiori alle 2.30 ore per mappa, il che dà sufficienti garanzie che il completamento dell'acquisizione dell'intero AIS possa avvenire in circa due anni/uomo.

ALICE BETTI<sup>1</sup>

<sup>1</sup>Università degli Studi di Pavia

## Archivi orali e socioprosodia: un'analisi dell'interrogativa globale a Ripacandida (PZ) e Siena (SI)

**Obiettivo:** Questo lavoro si propone di indagare l'intonazione delle interrogative globali (d'ora in avanti IG) di Ripacandida (PZ) e Siena (SI) allo scopo di verificare l'esistenza di parametri soprasegmentali marcati diatopicamente e connessi a variazioni di tipo segmentale che permettano di identificare il parlante dal punto di vista diatopico. Questa analisi si è dunque voluta inserire all'interno di un più ampio ventaglio di ricerche, sviluppatosi a partire dagli ultimi anni del secolo scorso, che ruotano attorno alla variazione geoprosodica (cfr. Sorianello, 2006). Nello specifico si risponderà alle seguenti domande di ricerca:

È possibile individuare una variazione diatopica nell'intonazione dell'italiano regionale parlato a Ripacandida (PZ) e Siena (SI)?

Se sì, questa variazione diatopica dell'intonazione è accompagnata anche da una variazione a livello segmentale e, in particolare, nello spazio vocalico utilizzato dai parlanti delle due aree geografiche?

**I dati:** Inserendosi in un paradigma sociofonetico (cfr. Calamai, 2015) e allo scopo di ottenere dei dati il più possibile rappresentativi del parlato regionale in questione, oltre alla dimensione diatopica si è tenuto conto anche delle variazioni diafasiche e diastratiche; sono infatti stati elicitati sia dati semi-spontanei (conversazioni guidate) sia dati di parlato letto prodotti da parlanti, bilanciati per quanti riguarda il sesso (4M-4F). Durante le registrazioni è stato utilizzato un registratore Eivotor (modello EIVOTOR PORO65KE12), in formato .wav, con una frequenza di campionamento di 44.1 kHz e un sampling rate a 18bit.

Ciascuna registrazione è stata articolata in 3 fasi distinte a seconda del metodo di elicitazione utilizzato e, dunque, del tipo di dato linguistico ricavato. In ciascuna parte di questo lavoro i parlanti erano consapevoli di essere registrati e hanno firmato l'apposito modulo di consenso e divulgazione anonima. La prima fase delle registrazioni consisteva in una vera e propria intervista semi-strutturata a ciascun informante preso singolarmente. Gli argomenti trattati nelle interviste sono stati vari; a partire da dati più formali come le informazioni biografiche del parlante – l'età, gli studi condotti e i viaggi fatti – si è arrivati a suscitare risposte più personali ed emozionali attraverso la richiesta di raccontare, in pieno stile laboviano classico (cfr. Labov, 1963), un'esperienza piacevole e una triste. I dati

ottenuti durante le interviste individuali non sono stati oggetto di analisi socioprosodica ma sono stati utilizzati per tracciare il profilo sociolinguistico del parlante e per permettere al soggetto di sentirsi a proprio agio con il registratore e il ricercatore superando, per quanto possibile, il *paradosso dell'osservatore* (cfr. Labov, 1972). La seconda fase è consistita nella lettura di una lista di frasi, elicitando un parlato letto, mentre l'ultima fase della registrazione è stata caratterizzata dall'uso di *Map Task*, un particolare metodo di registrazione che permette di elicitare parlato semi-spontaneo all'interno di una situazione di gioco nella quale interagiscono due parlanti alla volta.

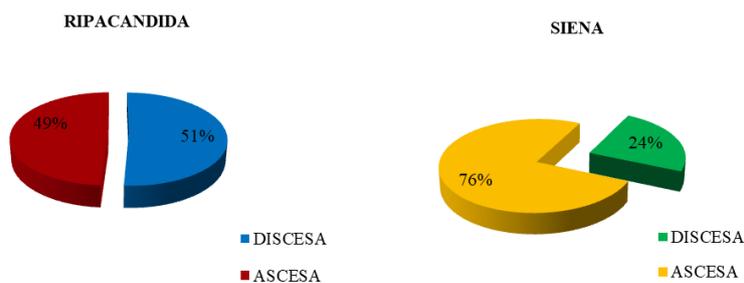
Sono state realizzate 32 registrazioni per un totale di 05h 32' 53", con una media di 01h 45' 13" per soggetto. I file audio sono stati catalogati in base ai seguenti parametri: area geografica di provenienza, fase di elicitazione, sesso e informante. Dei file audio elicitati mediante *Map Task* è stato deciso di prendere in considerazione esclusivamente la modalità frasale oggetto della presente analisi, dunque l'IG, escludendo qualsiasi altra tipologia di frase. Ciò è stato possibile attraverso una linea di annotazione (TextGrid) del *software* PRAAT (cfr. Boersma & Weenick, 2018), dove sono state annotate ortograficamente, previa accurata selezione, solo le IG pronunciate dagli informanti. Con queste produzioni è stato costituito il *Corpus Map Task (CMT)*, separato e distinto da quello composto dalle IG prodotte durante la seconda fase, quella della lettura della lista di frasi, definito per una corretta gestione degli archivi orali *Corpus Frasi Fisse (CFF)*. I dati dei *corpora* costituiscono dunque l'archivio orale *CIGReS*, già disponibile sulla pagina <https://github.com/AliceBetti/CIGReS>.

**Analisi:** I dati contenuti nell'archivio orale *CIGReS* sono stati analizzati in chiave sia qualitativa che quantitativa: per quanto riguarda l'analisi qualitativa, condotta attraverso l'osservazione della curva intonativa e per mezzo dell'applicazione totalmente automatica delle etichette Intsint (cfr. Hirst, 2006, 2007; Hirst & Di Cristo, 1998) attraverso l'apposito *plugin\_momel-intsint (version 2016-03-29)* di PRAAT (cfr. Hirst, 2007), sono stati presi in considerazione sia il *Corpus Frasi Fisse* sia quello contenente le IG elicitate mediante il *Map Task*.

Attraverso l'analisi e il confronto dell'ultima etichetta Intsint (assegnata automaticamente dal *plugin*) di ciascuna IG contenuta nel *Corpus Frasi Fisse* è stato possibile osservare, ad esempio, come le produzioni di Siena, in linea anche con quanto osservato da studi intonativi precedenti ma di diverso impianto teorico (cfr. Marotta & Sorianello, 2001), presentino un'ascesa finale della curva di F0 molto forte e sistematica, a differenza del contorno melodico terminale delle IG prodotte a Ripacandida che invece non si caratterizza per una così elevata ascesa finale, in termini hertziali, di F0. Inoltre è stato osservato che le produzioni di Ripacandida si caratterizzano per una tendenza a un contorno melodico terminale discendente, non in linea con quello previsto di *default* in letteratura per questa modalità frasale. Di fatto, nel *CFF* il 36% delle IG di Ripacandida presenta un profilo intonativo finale con valori di F0 in discesa, rispetto al 9% delle produzioni senesi, mentre nel *CMT* questa tendenza a un profilo terminale

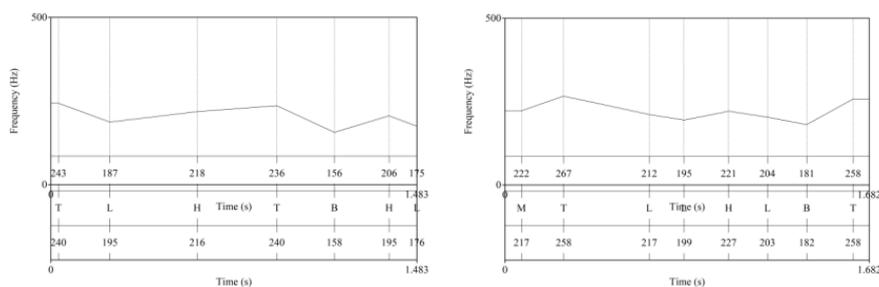
discendente aumenta notevolmente, tant'è che nella Figura 1 si può osservare come più della metà delle IG prodotte dai parlanti di Ripacandida nella terza fase di elicitazione presenta un contorno melodico terminale discendente.

Fig. 1 *La distribuzione della qualità del profilo intonativo terminale delle IG contenute nel Corpus Map Task*



I due grafici sottostanti (Fig. 2) esemplificano bene quanto affermato: infatti, a parità di condizione diastratica e diafasica, le due produzioni presentano un contorno melodico terminale molto differente.

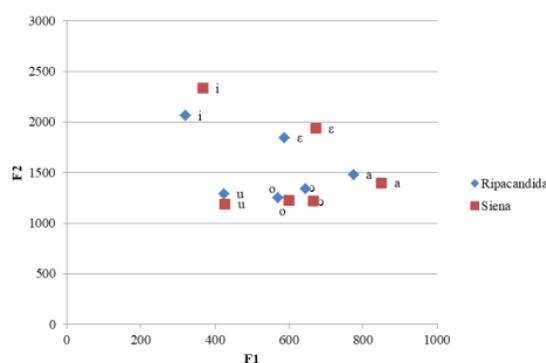
Fig. 2 *A sinistra il profilo intonativo di un parlante di Ripacandida, a destra quello prodotto da un senese*



L'analisi quantitativa invece è stata effettuata, per rendere possibile un confronto preciso tra i valori ottenuti, esclusivamente sui dati linguistici contenuti all'interno del *CFE*. In questo caso per ogni file audio sono state create tre linee di annotazione (TextGrid) attraverso l'uso di PRAAT: una dedicata alla frase IG, una alle singole parole lette e l'ultima alle vocali toniche. Nello specifico, per analizzare lo spazio vocalico in diatopia l'analisi è stata condotta sui valori della prima e della seconda formante, calcolati sul punto medio delle vocali toniche tramite uno script di PRAAT appositamente modificato per il presente lavoro e disponibile all'interno del corpus *CIGReS* su Git.hub. I valori di F1 e F2 sono stati rappresentati in grafici a dispersione che hanno permesso di osservare (Fig. 3) come lo spazio vocalico vari in relazione a fattori diatopici: infatti lo spazio vocalico dei parlanti di Ripacandida risulta essere molto più ristretto e arretrato

rispetto a quello dei parlanti di Siena, soprattutto per ciò che riguarda le vocali anteriori. Si è potuto poi osservare, grazie all'analisi statistica condotta tramite il *software* IBM SPSS20, come vi sia una variazione statisticamente significativa, legata a fattori diatopici, riguardante i valori di F1 e F2 delle vocali /a/ e /i/. Per ciò che concerne la vocale /a/ i parlanti di Ripacandida (M=774,53 Hz) presentano una F1 significativamente inferiore di 75,35 Hz rispetto ai parlanti di Siena (M=849,88 Hz), con  $p=0,002$  e indice  $d$  di Cohen pari a  $-0,515$ . La seconda formante invece risulta essere significativamente maggiore di 82,11 Hz nelle produzioni dei parlanti di Ripacandida (M=1479,88 Hz) rispetto ai parlanti di Siena (M=1397,77 Hz), con  $p=0,002$  e indice  $d$  di Cohen uguale a  $0,573$ . Passando poi alla vocale /i/, per quanto riguarda F1 è stato osservato che le produzioni di Siena (M=367,96 Hz) presentano un valore medio significativamente superiore a quelle di Ripacandida (M=320,32 Hz) di 47,64 Hz, con  $p=0,013$  e indice  $d$  di Cohen di  $-0,749$ . Anche per quanto riguarda F2, l'analisi statistica ha evidenziato valori significativamente maggiori di 271,28 Hz nelle produzioni di Siena (M=2332,36 Hz) rispetto a quelle di Ripacandida (M=2061,08 Hz), con  $p=0,005$  e indice  $d$  di Cohen pari a  $-0,846$ .

Fig. 3 Grafico a dispersione raffigurante lo spazio vocalico tonico dei parlanti di Siena e Ripacandida nella produzione delle IG



**Conclusioni e prospettive:** I risultati ottenuti dimostrano l'esistenza di una sistematica variazione nell'intonazione delle IG dei due italiani regionali, accompagnata anche da una variazione (segmentale) dei valori delle variabili continue F1 e F2 e dunque dello spazio vocalico usato dai parlanti delle due diverse aree geografiche; una puntuale indagine articolatoria tramite UTI potrebbe dunque aiutare a comprendere la causa di queste variazioni acustiche.

### *Riferimenti bibliografici*

- BOERSMA, P. & WEENINK, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved from <http://www.praat.org/>.
- CALAMAI, S. (2015). *Introduzione alla sociofonetica*. Roma: Carocci editore.
- CELATA, C. & CALAMAI, S. (2014). *Advances in Sociophonetics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- FELLONI, M. C. (2011). *Prosodia sociofonetica. L'italiano parlato e percepito a Parma*. Milano: FrancoAngeli.
- HIRST, D. (2006). Phonetic and Phonological Annotation of Speech Prosody. In SAVY, R. & CROCCO, C. (Eds.). *Analisi Prosodica. Teorie, modelli e sistemi di annotazione*, Atti del II Convegno Nazionale della Società Italiana di Scienze della Voce (AISV), Università degli Studi di Salerno, 30 novembre-2 dicembre 2005. Torriana: EDK, 24-29.
- HIRST, D. (2007). A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. In TROUVAIN, J. (Ed.) *Proceedings of the XVIth International Conference of Phonetic Sciences*, Saarbrücken, 6-10 agosto 2007. Saarbrücken: Universität des Saarlandes, 1233-1236.
- HIRST, D. & DI CRISTO, A. (Eds.) (1998). *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- LABOV, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- LABOV, W. (1973). Lo studio del linguaggio nel suo contesto sociale. In GIGLIOLI, P. P. & FELE, G. (Eds.) (2000). *Linguaggio e contesto sociale*. Bologna: il Mulino, 207-232.
- LABOV, W. (1994). *Principles of Linguistic Change: Internal Factors*, London: Blackwell, vol. 1.
- MAROTTA, G. & SORIANELLO, P. (2001). La teoria autosegmentale nell'analisi dell'intonazione interrogativa in due varietà di toscano (Lucca e Siena). In ALBANO LEONI, F., KROSSBAKKEN, E. S., SORNICOLA, R. & STROMBOLI, C. (Eds.). *Dati empirici e teorie linguistiche*, Atti del XXXIII Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI), Napoli, 28-30 ottobre 1999. Roma: Bulzoni, 177-204.
- MAROTTA, G. & SARDELLI, E. (2009). Prosodiatopia: parametri prosodici per un modello di riconoscimento diatopico. In FERRARI, G., BENATTI, R. & MOSCA, M. (Eds.). *Linguistica e modelli tecnologici di ricerca*, Atti del XL congresso internazionale di studi della società di linguistica italiana (SLI). Vercelli, 21-23 settembre 2006. Roma: Bulzoni, 411-435.
- SORIANELLO, P. (2006). *Prosodia*, Roma: Carocci editore.
- VAN DER HARST, S. (2011). *The Vowel Space Paradox. A Sociophonetic Study on Dutch*. Utrecht: LOT.

CAMILLA BERNARDASCI<sup>1</sup>, DIETER STUDER-JOHO<sup>1</sup>

<sup>1</sup>Università di Zurigo

## L'Archivio fonografico zurighese tra passato e futuro

L'Archivio fonografico dell'Università di Zurigo (*Phonogrammarchiv der Universität Zürich, PAZ*), fondato nel 1909 dal germanista Albert Bachmann (1863-1934) con la collaborazione di Louis Gauchat (1866-1942) e di Robert von Planta (1864-1937), è l'archivio di fonti orali più antico della Svizzera. Ad oggi accoglie quasi 4400 supporti (dischi di cera e di gommalacca, nastri a filo magnetico, nastri magnetici e audiocassette) con registrazioni provenienti dalle quattro regioni linguistiche della Svizzera. Tra gli scopi dell'Archivio vi sono la digitalizzazione, la trascrizione e la pubblicazione del patrimonio sonoro svizzero, per garantire da una parte la sua conservazione, dall'altra la sua fruibilità in ambito accademico, ma non solo.

Fin dalla sua fondazione, il *PAZ* collabora con il *Phonogrammarchiv* di Vienna, che possedeva uno dei primi registratori in circolazione, il fonografo viennese, uno strumento meccanico che permetteva di incidere su dischi di cera. Le campagne d'inchiesta del *PAZ* non si concentravano esclusivamente sulle parlate dei grandi centri, ma cercavano di documentare anche i dialetti delle zone rurali della Svizzera. La prima pubblicazione risale al 1914 quando, in occasione dell'Esposizione nazionale tenutasi a Berna, fu pubblicata nella collana dell'Archivio viennese una scelta di trascrizioni di testi dialettali svizzeri (Gröger, 1914). Dopo la collaborazione con Vienna, non disponendo ancora di uno strumento di registrazione proprio, il *PAZ* intraprese una nuova sinergia con la *Lautabteilung der Preussischen Staatsbibliothek* di Berlino, diretta da Wilhelm Doegen (1877-1967), che disponeva di un grammofono a motore. Grazie al supporto tecnico di Berlino, il *PAZ* poté registrare, tra il 1924 e il 1929, ca. 225 dischi (cfr. Studer-Joho, 2015). Dal 1932 il *PAZ* comincia ad acquistare registratori propri: prima un *Domofon*, che incideva su dischi di gelatina e, in seguito, un registratore a filo magnetico *Webster Wire Recorder*. Dopo la metà degli anni Cinquanta, al passo con il progresso tecnologico, il *PAZ* si procura dei registratori che incidevano su nastro magnetico (*Revox A36, A 77, B 77, Uber Report Serie, Stellavox SP 7*), con l'ausilio dei quali fu possibile organizzare, negli anni seguenti, svariate campagne d'inchiesta. L'inizio dell'era digitale ha comportato, già a partire dagli anni Novanta, una notevole svolta nell'ambito delle tecniche di registrazione: con il superamento delle cassette audio si è cominciato a registrare su supporti DAT (*Digital Audio Tape*) e, in seguito, con registratori (ad es. *Fostex, Marantz e Zoom*)

che permettessero di trasferire i dati digitalmente su un computer, senza perdita di qualità.

Nei suoi primi 90 anni di vita, quindi, il *PAZ* si occupa principalmente di raccogliere nuovo materiale sonoro, di trascriverlo e di pubblicarlo. «Ora [a partire dagli anni Duemila], le nuove possibilità tecnologiche prevedono inoltre nell'agenda dell'Archivio Fonografico un nuovo compito, e cioè la necessità di garantire la salvaguardia dei documenti storici mediante il loro riversamento su supporti digitali» (Gadmer, Schmid & Sigg, 2003: 4). Nel 2008 ha quindi preso il via un progetto di catalogazione in forma elettronica dei fondi archivistici, con lo scopo, da una parte, di facilitare l'amministrazione e l'inventariazione dei fondi, dell'altra di rendere più semplice l'accesso ai documenti orali sia per gli specialisti, sia per il grande pubblico. Dal 2010 si è proceduto alla digitalizzazione sistematica del materiale d'archivio: sia dei supporti sonori (iniziata, in realtà, già alcuni anni prima grazie al sostegno degli archivi di Vienna e Berlino), sia dei documenti cartacei che permettono di documentare la storia del *PAZ* (tramite la corrispondenza e i verbali delle riunioni della commissione) e di ricostruire la genesi delle diverse campagne d'inchiesta (grazie ai verbali delle inchieste e alle trascrizioni delle registrazioni). Parallelamente ci si è occupati della (ri)edizione di due serie di registrazioni, una risalente alla collaborazione viennese (*Schweizer Aufnahmen*, 2002) e una incisa in occasione dell'Esposizione nazionale del 1939 (*Voci della Svizzera*, 2012, già pubblicata nel 1939 col titolo *Voci della Patria*, ma senza trascrizione fonetica IPA).

La presenza in formato digitale delle registrazioni e delle relative trascrizioni ha permesso di allestire il *corpus* PAZTeK (*Phonogrammarchiv Zürich Text-Korpus*) – che comprende diversi *sottocorpus* di testi dialettali (come ad esempio il *corpus* LM, cfr. Bernardasci & Schwarzenbach, 2016) –, interrogabile *online*, dove è possibile consultare i *files* sonori allineati alle rispettive trascrizioni (<https://www.uzh.ch/phonogrammarchiv/static/PAZTeK/>).

Lo scopo del nostro contributo è quindi quello di far conoscere il *PAZ* e le sue risorse ancora troppo poco note e sfruttate, ma utili, in particolare, per la ricerca linguistica, in quanto documentano stadi antichi di un grande numero di dialetti svizzeri. Si potranno inoltre ascoltare alcune registrazioni rappresentative delle diverse tecniche di registrazione e delle varietà linguistiche conservate. Al tempo stesso, si metteranno in evidenza le difficoltà incontrate nel processo di digitalizzazione (e le soluzioni adottate), nonché le tappe che hanno portato all'allestimento del PAZTeK.

### *Riferimenti bibliografici*

BERNARDASCI, C. & SCHWARZENBACH, M. (Eds.) (2016). *St ri, stralùsc e stremizzi. RegISTRAZIONI dialettali nella Svizzera italiana (1929)*. Quaderni del Bollettino Storico della Svizzera italiana, 13. Bellinzona: Salvioni.

FLEISCHER, J. & GADMER, T. (Eds.) (2002). *Schweizer Aufnahmen (Series 6/1 e 6/2), Enregistrements Suisses, Ricordi sonori Svizzeri, RegISTRAZIUNS Svizras (français, italiano, rumantsch) (Series 6/3)*. Wien: Tondokumente aus dem Phonogrammarchiv der Österreichischen Akademie der Wissenschaften.

GADMER, T., SCHMID, S. & SIGG, R. (Eds.) (2003). I tesori sonori dell'Archivio Fonografico dell'Università di Zurigo. In DE DOMINICIS, A., MORI, L. & STEFANI, M. (Eds.). *Costituzione, gestione e restauro di corpora vocali*. Roma, 99-104.

GRÖGER, O. (1914). *Schweizer Mundarten*. Wien: XXXVI. Mitteilung der PhonogrammArchivs-Kommission der kaiserl. Akademie der Wissenschaften in Wien. *Stimmen der Schweiz = Voix de la Suisse = Voci della Svizzera = Vuschs da la Svizra*. In *historischen Aufnahmen. In Text und Ton herausgegeben für die Landi 1939* (2012), riedizione a cura di GLASER, E. & LOPORCARO, M., rielaborate da STUDER-JOHO, D., SCHWARZENBACH, M., GRÜNERT, M., MAÎTRE, R. & DIÉMOZ, F. in collaborazione con SCHMID, S. Frauenfeld.

STUDER-JOHO, D. (2015). The Early Swiss Dialect Recording Collection "LA" (1924-1927): A description and a Work Plan for Its Comprehensive Edition. In HOPFMANN, R. & TROUVAIN, J. (Eds.). *First International Workshop on the History of Speech Communication Research, Dresden, September 4-5, 2015*. Dresden, 87-94.

MICHELE TOSS<sup>1</sup>

<sup>1</sup>Fondazione Museo storico del Trentino

## **Conservazione e schedatura delle fonti orali: alcune riflessioni**

La conservazione delle fonti orali e la loro valorizzazione costituiscono temi di grande attualità. Concreta - e necessaria - è l'urgenza di salvaguardare su supporto digitale il vasto patrimonio orale raccolto non solo dai pionieri della storia orale ma anche da parte dei "ricercatori minori", i cui archivi sonori rischiano di andare persi o di diventare illeggibili a causa del degrado del loro supporto fisico. Solitamente le interviste venivano utilizzate da chi le aveva realizzate e solo recentemente si è iniziato a porre la questione, in tutta la sua complessità e delicatezza, del loro uso da parte di "terze persone", del loro deposito negli istituti di conservazione e, di conseguenza, dell'accesso a tale documentazione. I ricercatori di oggi (e, in maniera ancora più preponderante, quelli di domani) hanno/avranno la necessità di consultare queste tipologie di documenti. L'irruzione infine nella disciplina storica di un approccio legato alla Public History rappresenta un ulteriore elemento di riflessione per la valorizzazione della fonte orale. I documenti audiovisivi, grazie anche alla loro duttilità, sono ampiamente utilizzati non solo all'interno delle pubblicazioni e dei contributi scientifici, ma anche in quelle forme di restituzione che presuppongono linguaggi differenti, come ad esempio i percorsi espositivi, i documentari, le serie televisive, le piattaforme online della storia e della memoria, i laboratori didattici, le applicazioni per smartphone. La creazione di strumenti di corredo che permettono di orientarsi più facilmente e più velocemente all'interno delle campagne di interviste o degli archivi sonori diventa quindi una necessità.

Tramite un bando della Fondazione Caritro di Trento, la Fondazione Museo Storico del Trentino ha avviato un progetto – che è ancora in corso – finalizzato alla consultazione online e alla schedatura delle fonti orali presenti all'interno del proprio archivio (si tratta di più di 1000 ore di registrazione). Nella prima fase del lavoro la maggior parte del patrimonio audiovisivo è stato digitalizzato e messo in sicurezza su server dotati di un sistema di backup. A questo proposito la conservazione della fonte orale digitale presuppone una serie di problematiche non secondarie: da un lato l'utilizzo di formati che garantiscano la possibilità di lettura nel tempo, dall'altro la messa a disposizione di spazi virtuali, in continua espansione, dove depositare i file digitali. A causa del miglioramento delle attrezzature tecniche per le registrazioni le dimensioni delle interviste realizzate digitalmente sono sempre maggiori e ciò comporta lo stoccaggio in server sempre più grandi.

Il progetto si è quindi concentrato sulle modalità da adottare per la descrizione prima e successivamente per la messa a disposizione sul web (seppur in modo limitato) di una tipologia di materiale così sensibile e che pone importanti questioni legate alla privacy degli intervistati e alla responsabilità etica e professionale dell'istituzione che conserva la fonte orale. Durante l'intervento si metteranno in luce, problematizzandoli, tre aspetti principali:

**presentazione dei criteri adottati per la catalogazione e descrizione dei campi che compongono la schedatura delle interviste:**

L'intento è di offrire uno strumento di consultazione che sia maneggevole e, al tempo stesso, rigoroso. La schedatura si divide in quattro sezioni: la prima parte è dedicata alle informazioni di contesto della registrazione, la seconda agli aspetti tecnici, la terza ai dati biografici dell'intervistato mentre la quarta - forse la più delicata e complessa - si focalizza sugli aspetti contenutistici. In quest'ultimo caso si è deciso di lavorare su differenti livelli, passando dal generale (identificazione di alcune parole chiave) al particolare (abstract dell'intervista e una scaletta tematica corredata di minutaggio). A tal proposito è stata attivata una collaborazione con l'Università degli Studi di Trento, ed in particolare con il prof. Andrea Giorgi (professore di archivistica e socio AISO) per individuare i principali nodi teorici e testare la schedatura adottata.

**il lavoro di schedatura:** le registrazioni, nella maggior parte dei casi, non contengono informazioni preliminari (alcune, inoltre, non sono corredate da liberatorie) e nessuno ne conosce il contenuto esatto; questo comporta un considerevole impiego di risorse per ascoltare in tempo reale centinaia e centinaia di ore di interviste.

**messa a disposizione del materiale online:**

le schedature realizzate sono state inserite in un database informatico creato ad hoc e successivamente messe a disposizione della collettività (<http://900trentino.museostorico.it/Leinterviste-dalla-Cineteca-della-FMST>) su una piattaforma digitale che contiene altri database e differenti fonti storiche che dialogano tra loro (<http://900trentino.museostorico.it/>). Si è voluto mettere a disposizione un agile strumento di consultazione dove l'utente ha la possibilità di conoscere le informazioni legate all'intervista, di svolgere una serie di indagini tramite apposite maschere di ricerca e di visionare qualche minuto della registrazione.

Questo modello di gestione degli archivi orali con la messa online di informazioni sensibili comporta un'importante riflessione da parte degli storici orali e degli istituti di conservazione. Oltre alle considerazioni legate alla presenza o meno della liberatoria che accompagna la fonte orale, ci sono anche altre questioni che devono essere poste. I mezzi informatici ci consentono di usufruire, direttamente da casa, di una quantità di informazioni e di archivi sempre maggiore, ma la fonte orale può essere messa a disposizione a chiunque? cosa si può mostrare in pubblico? Le interviste, spesso, contengono osservazioni private e giudizi personali: possono essere affidati al web liberamente?

Il database informatico e la schedatura dei materiali sono in continuo aggiornamento con lo scopo di mettere a disposizione del pubblico, in maniera consapevole, un ricco patrimonio di documenti legato alle fonti orali che rappresenta, sempre di più, un bene comune.

HYEON-YEOL IM<sup>1</sup>

<sup>1</sup>Chung-ang University

## **Study on the Pronunciation of Korean TTS, conditioned on the presence or absence of a morpheme boundary**

The purpose of the study is to examine how well the Korean TTS reflects the phonological phenomenon and to suggest some ways to improve the Korean TTS in terms of the phonological phenomenon. In particular, this study examines the differences between a case that the morpheme boundary affects a phonological phenomenon and a case that does not. Based on these examination results, this study represents that the current TTS algorithm needs to be improved. Mainly, it will be focused that the pronunciation of string of /n/ and /l/, which vary depending on whether or not the morpheme boundary is present.

In this study, the words related to the /n/ and /l/ string were composed of sentences, and each sentence was put into Korean TTS to listen to the pronunciation result. The number of words related to the /n/ and /l/ string was 232. In order to extract this word list, this study used the frequency survey data of Cho (2003). In Cho (2003), 58,437 words actually used in the corpus were presented. Among these words, there were 544 words related to the /n/ and /l/ string. Among them, 232 words remained except for words sharing the same word base.

The TTS device for the experiment utilized Naver Papago's TTS(NP) and Google Translator's TTS(GT). In Korean TTS, Naver Papago's TTS and Google Translator's TTS are most commonly used. This is because NP and GT are expected to be the best in technology. We did not use special equipment to record the pronunciation of TTS. This study confirmed the pronunciation of TTS by listening to the sound delivered by the earphone. This is because the study needs to pay attention to the difference in the level of phonemic variation rather than the slight difference in the level of phonetic variation.

The pronunciation result was summarized by calculating the ratio of the number of words pronounced by standard pronunciation to the number of non-standard pronunciation. By type, this study classified the cases into the cases of lateralization and nasalization.

First, let's look at how NP and GT pronounced the words related to the /n/ and /l/ string with standard pronunciation. In NP, 39 out of 232 words produced non-standard pronunciation. Of the 39 non-standard pronunciations, 6 /ll/ errors and 33 /nn/ errors were counted. GT has produced non-standard

pronunciations in 53 of 232 words. Of the 53 non-standard pronunciations, /ll/ errors were 19, and /nn/ errors were 34. Overall, the number of non-standard pronunciation of GT compared to NP many times. There were a lot of GT in the /ll/ error. There were 26 items that showed the same error on NP and GT. There were 13 non-standard pronunciation in NP, but standard pronunciation in GT. Conversely, GT had a non-standard pronunciation, but NP had 25 standard pronunciation. NP and GT showed the same error in the same item, but NP and GT showed different errors in different items.

NP had 6 /ll/ errors among 39 non-standard pronunciation. However, there were 33 /nn/ errors. The /nn/ error appeared mainly at the words including a Sino-Korean lexeme used like suffix. Mostly, there were 23 errors in the words that used '론(/lon/, a theory)'. There were many errors in '량(/lyang/, quantity)(4)', '력(/lyeok/, force)(3)', '령(/lyeong/, order)(2)' and '료(/lyo/, rate)(1)'. If so, does all the words that include the Sino-Korean lexeme used like suffixes have problem? However, test results do not seem to be such a problem. Even if we look at other words in which '론(/lon/, theory)' is used, 8 words produce standard pronunciation. /ll/ error appeared only in a few words. But all of these words appear only in very infrequent words.

GT had 19 /ll/ errors among 53 non-standard pronunciation. However, there were 35 /nn/ errors. /ll/ errors appeared much higher on GT than on NP. It can appear in words that are used infrequently. But, it can be seen that non-standard pronunciations are also produced in words that are frequently used, such as '현란하다(/hyeonlanhada/, dazzling)', '판로(/panlo/, market)', '선로(/seonlo/, track)', and

'마천루(/macheonlu/, skyscraper)'. Also, the number of /nn/ errors was 34, similar to NP's /nn/ errors.

Here again, 20 non-standard pronunciations appeared in the words related to '론(/lon/, theory)'. However, there were also 9 non-standard pronunciations in words related to '력(/lyeok/, force)'. Other than that, non-standard pronunciation appeared in '령(/lyeong/, order)(2)', '량(/lyang/, quantity)(1)', '료(/lyo/, rate)(1)', '룡(/lyong/, dragon)(1)'.

If we look at the above test results, we can confirm the current status of pronunciation related to the /n/ and /l/ string as follows.

First of all, the /n/ and /l/ string were producing largely appropriate pronunciation in both NP and GT, despite the possibility of two different types of pronunciation depending on the environment. However, both NP and GT were producing non-standard pronunciations in some words. However, there were many /nn/ errors compared to /ll/ errors. It is presumed that the processing of /n/ and /l/ is basically calculated as /ll/, and that /nn/ is calculated in a specific environment. That is, basically, the /n/ and

/l/ string is processed as lateralization, and if there is a morpheme boundary, it is treated as nasalization. Therefore, it is very important to recognize whether a morpheme boundary is included between the /n/ and /l/ string. In a fairly

large number of words, it is appropriate to perceive it, because in some words morpheme boundary fail to recognize and /nn/ is pronounced as /ll/ according to the general rule.

Then, considering the above two situations, NP and GT seem to need two measures to improve the accuracy of TTS. In the words related to the string of /n/ and /l/, general words without morpheme boundary should be extended to word-based standard pronunciation data. Next, the input of morphological information for the lexical phrases such as '론(/lon/, theory)', '력(/lyeok/, power)',

'량(/lyang/, quantity)' and '료(/lyo/, rate)', which are frequently used as a suffix, becomes more clear, So that they can be converted to /nn/ in a combined word.

However, this study did not deal with the pronunciation of loanwords that do not have the standard pronunciation in Korean. This means that it is necessary to research separately the pronunciation of loanwords of Korean TTS pronunciation.

PIETRO BRAIONE<sup>1 2</sup>, CARLO CECCHETTO<sup>3 1</sup>, CARLO GERACI<sup>4 3 5 6</sup>, ROLAND PFAU<sup>7</sup>, JOSEP QUER<sup>8 9</sup>

<sup>1</sup>Università di Milano-Bicocca, <sup>2</sup>CINI, <sup>3</sup>CNRS, <sup>4</sup>IJN, <sup>5</sup>DEC, <sup>6</sup>ENS, <sup>7</sup>University of Amsterdam, <sup>8</sup>ICREA, <sup>9</sup>Universitat Pompeu Fabra

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

**Background.** Video technology has been widely available for several decades, and since the late 1990s also in digital format (Castells 2001). This technological breakthrough opened huge opportunities for the study of sign languages (SLs) in all domains of social sciences (Kusters et al. 2017), which in turn speeded up the process of recognition of local SLs by national governments (DeMeulder et al. in press). The video materials document the oral memories of visual languages (i.e., SL).

**The project.** The (EU funded) Sign-Hub project aims at providing a digital platform to archive and give access to: on-line SL grammars, SL assessment tools, SL atlas of linguistic structures, and life narratives of elderly signers documenting the history of Deaf communities. Sign-Hub is generating a huge amount of video data that are and will be used in various ways.

**Goals.** We focus here on the following topics:

- Opportunities and challenges of use and re-use of historical narratives for language documentation and culture preservation purposes
- Risks and challenges associated with long-term preservation and streaming of video materials and their metadata

**Discussion.** The Sign-Hub platform is designed to fulfill two requirements: it must 1) provide content creators (grammar writers, atlas compilers and SL historians) with a user-friendly interface to upload and manage their contents (Fig. 1); and 2) offer a pleasant experience to the final users (a.o., linguists, students, historians, Deaf people, general audience, etc.).

As a case study, we discuss **life narratives of elderly signers**. In the course of the second half of the 20th century, several events impacted the lives of deaf people and the destiny of European Deaf communities. Some are part of the World history (e.g., WW2, Shoah, Civil Wars or Cold War). Others are specific to Deaf communities (e.g., experiences in Deaf institutes, situation of ethnic, or gay/lesbian minorities) or the lives of Deaf individuals (e.g., first encounter with a deaf person, cases of language discrimination).

Data for this part of the project come from already existing recordings (for France and Israel) or from newly recorded interviews (for Germany, Italy,

Netherlands, Spain, Turkey). As for the former, we provide: 1) a digitization of old videotapes, 2) subtitling and voicing in the local language, 3) English subtitles. As the latter, we provide 1) subtitling in the local spoken language and in English, 2) partial linguistic annotation in ELAN (Crasborn & Sloetjes 2008; see Fig. 2), 3) Metadata in CMDI format specific to SL (Crasborn 2017).

While the memory of 20<sup>th</sup>-century events from the perspective of Deaf communities will be preserved, the very same material will also be used as:

- documents about the history of Deaf communities
- samples of the language of elderly signers
- the basis to study linguistic change across Deaf generations
- material for other social sciences (e.g., anthropology, sociology)
- materials for documentaries for the general public

Digital format is extremely perishable and subject to fast technological changes. We are faced with a double challenge: 1) archiving the video materials for long-term preservation; 2) making them accessible via web-interfaces. Our solution is to use a double storage policy: the same video is generated in two formats: a high quality format is stored in public archives for long-term preservation; a compressed form is used for streaming and immediate access to content.

Figure 1: Screenshot of the workspace for content creator

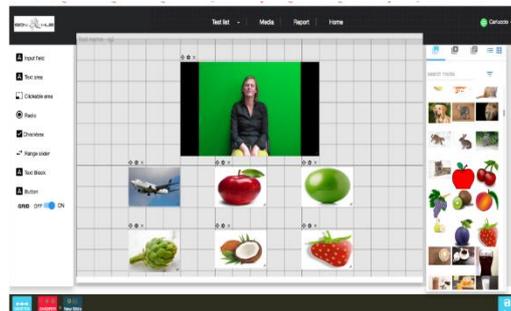
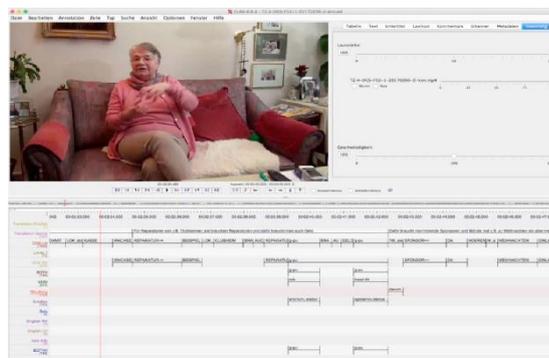


Figure 2: Screenshot of ELAN workspace



*Riferimenti bibliografici*

CASTELLS (2001). *The internet galaxy*. OUP.

CRASBORN & SLOETJES (2008). Enhanced ELAN functionality for sign language corpora. In *LREC*.

CRASBORN (2017). *SL-TL: Sign language data archived at The Lang. Archive*.

DEMEULDER *et al* (In press). *Recognizing Sign Languages*. Multilingual Matters.

KUSTERS *et al.* (2017). *Innovations in Deaf Studies: The Role of Deaf Scholars*. OUP.

FABIO ARDOLINO<sup>1</sup>

<sup>1</sup>Università di Pisa

## **Adattare i protocolli IAT allo studio del parlato: criticità e proposte metodologiche**

La collezione di informazioni circa l'esistenza, l'entità e le tendenze degli atteggiamenti psicologici (espliciti ed impliciti) costituisce una forma di documentazione centrale nel metodo di numerose discipline (dalla psicologia alla sociologia, dalla linguistica alla psichiatria). Nella relativa letteratura, si circoscrive, con il termine "atteggiamento", la predisposizione comportamentale di un individuo rispetto ad un topic qualsiasi (Allport, 1967); una predisposizione che si motiva, in prima istanza, a partire dal complesso di valori, credenze ed esperienze costituito dal soggetto nel corso della sua esistenza.

Lo studio degli atteggiamenti è cruciale poiché è a partire da tali impianti che il singolo attua e regola il proprio comportamento, tanto all'interno della propria società quanto, eventualmente, al suo esterno (rispetto cioè a collettività e ad individui altri rispetto a quelli che gli sono familiari). Tuttavia, la raccolta di informazioni in proposito si scontra spesso con difficoltà consistenti dovute, soprattutto, a due aspetti: da un lato, la naturale ritrosia dei soggetti a rivelare apertamente propri impianti giudiziari (e pregiudiziali) e, dall'altro, il fatto che tali impianti non affiorano, il più delle volte, al livello cosciente, restando sottesi a quei comportamenti impliciti, rapidi ed automatici (vedi Greenwald & Banaji, 1995) che acquisiscono il maggior peso nell'attuazione di una condotta. Si tratta di nodi problematici che gravano in maniera sostanziale sull'applicabilità e sulla validità di quei protocolli sperimentali diretti (su tutti, la scala Likert, cfr. Likert, 1939) basati sulla sottomissione di questionari espliciti, costituiti cioè da domande dirette circa l'esistenza di giudizi di apprezzamento/disprezzo verso l'oggetto di studio.

Lo sviluppo e l'applicazione sul campo dei protocolli sperimentali definiti IAT (Implicit Association Tests, cfr. Greenwald *et al.*, 1998) ha impresso un deciso sviluppo alla metodologia d'indagine delle scienze sociali e psicologicosociali, consentendo il superamento di quelle tare sperimentali sopra esposte. Un test IAT, infatti, si pone lo scopo di registrare l'atteggiamento implicito di un soggetto elicitando nello stesso una forma di comportamento automatica, non cosciente o comunque non razionalmente elaborata. Protocolli IAT, o da essi derivati, sono stati applicati con risultati positivi in studi sull'autostima (Greenwald & Farnham, 2000), sull'ansia sociale (Egloff & Schmukle, 2002), sul pregiudizio di genere (Fernandez *et al.*, 2014), sul pregiudizio

razziale nell'infanzia (Baron & Banaji, 2006) sulla depressione e le tendenze suicide (Nock *et al.*, 2010) e sulle tendenze sessuali (Anselmi *et al.*, 2013).

L'utilità rivelata dallo IAT nella collezione di dati sugli impianti pregiudiziali presenti fra sottogruppi sociali ha motivato l'interesse con il quale tale protocollo è stato accolto nell'alveo degli studi sociolinguistici (cfr. Campbell-Kibler, 2012), con particolare riguardo all'indagine del pregiudizio nei confronti degli accenti stranieri (Pantos & Perkins, 2012; Álvarez-Mosquera & Marín-Gutiérrez, 2017), delle varietà (Rosseel *et al.*, 2017) e dell'innovazione linguistica (McKenzie & Carrie, 2018).

Il contributo qui presentato si pone un duplice scopo: da un lato, valutare l'applicabilità dei protocolli IAT allo studio dei principali aspetti socio- e psicolinguistici della lingua parlata, valutandone vantaggi e criticità; dall'altro, proporre e discutere, alla luce delle esperienze sul campo, modalità sperimentali e cautele adeguate alla produzione di risultati utili alla ricerca. Fra questi temi si pongono in evidenza:

1. Con quali modalità produrre lo stimolo linguistico, sia sfruttando strumenti di sintesi del parlato sia reclutando parlanti e, in tal caso, a quali caratteristiche porre attenzione al fine di evitare interferenze sperimentali (elicitazione di atteggiamenti non pertinenti prodotti dalle caratteristiche vocali del parlante).
2. Come costituire lo stimolo linguistico, contrapponendo l'uso di nonparole all'uso di parole significanti e, in questo caso, come evitare effetti di sovrapposizione emotive bilanciando lo stimolo in termini di composizione sillabica e caratteristiche segmentali.
3. Come sfruttare una o più fasi di pre-testing per affinare la qualità dello stimolo.
4. Come ovviare alla criticità derivante dall'impiego di stimoli estesi nel tempo in una modalità di test nel quale parte del risultato è computato sul tempo di reazione del soggetto. Anche alla luce della letteratura precedente, si discuteranno i limiti temporali e le caratteristiche in grado di garantire una buona riuscita del protocollo.
5. Quali cautele usare nel coinvolgimento dell'informante, nel trattamento del dato raccolto e nella divulgazione dei risultati.

Gli spunti discussi verranno inoltre contestualizzati, mostrando la fase di sviluppo e, coerentemente con la prosecuzione dei relativi progetti, di applicazione, di due protocolli IAT elaborati dal proponente e finalizzati a produrre dati per due studi, aventi come oggetto:

1. L'atteggiamento dei professori di scuola superiore italiana nei confronti delle parlate italiane con accento straniero, e cinese in particolare, in istituti toscani; lo IAT ha in questo caso rilevato a presenza di una forma di lieve impianto pregiudiziali nei confronti del parlato con accento cinese.
2. Il sentimento di integrazione/alienazione sociale di una comunità immigrata, che si assume potenzialmente correlato all'esistenza di un

atteggiamento implicito positivo/negativo nei confronti della lingua veicolare in uso presso la comunità ospitante.

### *Riferimenti bibliografici*

- ALLPORT, G.W. (1967). Attitudes. In FISHBEIN, M. (Ed.). *Readings in Attitude Theory and Measurement*. New York: Academic Press, 8.
- ÁLVAREZ-MOSQUERA, P. & MARÍN-GUTIÉRREZ, A. (2017). Implicit language attitudes toward historically white accents in the South African context. In *Journal of Language and Social Psychology*, 37(2), 238-248.
- ANSEMI, 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(11).
- BARGH, J.D. & CHARTRAND, T. (2000). The mind in the middle: a practical guide to priming and automaticity research. In REIS, H.T. & JUDD, C.M. (Eds.). *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press, 253-285.
- 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(1), 53-58.
- CAMPBELL-KIBLER, K. (2012). The Implicit Association Test and sociolinguistic meaning. In *Lingua*, 122(7), 753-763.
- EGLOFF, B. & SCHMUKLE, S.C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. In *Journal of Personality and Social Psychology*, 83(6), 1441-1455.
- 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(6), 1464-1480.
- GREENWALD, A.G. & BANAJI, M.R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. In *Psychology Review*, 2(1), 4-27.
- GREENWALD, A.G. & FARNHAM, S.D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. In *Journal of Personality and Social Psychology*, 79(6), 1022-1038.
- LIKERT, R. (1932). Technique for the measure of attitudes. In *Archives of Scientific Psychology*, 22(17), 5-55.
- MCKENZIE, R.M. & CARRIE, E. (forthcoming). Implicit-explicit attitudinal discrepancy and the investigation of language attitude change in progress (early view). In *Journal of Multilingual and Multicultural Development*.
- 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(4), 511-517.
- PANTOS, A.J. & PERKINS, A.W. (2012). Measuring implicit and explicit attitudes toward foreign accented speech. In *Journal of Language and Social Psychology*, 32(1), 27-32.

ROSSEEL, L., GEERAERTS, D. & SPEELMAN, D. (2017). Measuring language attitudes in context: exploring the possibilities of the PIAT, a case study on regional varieties of Dutch in Belgium. In *Methods in Dialectology XVI* (abstract).

MARIA FRANCESCA STAMULI<sup>1</sup>

<sup>1</sup>Soprintendenza Archivistica e Bibliografica della Toscana

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

Le *fonti orali* sono considerate fonti primarie per tutte quelle discipline umanistiche che si fondano sul metodo della 'ricerca sul campo'. Sin dalla fine del secolo XIX, la linguistica, l'antropologia culturale, la storia - e non soltanto quella 'orale' - e la storiografia hanno rilevato precocemente, nell'ambito di una riflessione metodologica non di rado conflittuale, gli elementi critici di queste fonti, rispetto tanto alla loro definizione quanto alla loro costituzione, reperimento e uso.

L'approccio archivistico a questi documenti permette di avere, con un posizionamento sovradisciplinare, una visione utile ad affrontarne le specificità. L'archivistica, infatti, è a sua volta portatrice di una propria metodologia, dettata da una finalità chiara quanto specifica: la realizzazione di archivi intesi come l'insieme dei documenti redatti (e ricevuti) da una persona fisica o giuridica come strumento e residuo della propria attività professionale, creativa o istituzionale e conservati per 'riferimento' proprio o pubblico. Il punto di vista dell'archivista, in particolare, rileva tre aspetti:

a. le *fonti orali* sono di definizione controversa. Con Contini si potrebbe dire che "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). D'altra parte, fino all'invenzione e alla diffusione dell'uso di apparecchiature di registrazione della voce e dell'immagine nelle ricerche 'etnologiche' e 'linguistiche', per 'fonti orali' si intendevano le 'testimonianze orali' fornite da vive voci e 'fissate' (o 'nuovamente mediate' nelle forme della scrittura).

b. le *fonti orali* sono difficilmente determinabili. La lunga conformazione in scrittura assunta dalle *fonti orali* nei diversi ambiti disciplinari che in Italia ne facevano uso ha probabilmente contribuito a determinare la difficoltà del riconoscimento del 'documento - testimonianza' come documento informativamente autonomo. Poiché le *fonti orali* si presentano molto spesso come un documento prodotto nell'ambito di uno specifico percorso di ricerca, esse, non di rado - e soprattutto in passato, sono state percepite come uno strumento di lavoro ad uso del solo ricercatore, come documenti privati o non pubblicizzabili. Dal punto di vista archivistico, tali documenti dipendono intrinsecamente dal più ampio percorso di ricerca, ne sono, appunto, 'strumento e residuo', andando a costituire l'*archivio* di quella specifica attività di

ricerca: nell'ambito dell'insieme archivistico, tali documenti si stratificano organicamente e 'vivono' come fonti informative autonome.

a. le *fonti orali* sono molto spesso documenti ad alto rischio di conservazione. Cilindri fonografici, dischi, nastri magnetici, CD o nastri a incisione digitale: a tutti questi supporti, propri di documenti sonori e cinematografici, è riconosciuta una notevole deperibilità, legata all'uso stesso, alle difficoltà di garantire adeguati ambienti di conservazione, all'obsolescenza delle macchine utili a riprodurre e 'leggere' tali documenti. Si aggiunga che i costi per campagne di riversaggio su supporti e forme attuali (in questo momento, *files* in formati standardizzati con conservazione in server) sono altissimi, sicuramente non sostenibili per la maggior parte degli istituti di conservazione di documenti sonori e cinematografici in Italia. Un altro importante elemento di rischio per la conservazione delle *fonti orali* non è legato alla deperibilità dei supporti ma all'aspetto che si è definito 'archivistico - strumentale' dei documenti che veicolano tali fonti: la difficoltà di riconoscere al documento legato all'oralità uno 'statuto autonomo' rispetto agli obiettivi del più ampio percorso di ricerca nel quale l'elicitazione della fonte si colloca non di rado ne ha determinato la dispersione, se non, in alcuni casi, la distruzione. Non raramente, in passato, alcuni ricercatori, visti i costi dei supporti, hanno 'cancellato' le registrazioni conservandone solo la trascrizione: esemplare è il caso dell'Oral History Research della Columbia University, che ha conservato solo le trascrizioni delle fonti sonore, distrutte.

Gli aspetti fin qui rilevati, nelle loro diverse articolazioni, convergono verso due direzioni:

1. un'ulteriore riflessione terminologica sulla definizione di *fonte orale*, con un percorso di analisi che scandagli la distinzione tra 'informazione' e 'oggetto informativo' e che individui la *fonte orale* non nel suo *medium* attuale ma nella modalità di rilievo e produzione 'primaria' che ha caratterizzato la raccolta di quel contenuto informativo. Solo quando questo dualismo sarà ben definito sarà possibile fornire una efficace definizione di *fonte orale*, con una sua proficua e necessaria declinazione tipologica (documento scritto-illustrato, documento sonoro, documento cinetico o filmico, e simili).
2. costituire tali fonti in *archivi*, ovvero in luoghi in cui i documenti informativi siano sempre contestualizzati rispetto al loro farsi e siano descritti in modo standardizzato per un accesso efficace.

La necessità di avere 'note' di contesto circa l'elicitazione delle fonti orali è sottolineata concordemente dagli antropologi, dai linguisti e dagli storici. Il contesto, d'altra parte, è uno dei cardini intorno ai quali si è formata epistemologicamente la disciplina archivistica: il 'vincolo archivistico', il legame che mette in relazione i documenti tra loro e con la prassi che organicamente li produce, è quel fattore indispensabile a determinare la possibilità che singoli oggetti e segmenti informativi siano individuabili e accessibili. Con l'invenzione della rete e, soprattutto, con l'ulteriore sviluppo delle sue funzionalità nel modello

del web semantico, il contesto ha assunto ancora maggiore rilievo: al 2016 risale la proposta dell'*International Council of Archives* di far confluire i diversi standard di descrizione delle entità archivistiche in un unico modello di registrazione dei contesti di riferimento (si veda il modello di *Record in Contexts*). La creazione di archivi organizzati in *repositories* standardizzati di sistemi informativi assume, con questo approccio, un più immediato ed evidente interesse interdisciplinare: l'esistenza di archivi di *fonti orali* si pone non solo come una necessità per la valorizzazione e la conservazione del patrimonio immateriale, in questo comprendendo anche il patrimonio linguistico, ma anche come una sfida metodologicamente significativa.

La progettazione di Archivio Vi.Vo, archivio delle fonti orali della Toscana, in cui i documenti di testo si integrino a quelli orali e cinematografici nel contesto dei soggetti e dei sistemi di ricerca che li hanno prodotti (progettazione che è nata dalla collaborazione tra Soprintendenza Archivistica della Toscana, Università di Siena e soggetti portatori di 'fonti orali' disseminati sul territorio) vuole andare in questa direzione.

#### *Riferimenti bibliografici*

INTERNATIONAL COUNCIL ON ARCHIVES. EXPERTS GROUP ON ARCHIVAL DESCRIPTION (2016). *Record in Contexts. A conceptual model for Archival Description. Consultation Draft v.01*. Consultato da <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf> il 24.09.2018

ANDREINI, A. & CLEMENTE, P. (Eds.) (2007). *I custodi delle voci. Archivi orali in Toscana: primo censimento*. Toscana Beni culturali, vol 8. 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.

CARUCCI, P. & CONTINI, G. (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.

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.it*, 4(1), 213-226. DOI: <http://dx.doi.org/10.4403/jlis.it-5474> )

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*, 795-820. Roma: Ministero per i beni e le attività culturali. Dipartimento per i beni archivistici e librari. Direzione Generale per gli Archivi.

DURANTI, L. (2015). Il documento archivistico. In GIUVA, L. & GUERCIO, M. (Eds.) *Archivistica. Teorie, metodi e pratiche*, 19-34. Roma: Carocci editore.

- IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. (2009). *Functional requirements for bibliographic records: final report*. Consultato da [https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf) il 24.09.2018.
- LEYDI, R. & TELLMON, 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)*, 17-34. Roma: Ministero per i beni e le attività culturali.
- MATRANGA, V. (2002). Come si fa un'indagine dialettale sul campo. In CORTELAZZO, M., MARCATO, C. & DE BLASI, N. *I Dialetti Italiani*, 64-80. Torino: UTET.
- MC LUHAN, M. (1967). *Gli strumenti del comunicare*. Milano: Il Saggiatore.
- MAGNANI, F. (2018) Le documentazioni audiovisive nelle schede di catalogo per i beni culturali demoetnoantropologici. In TUCCI, R. (2018). *Le voci, le opere e le cose. La catalogazione dei beni culturali demoetnoantropologici*, 262-274. Roma: Ministero dei beni e delle attività culturali e del turismo, Istituto centrale per il catalogo e la documentazione.
- ONG, W. J. (1986). *Oralità e scrittura. Le tecnologie della parola*. Bologna: Il Mulino.
- WIKIPEDIA CONTRIBUTORS. (2018, July 21). The Human Library. In *Wikipedia, The Free Encyclopedia*. Consultato da [https://en.wikipedia.org/w/index.php?title=The\\_Human\\_Library&oldid=851365517](https://en.wikipedia.org/w/index.php?title=The_Human_Library&oldid=851365517) il 24.09.2018.

PHILIPPE MARTIN<sup>1 2 3</sup>

<sup>1</sup>LLF, <sup>2</sup>UFRL, <sup>3</sup>Université Paris Diderot

## Prosodic annotation of oral archives (1913-2018)

Oral archives found in 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) make intonation analysis difficult and unreliable for most pitch tracking algorithms, such as the ones found in the popular software program Praat. Specific 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.

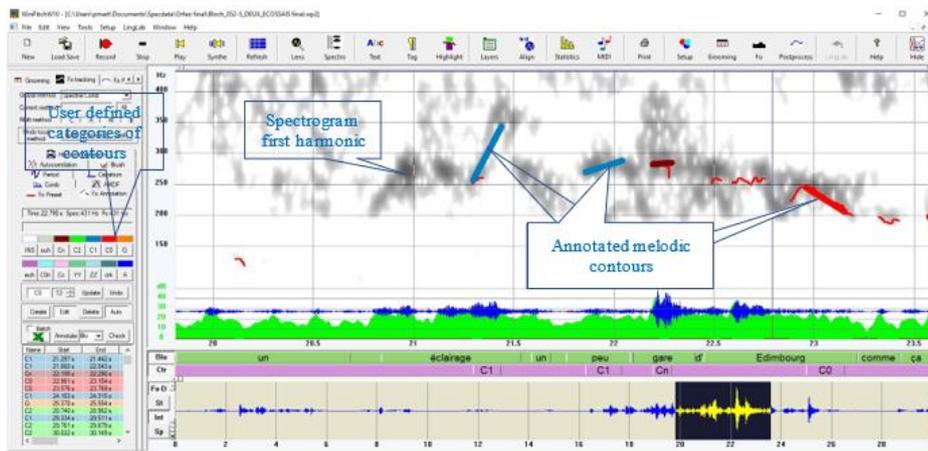
To address these limitations and pursue prosodic research on valuable speech archives, a set of dedicated functions have been implemented in the software program WinPitch. These functions are based on the visual validation of the pitch curve against the simultaneous display of a frequency aligned narrow-band spectrogram first (or second) harmonic. Using one of the seven tracking algorithms available in WinPitch (autocorrelation, AMDF, spectral comb, spectral brush, Cepstrum...), the user is not only able to visually check the validity of a given pitch track section against the corresponding spectrogram first harmonic segment, but also to select graphically the F0 curve section of interest and enter its characteristics by graphic commands.

These user-friendly commands using only mouse controls allow to position graphic piecewise linear segments approximating any shape of the fundamental frequency curve. Furthermore, 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 (Fig. 1). The 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 C0 (terminal falling and low for declarative sentences), C1 (rising above the glissando threshold), C2 (falling above the glissando threshold) and Cn (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. Using ToBI notation such as

H\*H%, L\*L-, , L+H\* for FTtoBI, it is equally easy to predefine as actual acoustic melodic movements are then approximated in terms of tone targets. Labelling can then be done manually or automatically according to the user criteria. The integrated system has been applied to the analysis of a large number of speech archives, both old and recent. Examples pertaining to the changes occurred in the realization of melodic contours in French in the period 1913-2018 will be given, 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.

Fig. 1. 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.



### Riferimenti bibliografici

MARTIN, Ph. (2018) *Intonation, structure prosodique et ondes cérébrales*, London: ISTE, 332.

WINPITCH: [www.winpitch.com](http://www.winpitch.com)

ALESSANDRO RUSSO<sup>1</sup>, NICCOLO' PRETTO<sup>1</sup>, ANTONIO RODÀ<sup>1</sup>, SERGIO CANAZZA<sup>1</sup>

<sup>1</sup>Università degli Studi di Padova

## **Information engineering methodologies for managing and preserving audio (speech and music) archives**

The lack of a standard preservation methodology is an important issue that digital archives have to face nowadays. In fact, the huge amount of information that needs to be stored, data and metadata, requires the implementation of a methodological framework, in order to prevent significant cultural losses. Active preservation and restoration of audio documents is one of the main research topics at Centro di Sonologia Computazionale (CSC) of the Università di Padova, Department of

Information Engineering (DEI). The laboratory, during its years of activity, cooperated with many local and international archives, dealing in particular with audio documents, more often magnetic tapes, containing speech recordings, electro/acoustic music and linguistic and ethnographical studies. In this scenario, a multidisciplinary approach is fundamental, due to the complex nature of audio documents. As a matter of fact, even if the stored audio information is the main center of importance, also the carrier has to be preserved and restored, with procedures that require skills in the field of information engineering, material science, chemistry and musicology. This is particularly important, for example, dealing with electronic music, where the carrier is a part of the artwork itself, due to signs and annotations that the composers usually took onto the tape. The research activity led to the development of a scientific methodology that aims to answer the requirements of audio archives in the modern scenario, that in many cases still do not dispose of adequate instruments for managing a complex situation as a digitized audio corpus is. In fact, it's evident that common issues, like chemico-physical and digital degradation, obsolescence of technologies (carriers and playback systems) and digital formats can't be faced with a classic "librarian" approach by the archivist, but require specific protocols. In addition, regarding the data access, it is important to remind that stored data have to be reached by the user in every moment and from everywhere. New technologies, like the web-based ones, may be helpful for achieving the abovementioned purpose and have to be considered by modern archives. Nevertheless, due to the incremental progress in the field of artificial intelligence, like machine learning

techniques, the framework is still a work in progress at the CSC and the methodology is always improved with new additions, both in the preservation and in the access procedures. Major improvements derived from machine learning-based techniques, like automatic systems for detecting the right equalization curve to apply to a tape during the signal transfer and systems for the automatic recognition of discontinuities on the tape (splices, lacks of magnetic paste, presence of signature etc.) are just some of the last results of the research activity of the laboratory that may be really helpful in the modern scenario. In fact, the precious information extracted thanks to the use of these techniques may be helpful to prevent common human mistakes and data loss, leading to a huge savings in time and expense for the archive. In the paper the authors will introduce the developed methodology, explaining it in all its steps, focusing both on the preservation and the access procedures, with a particular emphasis on the concept of metadata and its primary importance in the storing procedure, the philological accuracy of the process, some machine learning-based techniques applied for the recognition of equalization curves and discontinuity on the tape and the benefits that audio restoration algorithms can bring to digitized oral archives, in particular for improving the intelligibility of speech recordings, even providing some examples.

## Le consonanti occlusive sorde aspirate in Calabria: un confronto tra aree dialettali

L'obiettivo di questo studio è oggettivare acusticamente la presenza di post-aspirazione nelle consonanti occlusive sorde in contesto post-nasale per alcuni dialetti calabresi appartenenti alla terza, quarta e quinta area dialettale della regione, sulla base della partizione proposta da Trumper, Maddalon (1988) e Trumper (2016). Il presente lavoro si inserisce in una più ampia ricerca che mira alla mappatura del fenomeno in esame nell'intero contesto calabrese.

L'aspirazione delle occlusive sorde in Calabria è un fenomeno che si manifesta in modo irregolare sul territorio. Questo fenomeno 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 quando geminate o a seguito di /N/ o /r/, mentre il fenomeno risulta meno frequente, in tutte le posizioni accentuali, se l'occlusiva sorda è preceduta da /l/.

Relativamente ai contesti post-nasali, soggetti ad esiti alquanto eterogenei nelle diverse aree esaminate dagli autori sopra menzionati, i riscontri acustici finora attestati riguardano esclusivamente: i) l'italiano regionale e il dialetto di parlanti adulti dell'area urbana di Cosenza (Sorianello, 1996) e del paese di San Giovanni in Fiore (Mele & Schmid, 2009; Mele, 2009; Romito, Ciardullo & Tarasi, 2015); ii) la varietà regionale di adolescenti e anziani dell'area catanzarese, rispettivamente indagati da Nodari (2015) e nello studio pilota di Frontera (2018); iii) il parlato spontaneo (dal corpus HLVC di Toronto) di calabresi emigrati, analizzato da Nagy, Nodari, Celata (2018) e Nodari, Celata e Nagy (*forthcoming*). 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 (area due) tendono all'utilizzo delle rispettive omorganiche sonore, per effetto di processi fonologici di sonorizzazione (Trumper & Ortale, 1978; Sorianello, 1996): seppure con una forte variabilità intra-parlatore, nel dialetto sangiovanese si attestano realizzazioni *long lag* (Lisker & Abramson, 1964) fino a 75 ms per il nesso -nk- in posizione post-tonica interna di parossitoni e valori medi di

VOT uguali a 57 ms per lo stesso nesso in posizione pretonica iniziale di parossitone (Romito, Ciardullo & Tarasi, 2015). Tali valori accomunano tutte le fasce di età considerate ed entrambi i sessi. Risulta interessante, perciò, per questa ricerca, verificare se e quanto differiscano, 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. Nel presente lavoro si concentrerà l'attenzione sui contesti fonetici post-nasali -nk, -nt, -mp. Le domande a cui si cercherà di rispondere sono, dunque, le seguenti:

- 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 dialettologiche?
- Tali valori si uniformano ai *trend*, riscontrati in letteratura, inerenti alle variazioni fra contesti fonetici e accentuali di occorrenza dei suoni occlusivi sordi?
- Esiste un'area, fra quelle esaminate, in cui i risultati riportati siano significativamente divergenti rispetto alle altre aree?
- Esistono differenze significative nei livelli di aspirazione, laddove presente, prodotta da parlanti più e meno giovani?

Si procederà, pertanto, all'analisi dei valori di VOT prodotti nelle consonanti occlusive sorde /p t k/ in contesto post-nasale di parlato dialettale, bilanciando il numero di occorrenze sulla base delle consonanti analizzate, del timbro delle vocali ad esse adiacenti (/a i u/) e del contesto accentuale in cui compaiono (posizione pretonica/postonica, rispettivamente in parossitono e proparossitono). I dati verranno elicitati tramite *task* consistenti nella lettura di 36 parole target inserite in contesto frasale, da tradurre nella varietà dialettale di appartenenza. Il campione comprenderà un minimo di 36 soggetti, di sesso maschile e femminile (18+18) distribuiti su 3 fasce di età, in un *range* compreso fra 18 e +65 anni e provenienti dalle aree dialettologiche tre, quattro e cinque, ciascuna delle quali sarà rappresentata da due paesi. I valori di VOT saranno misurati a partire dal primo *burst* prodotto nella fase di rilascio, fino all'*onset* periodico della vocale successiva (Harrington, 2013). Tutti i suoni target saranno etichettati e annotati foneticamente. I valori dei singoli VOT saranno estratti automaticamente e verificati manualmente tramite il software Praat, come somma della fase di rilascio (REL) e aspirazione (ASP) di ciascuna consonante, seguendo la procedura adottata da Abramson e Whalen (2017). Si effettuerà il calcolo delle medie totali ottenute da ciascun gruppo, esaminate sulla base di tutte le variabili fonetiche e sociolinguistiche coinvolte (qualità consonantica, vocali adiacenti, contesti accentuali, differenze tra fasce d'età, sessi e aree

dialettologiche). Inoltre, i dati verranno trattati statisticamente al fine di individuare differenze significative interne ed esterne ai gruppi, nonché eventuali combinazioni di variabili più favorevoli a produrre aspirazione.

I primi risultati confermano i dati riscontrati in letteratura riguardo alla variazione del VOT in relazione ai contesti fonetici di occorrenza, ovvero durate maggiori nelle occlusive arretrate rispetto alle più anteriori, nelle vocali alte rispetto a quelle basse e in contesto postonico. Il picco maggiore sembra essere riscontrato nei parlanti più anziani di area tre e quattro, i quali producono valori di aspirazione mediamente maggiori o uguali a 60 ms di durata (*long lag*). Nei dati analizzati fino a questo momento, i valori di VOT ottenuti non solo sono mediamente superiori rispetto a quelli presenti nelle aree dialettologiche indagate in passato, ma quelli di parlanti anziani superano notevolmente anche le durate ottenute nelle fasce d'età dei più giovani.

Ad oggi, in ultimo, non esiste una spiegazione univoca che dia contezza della distribuzione eterogenea dei suoni in esame sul territorio calabrese, nonostante siano state avanzate nel tempo alcune ipotesi (cfr. ad esempio Falcone, 1977; Piccardi, 2018): i risultati raggiunti potranno fornire nuovi elementi di raffronto nella riflessione di natura linguistico-storica tuttora aperta.

### *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.
- CANEPARI, L. (1986). *Italiano standard e pronunce regionali*, CLEUP: Padova.
- FALCONE, G. (1976). Calabria. In CORTELAZZO, M. (Ed.). *Profilo dei dialetti italiani*, Pacini: Pisa.
- 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*, 28-30 August, Paris, France, 33-36.
- HARRINGTON, J. (2013). Acoustic Phonetics. In HARDCASTLE, W. J., LAVER, J. & GIBBON, F. E. (Eds.). *The Handbook of Phonetic Sciences*, (2nd ed.). Blackwell Publishing: Chichester, 81-129.
- LISKER, L. & ABRAMSON, A.S., (1964). A cross-language study of voicing in initial stops: Acoustical measurements. In *Word*, 20(3), 527-565.
- 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.
- MELE, B. (2009). *Fonetica e fonologia del dialetto di San Giovanni in Fiore*. Tübingen-Basel: Francke Verlag.
- NAGY, N., NODARI, R. & CELATA, C. (2018). A variationist analysis of Heritage Italian VOT: Phonetic but not phonological fidelity, *NWAV 47*, NY, NY, 18-21 Oct. 2018.

- 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., CELATA, C. & NAGY, N. (*forthcoming*). Socio-indexical phonetic features in the heritage language context: VOT in the Calabrian community in Toronto. In *Journal of Phonetics*.
- PICCARDI, D. (2018). L'aspirazione nei dialetti italiani meridionali e greci moderni. Tracce di una strategia di preservazione?. In *Lingua e stile*, LIII (1), 27-55.
- ROHLFS, G. (1966). *Grammatica storica della lingua italiana e i suoi dialetti (I)*, 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.). *Il 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.
- SORIANELLO, P. (1996). Indici fonetici delle occlusive sorde nel cosentino. In *Rivista Italiana di Dialettologia*, 20, 123-159.
- TRUMPER, J. & MADDALON, M. (1988). Converging divergence and diverging convergence: the dialect-language conflict and contrasting evolutionary trends in modern Italy. In *Variation and convergence*. Berlin-New York: W. De Gruyter, 217-259.
- TRUMPER, J. (2016), *Geostoria linguistica della Calabria*, Roma: Aracne editore.
- TRUMPER, J. & ORTALE, R. (1978), Analisi preliminare del sistema consonantico del dialetto di Cosenza. In *Lingua e contesto*, 5, 382.

VALENTINA COLONNA<sup>1</sup> <sup>2</sup>, ANTONIO ROMANO<sup>2</sup>

<sup>1</sup>Università degli Studi di Genova, <sup>2</sup>Università degli Studi di Torino

## VIP: un archivio per le voci della poesia italiana

La lettura della poesia costituisce uno degli aspetti più trascurati e interessanti per osservare la dimensione sonora della poesia. A tal fine la creazione di archivi vocali che raccolgano registrazioni di letture poetiche rappresenta non solo una realtà recente e che nel panorama internazionale è andata affermandosi in forme differenti, ma anche un elemento essenziale per un tipo di studio che, in questo specifico ambito, vede ancora uno sviluppo limitato. Se infatti importanti database di letture originali di autori contemporanei (in lingua originale e in traduzione) di numerosi Paesi e in numerose lingue costituiscono progetti internazionali imponenti, con un ventaglio molto largo di dati raccolti (si pensi ad esempio all'archivio Lyrikline, forse il più rappresentativo e completo in questa direzione), appare importante, dall'altra parte, lo sviluppo di archivi nazionali vocali più mirati, che raccolgano registrazioni della letteratura poetica nazionale e possano avere, oltre a un fine documentale, l'obiettivo di rappresentare materiale di ricerca scientifica. Su questo fronte, alcune realtà geografiche sono da più tempo sensibili alla questione e si sono dedicate alla raccolta di materiale nel corso di questi anni: possiamo pensare all'area ispanofona (con i rappresentativi *A media voz*, *Voces que dejan huellas* - Cecilia.com.mx, Biblioteca nacional digital de Chile, *The Booksmovie*) e all'area anglofona, dove molti sono gli archivi e tra questi è particolarmente ampio, ad esempio, il database della Library of Congress di 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.<sup>1</sup> Offre un ampio archivio di materiale anche il generale The Internet Archive, nella sua sezione specifica *Audio Books and Poetry*. A questi si possono aggiungere nelle varie realtà nazionali gli archivi delle rispettive radio-televisioni di Stato e private, insieme ai numerosi centri studi dedicati ai singoli autori. Nel panorama degli archivi italiani sarebbe meritevole

---

<sup>1</sup> Citiamo alcune tra le più significative piattaforme che raccolgono registrazioni di poesia in lingua inglese: Archive of the Now, PennSound, UbuWeb, Woodberry Poetry Room, The Poetry Archive, Raleigh Review, Academy of American Poets, The Cortland Review, Essential American Poets, From the Fishhouse, The Internet Poetry 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. Si considerino in questo elenco cataloghi di letture singole degli autori o interpretate da attori e ascoltabili online.

di menzione il progetto Phonodia dell'Università degli Studi di Venezia (Dipartimento di Filosofia e Beni Culturali).

Globalmente gli archivi citati rappresentano banche di dati, frutto di digitalizzazioni o dati raccolti appositamente, al punto da costituire vere e proprie audiotecche, senza specifiche finalità di studio fonetico. Progetto di rilievo nel panorama europeo degli studi afferenti alla lettura del testo è 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*.

All'interno di una situazione internazionale e nazionale che vede dunque gli studi fonetici rivolti alla lettura della poesia e alla sua questione ritmico-melodica alquanto limitati (fatti salvi i contributi pionieristici di Bertinetto, 1973, 1978), è emersa la necessità di dedicare uno studio mirato, che prenda in analisi un archivio rappresentativo del panorama poetico italiano contemporaneo. Se anche la Critica e la Linguistica italiane hanno spesso messo in secondo piano indagini sulla "musica" del testo poetico, per quanto autorevoli lavori di tipo metrico e stilistico siano stati condotti in una tradizione di studi a livello endofasico, fuori dall'Italia in diversi si sono cimentati nella ricerca applicata al campo dell'intonazione del verso poetico e della fonologia metrica. I numerosi contributi italiani in ambito prosodico si sono invece concentrati principalmente sull'intonazione del parlato e isolati sono i lavori di tipo fonetico concentrati sulla poesia (si pensi a Schirru, 2004).

A tal fine, *Voices of Italian Poets* vuole essere una piattaforma che accoglie 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 esplicita nella piattaforma stessa. L'archivio si propone di diventare una significativa banca-dati di letture poetiche, curate da poeti e professionisti della voce, e un riferimento per lo studio di tipo fonetico della poesia. Più nello specifico, *VIP* prende forma a partire dalle letture originali di autori del Novecento, reperite in archivi di vario tipo e in materiale inedito o fuori dai diritti commerciali e spazia poi nelle letture comparative dei medesimi testi e nelle letture inedite di poeti viventi che leggono proprie poesie. È possibile in questo progetto una doppia chiave interpretativa e di osservazione, che si sviluppa dunque in orizzontale, su un piano cronologico che permette di avere una panoramica dettagliata di vari stili di lettura nel tempo, e un piano verticale trasversale e comparativo, che consente il confronto tra diverse interpretazioni di un medesimo testo del Novecento.

Studi fonetici, dettagliati e comparativi, secondo la doppia linea che guida la creazione e la costituzione dell'archivio, possono risultare particolarmente utili per osservare aspetti non ancora esplorati scientificamente e possono ambire al tentativo di continuare, in una modalità differente e complementare, gli studi metrici finora svolti. In particolare, il caso del verso libero, centrale e prevalente

nelle composizioni contemporanee e presenti nelle letture del database, costituisce anche il fulcro delle osservazioni fonetiche, in cui la materia metrica si scontra con difficoltà di analisi e definizione. La ricerca della questione ritmica e melodica del testo per mezzo dell'osservazione e dell'analisi di dati acustici, ora reperibili e assenti invece nella documentazione di tutta una tradizione letteraria, è in grado di fornire informazioni utili a livello interpretativo, non individuabili a una sola lettura "silenziosa". Per contro, non sufficiente sarebbe una mera osservazione acustica del dato sonoro, che non tenga conto del testo scritto e delle variabili che esistono nelle differenti interpretazioni di lettori o anche solo di momenti diversi, così come avviene in musica nel rapporto composizione-interprete. Dunque, obiettivo di ricerca, che continua e amplia il primo lavoro condotto nella tesi magistrale di V. Colonna, dedicata a dodici letture del Congedo del viaggiatore cerimonioso di Giorgio Caproni, non è quello di dare una norma o indicare la più consona lettura adatta alla poesia, bensì quello di individuare i punti salienti che accomunano e distinguono varie interpretazioni, senza entrare mai nel limite sottile dell'indicibile dell'arte. Fine ultimo è quello di osservare la realizzazione concreta del ritmo poetico nel verso libero e analizzare il rapporto tra "sovra-testo" prosodico della lettura ad alta voce e testo scritto, utilizzando un approccio di tipo multidisciplinare.

La piattaforma Voices of Italian Poets è ospitata sul sito del Laboratorio di Fonetica Sperimentale "Arturo Genre" dell'Università di Torino e rappresenta un importante supporto di consultazione gratuita e accessibile a tutti, oltre che prestarsi a eventuali utilizzi futuri in campo didattico: [http://www.lfsag.unito.it/ricerca/VIP\\_index.html](http://www.lfsag.unito.it/ricerca/VIP_index.html).

### *Riferimenti bibliografici*

- ALBANO LEONI, F. (2009). *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: Il Mulino.
- 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. (1973). *Ritmo e modelli ritmici: analisi computazionale delle funzioni periodiche nella versificazione dantesca*. Torino: Rosenberg & Sellier.
- BERTINETTO, P. M. (1978). Strutture soprasedimentali e sistema metrico. In *Metrica*, I, 1-54.
- COHEN, J. (1966). *Structures du langage poétique*. Paris: Flammarion.
- FÓNAGY, I. (1982). *La ripetizione creativa. Ridondanze espressive nell'opera poetica*. Bari: Dedalo, 10-14.
- GRAMMONT, M. (1946). *La prononciation française. Traité pratique*. Paris: Delagrave.
- MORTARA GARAVELLI, B. (2003). *Prontuario di punteggiatura*. Roma-Bari: Laterza.

- PAMIES BERTRAN, 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.
- PHONODIA (UniVe) <http://phonodia.unive.it/> (ultima consultazione 24/09/2018).
- ROMANO, A. (2004). Indices acoustiques suprasegmentaux dans la caractérisation des langues romanes: identification de variétés linguistiques et description des traits prototypiques. In *Identification des langues et des variétés dialectales par les humains et par les machines*, Atti del convegno MIDL (Parigi, Francia, 29-30 nov. 2004). Paris: École Nationale Supérieure des Télécommunications, 91-92.
- ROMANO, A. (2008). *Inventari sonori delle lingue. Elementi descrittivi di sistemi e processi di variazione segmentali e sovrasegmentali*. Alessandria: Edizioni Dell'Orso.
- 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 AISV 2009, Università di Zurigo, Kollegengebäude, (4-6 Febbraio 2009). Torriana (RN): EDK, 45-75.
- SCARPA, R. (2004). *Sintassi e respiro nei sonetti di Giorgio Caproni*. Alessandria: Edizioni dell'Orso, Alessandria.
- SCHIRRU, G. (2004). Costituenza metrica e lingua poetica italiana. In ALBANO LEONI, F., CUTUGNO, F., PETTORINO, M. & SAVY, R. (Eds.). *Il parlato italiano* (Atti del convegno nazionale di Napoli 13-15 febbraio 2003). Napoli: D'Auria, testo F9 (25 pp.).

ANNE BREITBARTH<sup>1</sup>, MELISSA FARASYN<sup>1</sup>, ANNE-SOPHIE GHYSELEN<sup>1</sup>,  
JACQUES VAN KEYMEULEN<sup>1</sup>

<sup>1</sup>Ghent University

## What linguistic resources reveal about rusted bicycles and hoop making

In the 60s and 70s of the last century, a collection of 783 tape recordings (ca. 700h) of spontaneous dialect speech from all Dutch-speaking provinces in Belgium, Zeelandic Flanders (Netherlands) and French Flanders (France) was assembled. The speakers were all born around the turn of the 20th century (the oldest in 1871). The collection was meant to capture the unique linguistic features of the dialects before they would be lost. However, the recordings also form a unique window into the rural culture of the region around 100 years ago, as they form the biggest collection of life-stories of common, uneducated people living in the first half of the 20<sup>th</sup> century. They contain invaluable information about topics such as disappeared professions (e.g. rope and hoop making), the first and second World War, the introduction of electricity, bikes and cars and the typical coastal Flemish fishing expeditions to Iceland, the latter of which has barely been documented in writing.

The tapes have been digitised ([www.dialectloket.be](http://www.dialectloket.be)), but not yet digitally transcribed or linguistically annotated. With an eye on fast advancing dialect loss across Flanders (De Caluwe & Van Rentergem 2011; Ghyselen & Van Keymeulen 2014), it is an urgent desideratum that this wealth of data be transcribed, annotated and made available for linguistic and historical research, as there are now only very few people who are able to understand the recordings, let alone transcribe them. The current paper reports on an on-going project developing and testing a transcription and annotation standard for this important collection on a strategic sample of 40 recordings.

The recordings are transcribed in two tiers using the free ELAN tool (<https://tla.mpi.nl/tools/tla-tools/elan/>). One tier is closer to the dialect (cf. 1a, for instance capturing clitic clusters, marked with #), and one closer to Standard Dutch (2a), in order to facilitate further (semi-)automatic annotation, as well as searching through the data with minimal knowledge of the dialect.

(1) a neen#t... k#en ik ewrocht met een  
ploef  
b neen het... ik heb ik gewrocht met een  
ploeg  
no it... I have I worked with a plough  
'No (that is not the case), I have worked with a plough.'

The transcribed data are then tokenized, lemmatized, PoS-tagged, and parsed. We opt for enrichment of ELAN-xml, as this allows maintaining the association with the time codes/the audio. The data are PoS-tagged using FROG (<https://languagemachines.github.io/frog/>) and corrected manually, in order to ensure interoperability with other corpora of spoken Dutch, esp. the CGN (<https://ivdnt.org/downloads/tstc-corpus-gesproken-nederlands>).

Given the great potential of the collection for research into cultural and oral history, the next step is the addition of keywords to facilitate searching the corpus for topics, e.g. all passages relating to hoop making. Using crowdsourcing, we have set up a network of dialect-speaking volunteers across Flanders, who have created content summaries of the tapes, accompanied by time indications per topic. These are used to add an extra time-aligned tier with these topics within ELAN-xml. The topics will be standardised and hierarchically classified using a controlled vocabulary.

Ultimately, it is the intention to combine audio, aligned transcriptions and annotations in a sustainable and searchable online corpus, made available via CLARIN in collaboration with the *Instituut voor Nederlandse Taal* (INT).

#### *Riferimenti bibliografici*

DE CALUWE, J. & VAN RENTERGHEM, E. (2011). Regiolectisering en de opkomst van tussentaal in Vlaanderen. In *Taal en Tongval*, 63: 61-77.

GHYSELEN, A.S. & VAN KEYMEULEN, J. (2014). Dialectcompetentie en functionaliteit van het dialect in Vlaanderen anno 2013. In *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 130, 117-139.

FRANCESCO CANGEMI<sup>1</sup>, JESSICA FRÜNDT<sup>1</sup>, HARRIET HANEKAMP<sup>1</sup>,  
MARTINE GRICE<sup>1</sup>

<sup>1</sup>University of Cologne

## **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 combines the speed of automatic processing with the quality of manual transcription and segmentation. While recent evaluations suggest that read speech might be efficiently segmented without requiring further manual verification (Vagnini-Holbl & Draxler, 2018), in this submission we use highly spontaneous speech as test material. We show virtually no loss of quality (compared with manual output), while reducing manual processing time by an average of 75%.

Two non-native speakers of English were recorded while playing a videogame, using two head-mounted microphones (AKG C544L) connected through an audio-interface (Focusrite Scarlett 6i6) to a computer running an audio-processing software (REAPER, Cockos, 2018), with a sample rate of 44100 Hz and a bit depth of 24 bit. From the two separated mono channels, we extracted 8 audio files of 1 minute in duration. 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. In the semi-automatic workflow, the second author transformed the audio files into video files, by pairing them with an image. The video files were uploaded to a private channel on YouTube. The YouTube automatic transcription function was used to obtain a first pass of the orthographic transcription. The automatic transcription contains timestamps for suggested beginning and end of each interpausal unit. The transcription was exported as SBV file and transformed into a TextGrid file via a custom Praat script. Errors in the automatic transcription (including imprecise timestamps) were manually corrected in Praat. Wave and TextGrid files were further processed using WebMAUS (Kisler et al. 2017), using the G2P-MAUS-PHO2SYL pipeline. This pipeline 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 all unused tiers, and leaving only the orthographic transcription and the syllable segmentation. Syllabic boundaries were then manually corrected (separately) both by the second author and by the third author (who performed the manual segmentation). This strategy allowed to counter annotator bias, both when comparing the time necessary to the two workflows (by not presenting the same audio files twice to the same annotator, and thus introducing order effects) and when comparing the precision of the two segmentations (by not evaluating syllable segmentations provided by two different annotators).

Fig. 1 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 (Manual) and semi-automatic (Auto) workflows. A qualitative inspection suggests that the two workflows yield virtually undistinguishable quality. In order to quantify this assessment, we extracted the timestamps of 1094 syllables for each of the two segmentations (Manual and Auto, and evaluated distances between matching boundaries across the two segmentations. Fig. 2 shows the distribution of the interval durations between workflows, and indicates that 68% of the semi-automatic boundaries fall within  $\pm 20$ ms from the correspondent manual boundary. More importantly, Fig. 3 shows processing times for the two workflows, as performed (separately) by the two annotators. Depending on the files, compared to the manual workflow, the automated workflow requires 64%-85% shorter processing times.

Given the small amount of available data, we refrain from providing an evaluation based on inferential statistics. The results are nonetheless encouraging, since they suggest that, compared to the manual workflow, the semi-automatic workflow provides virtually undistinguishable precision, with a substantial time processing reduction. Note that these results are underestimating the efficacy of the semi-automatic workflow, since automatic orthographic transcriptions often needed correction either because of the non-native English of the speakers or because of the use of game-specific words, which are not featured in the WebMAUS dictionaries (e.g. ‘necroguards’). Automatic syllable segmentation was perceived to be relatively easy to perform; it often involved corrections relative to sounds with unusual durations – at least when compared to the reference phone models. This is for example the case of sounds which appear to be particularly long due to interactional reasons, such as the frication noise in a turn-opening ‘so’. Applying the semi-automatic procedure on native read speech with lemmatised words would surely yield even better results.

With an inflated ratio of approximately 9 minutes of processing to 1 minute of highly spontaneous non-native speech, this free and semi-automatic workflow for orthographic transcription and syllabic segmentation could be particularly

useful for the valorisation of existing speech corpora – provided it is legally possible to upload and process them on the Internet.

Fig. 1. Spectrogram and pitch track for a sound file excerpt, with orthographic transcription (Ortho). Syllabic annotation is provided separately for the automated workflow (Auto, in SAMPA) and for manual segmentation (Manual).

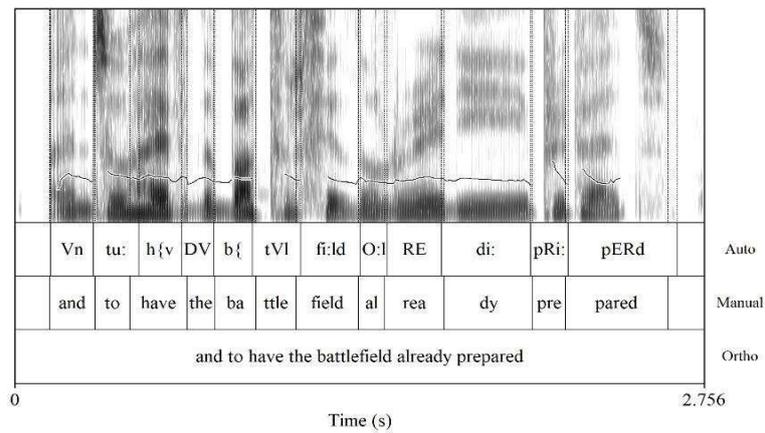


Fig. 2. Maximal temporal distance (in milliseconds) between boundaries in the automated and manual workflows. Negative numbers for late automatic boundaries. 45% of automated-workflow boundaries within  $\pm 10ms$  of boundary in the manual workflow; 68% within  $\pm 20ms$ ; 91% within  $\pm 50ms$ .

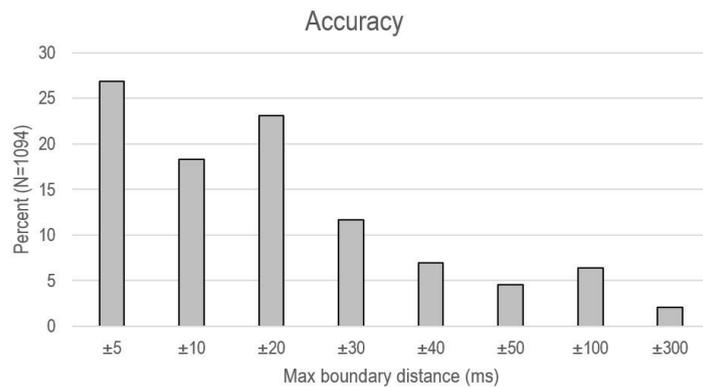
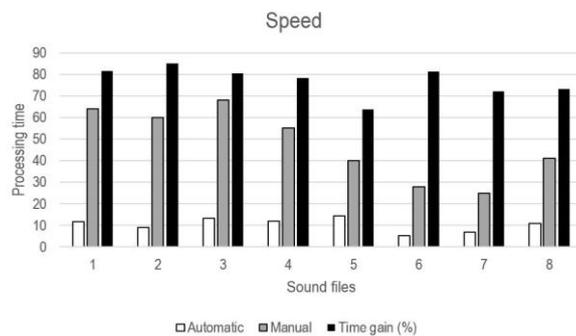


Fig. 3. Processing times (minutes) for each sound file in the Automatic workflow (white bars, ranging from 5 to 14 minutes) and in the Manual workflow (grey bars, ranging from 25 to 68 minutes). Black bars show the percent time gain when using Automatic workflow (ranging from 64 to 85%).



### Riferimenti bibliografici

BOERSMA, W. (2018). *Praat: doing phonetics by computer* [Computer program]. Retrieved from [www.praat.org](http://www.praat.org).

COCKOS (2018). *REAPER* [Computer program]. Retrieved from [www.reaper.fm](http://www.reaper.fm).

KISLER, REICHEL & SCHIEL (2017). Multilingual processing of speech via web services. In *Computer Speech & Language*, 45, 326–347.

VAGNINI-HOLBL & DRAXLER (2018). Comparing acoustic measurements from manual and automatic segmentations. *Talk at Phonetik und Phonologie im deutschsprachigen Raum*, Vienna, 6-7 September 2018.

FRANCA ORLETTI<sup>1</sup>, FRANCESCA SOCRATE<sup>2</sup>

<sup>1</sup>Università degli Studi Roma Tre, <sup>2</sup>AISO

## **Voci del 68: donne e uomini ridono in un corpus di storia orale**

### **Obiettivi**

Il nostro lavoro si propone di mostrare la rilevanza e la ricchezza euristica dell'interazione tra due discipline tradizionalmente distanti: la linguistica e la storia. In particolare intendiamo sottolineare la produttività di strumenti linguistici come la linguistica dei corpora e l'analisi della conversazione nell'analisi di corpora di narrazioni conversazionali, fonti su cui si basa la storia orale.

### **Metodo**

Il corpus è costituito dall'insieme delle interviste raccolte nell'arco dell'ultimo decennio a uomini e donne che hanno partecipato in varie sedi universitarie italiane al Sessantotto. Si tratta di interviste orali non strutturate che non hanno alle spalle la riflessione di chi scrive, corregge, pensa e ritorna sul suo testo. Le interviste sono state tutte digitalizzate e sono 63, pari a 842.378 occorrenze e a 37.443 parole; a ciascuna intervista sono associate una serie di note etnografiche riferite alla persona intervistata: il genere sessuale, l'anno di nascita, il luogo di nascita, la sede universitaria nel '68, la facoltà d'appartenenza, il titolo di studio superiore, la professione del padre, la professione della madre, la condizione professionale o lavorativa al momento dell'intervista, le origini regionali. Il corpus è stato analizzato per mezzo di un software (TaLTaC2) di analisi automatica del testo. Tale analisi consente di dare delle rappresentazioni del fenomeno studiato su base quantitativa sia a livello di unità di testo (parole) sia a livello di unità di contesto (frammenti/documenti), quindi come linguaggio utilizzato e come contenuti trattati nel testo. Il programma è stato ideato e elaborato da un team internazionale coordinato da Sergio Bolasco. TaLTaC2 utilizza risorse sia di tipo statistico sia di tipo linguistico, altamente integrate fra loro, e consente un'indagine a due livelli, lessicale e testuale, ovvero l'analisi del testo (text analysis) e il recupero e l'estrazione d'informazione, secondo i principi del data mining e del text mining. Il primo dato prodotto dal software è il vocabolario complessivo delle interviste: un elenco di tutte le parole presenti nel corpus, ordinato in una tabella secondo l'ordine di frequenza, ovvero secondo il numero di volte in cui quelle parole ricorrono all'interno del corpus complessivo delle interviste. Il vocabolario caratteristico o vocabolario specifico è un sottoprodotto del vocabolario complessivo del corpus e del suo incrocio con i tratti etnografici, ovvero il vocabolario specifico dei gruppi in cui può essere suddiviso il corpus: è la lista delle parole specifiche di quel gruppo, quelle che lo caratterizzano rispetto

agli altri. È basato su un algoritmo che ci dice se la presenza di una determinata parola in quella determinata partizione è sovradimensionata rispetto alle attese (ricorre con una frequenza più alta che altrove) senza che questo risultato sia dovuto al puro effetto del caso (sulla base di un valore che si chiama p-value, valore di probabilità: più basso è questo valore, più alta è la specificità della parola). Sono insomma le parole proprie di quel gruppo. Il vocabolario specifico è il dato primario su cui si è fondata la nostra analisi. Ogni volta sono state prese in considerazione le prime 500 parole di ciascuna lista delle specificità. Per ciascuna di quelle parole, attraverso lo strumento delle concordanze, è possibile risalire rapidamente al contesto testuale in cui è stata pronunciata e al nome della persona intervistata. È il momento in cui l'anonimato del vocabolario specifico di un gruppo si ritrae per restituire al soggetto la sua individualità e il suo racconto. Altri strumenti di analisi sono le co-occorrenze e il linguaggio peculiare. Le co-occorrenze sono le principali associazioni tra una coppia di termini che compaiono vicini nel testo entro un ridotto numero di parole: i due termini della coppia possono essere contigui o intercalati da altre parole. Le co-occorrenze ci dicono quali associazioni di parole sono più frequenti in un testo, o in un insieme di testi. Il linguaggio peculiare poggia sul valore di peculiarità che è dato, per ogni parola del corpus, dal confronto con un altro vocabolario sulla base dei rispettivi indici di frequenza. Nel caso di TaLTaC2 il vocabolario di confronto è costituito da dieci annate del quotidiano «la Repubblica» degli anni '90: un corpus quantitativamente rappresentativo di una forma di comunicazione (il testo giornalistico scritto) che più di altri si presta a rispecchiare con una certa ampiezza gli usi standard medio-alti dell'italiano. Se per collocazione sociale degli intervistati la lingua del corpus è sicuramente medio-alta, diverso è invece il registro comunicativo tra i due vocabolari: di fronte a questo limite sostanziale del confronto con il linguaggio peculiare, sono stati utilizzati i risultati delle peculiarità solo come valori indicativi. La trascrizione delle interviste, pur non avendo fatto ricorso alla trascrizione Jefferson, tipica dell'analisi della conversazione, si ispira agli stessi criteri filologici di quella. È stata adottata una trascrizione ortografica che permette la leggibilità del testo e sono stati trascritti i fenomeni paralinguistici delle pause, dei cambi di progetto come le false partenze o le interruzioni, e fenomeni caratteristici del parlato come le ripetizioni. Ciascuna trascrizione è integrata con delle glosse e dei commenti che descrivono comportamenti non verbali come le risate e i pianti nel continuum che va dalla risatina al riso pieno, dal silenzio di commozione al pianto.

#### **Oggetto dell'indagine**

Nell'analisi dei trascritti, attraverso il vocabolario specifico per appartenenza di genere, è emersa una occorrenza maggiore di risate nelle interviste alle donne, tanto che nelle specificità femminile compaiono “ride” al secondo posto e “ridono” al ventiseiesimo. Diverse le ipotesi interpretative: un segno di leggero distanziamento da ciò che si dice, un modo di sottrarsi a toni autocelebrativi, oppure una implicita, non verbalizzata complicità di genere, tra l'intervistatrice e

le intervistate? In ogni caso un tratto distintivo su cui intendiamo riflettere nel presente lavoro. Ci occuperemo pertanto del riso nelle diverse posizioni nei turni all'interno del dialogo prodotte dai partecipanti all'interazione, ovvero le intervistate, gli intervistati e l'intervistatrice. L'analisi della conversazione ha messo in evidenza il carattere estremamente pervasivo della risata nell'interazione umana e il suo ruolo multifunzionale a seconda della posizione all'interno dei turni della sequenza interazionale. In base al carattere condiviso o meno del riso, la risata può essere interpretata come espressione di affiliazione, intimità, resistenza, distacco e rappresentazione delle identità. Sulla base di un attento esame che, partendo dalle glosse di commento, ritorna sul dato primario della registrazione dell'intervista, si cercherà di individuare il ruolo svolto dall'espressione della commozione nel corpus orale analizzato. Il nostro obiettivo è pertanto mostrare il contributo che l'analisi di un fenomeno così fine dell'interazione può fornire nell'individuazione delle identità e della soggettività che intervistate e intervistati costruiscono attraverso il loro racconto, in un procedimento per cui dal microcosmo dell'interazione si risale al macrocosmo della società in un determinato periodo storico così come è vissuta nella memoria del singolo soggetto.

#### *Riferimenti bibliografici*

- BOLASCO, S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Roma: Carocci.
- GLENN, P. & HOLT, E. (Eds.) (2013). *Studies on Laughter in Interaction*. London: Bloomsbury.
- PORTELLI, A. (2017). *Storie orali, Racconto, immaginazione, dialogo*. Roma: Donzelli.

DANIELA-CARMEN STOICA<sup>1</sup>, ANYLA SARAÇI<sup>1</sup>

<sup>1</sup>“Fan S. Noli” University of Korçë

## **Oral Histories and the Dialectal Research in Albania**

The present article aims at discussing the advantages but also limitations of oral histories (both in the form of video recordings and in the form of interviews), as the sociolinguists' main tools for linguistic and dialectal research, as well as for the historians to document various historical moments in Albania. The article draws mainly on the empirical experience of the two authors, who have performed their research activity mainly in Korce 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 (Nevaci, 2011). 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 data bases 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 particular 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 nomad 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.

Last but not least, 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 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. In addition, their traditional nomad lifestyle contributed to a large extent to their spread in the Balkans. Our method of research combines the sociolinguistic interview with the ethnographic qualitative research, 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* and *dialogism* as relational processes (i.e. identity seen as a socially grounded process that involves a continuous confrontation of the self with others- Wilson, 1990; De Fina, 1995; Van Dijk, 2010).

Our study will show that certain phonetical traits and styles of speaking in the case of the Albanians living in Devoll area and Aromanians from Korce become symbolically associated with 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. From the collected material, it appears that in the case of the Aromanians living in Korce area, the frequency of pronouncing the apical *r* with several vibrations is more frequent with men, who, according to Capidan (1943: 78-79), have lost their uvular *r*, due to their social mobility, to their bilingualism or even trilingualism. 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, a situation which we are going to exemplify. On the other hand, in the case of the Devoll speech of the Tosk dialect, the oral histories belonging to speakers around 80 years old have highlighted phonetical traits, such as the use of the consonant groups [kl], [gl], as in *këlishë* (>kishë 'church'), *glu* (>gju 'knee'), present in the old Albanian<sup>1</sup> and today in the Arberesh speeches of Italy. Thus, these speakers are considered to be the only ones left still preserving the Devoll speech, while the oral histories of the other age groups present more phonetical characteristics of the standard variant.

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.

---

<sup>1</sup> The Albanian language before its separation into the two main dialects, Geg and Tosk, around the 15th century.

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 Korce 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 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.

The number of the informants was 60 (30 Albanians and 30 Aromanians), males and females in a balanced number, of various ages and having various education levels. The age groups we identified are the following: 60 to 80, 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 winter they used to lead them near the sea, where they would build shelters for them), many of them being without formal education or having only the basic level. 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.

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), whereas 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 Phonogramic 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.

### *Riferimenti bibliografici*

AKADEMIA E SHKENCAVE E SHQIPËRISË, INSTITUTI I GJUHËSISË DHE I LETËRSISË (2007, 2008) *Atlasi dialektologjik i gjuhës shqipe*, hartues prof. dr. J. GJINARI (drejtues), prof. dr. B. BECI, prof. dr. G. SHKURTAJ, prof. dr. X. GOSTURANI, Vëllimi I, II, Universita degli Studi di Napoli L'Orientale.

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

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, 351–75.

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

NEVACI, M. (2011). *Graiul aromânilor fărșeroși din Dobrogea*. București: Editura Universitară.

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

SHKURTAJ, GJ. (1999). *Sociolinguistika*. Tiranë: SHBLU.

VAN DIJK, T.A. (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.

LUKASZ STOLARSKI<sup>1</sup>

<sup>1</sup>Jan Kochanowski University

## **Building a Phonetic Corpus from Librivox Audiobooks**

“Librivox.org” is a non-profit library of free audiobooks recorded by volunteers. The website was initiated in 2005 by Hugh McGuire. It offers more than 12 000 finished projects, most of which are novels in English.

The current project aims to explore the possibilities of creating a specialized corpus on the basis of the aforementioned audiobooks library. The corpus would be especially useful for phoneticians and other linguists, but could be also applied in other domains related to natural language.

At the initial stage of the project, 105 English audiobooks were downloaded from ‘librivox.org’. In order to obtain a sample representing different dialects and genders, only these audiobooks were chosen which were read by groups of readers, rather than an individual person. After that, the corresponding text versions of the novels were found at ‘gutenberg.org’. With the use of a Python script, the texts were divided into syntactically and prosodically independent units. Next, these units were automatically aligned with the corresponding parts in audiobooks using *Aeneas*, which is a Python/C library designed to automatically synchronize audio and text. Additionally, with the use of various scripts written in Python, all the text units and the corresponding recordings were classified according to numerous criteria, such as context (narrator vs. dialogue), pragmatic function (statement/directive vs. question vs. exclamative statement), reader’s gender (female vs. male), reader’s dialect (American vs. British vs. Australian vs. non-native), author’s gender (female vs. male), number of words, syllables and phonemes, duration, etc.

On the basis of the database obtained, it is possible to create a free online corpus offering versatile functionality. A prototype version of such a tool is currently under development. It aims to enable searching for individual words and phrases in the audiobooks. The fragments containing such words or phrases will be available for audio playback directly from a web browser and for download as mp3 files. Moreover, search options will allow filtering the results according to many criteria, such as the ones mentioned previously.

Possible future developments include enlarging the database to about 1000 audiobooks, which would be 10 times more than in the current version. The predicted duration of audio materials for such a sample would be above 10 000 hours, and the number of words in the corresponding texts would be above 100 000 000. To the best of the author’s knowledge, this would result in the largest

phonetic corpus for the English language currently available. The tool would be comparable in volume to large textbased corpora, such as British National Corpus. Moreover, text alignment could additionally be performed on the word level and part-of-speech tagging could be introduced. This would allow the user to search for individual words and phrases in a much more flexible way.

ANNA-MARIA SICHANI<sup>1</sup>

<sup>1</sup>University of Sussex

## **Data documentation, modeling and analysis of a digital oral history archive: “The Connected Histories of the BBC” project as a case-study**

This paper aims to discuss various data-related challenges around the development of a digital oral history archive, using the “Connected Histories of the BBC” project as a case study.

The “Connected Histories of the BBC” project is an interdisciplinary project towards the creation of a digital oral history archive of the BBC. The project is hosted at the University of Sussex (School of Media, Film and Music & Sussex Digital Humanities Lab), will run for nearly five years in the lead-up to the Corporation’s centenary in 2022, and is being funded by a grant from the Arts and Humanities Research Council (AHRC-UK). The project is also supported by key partners in the field: the BBC itself, the Science Museum Group, the Mass Observation (MO), and the British Entertainment History Project (BEHP).

The project aims to tell the story of the BBC directly through the voices of the people who worked there. Their oral interviews offer unique accounts of how the BBC has developed the arts of broadcasting and seen the world of politics and culture. Yet, not only are they inaccessible to all but a select few; they are also unusable - scattered, un-catalogued, preserved in multiple formats, from videotape, audio files to crumbling paper.

This project aims to enrich Media History and Oral History with Digital Humanities methodologies and tools: from digitisation of the oral interviews, data management, curation and structuring to innovative data analysis techniques, in order to create a new digital catalogue of the entire oral history collection. Furthermore, the project seeks to employ 'linked open data' (LOD) technologies that would allow the diverse dataset to be interconnected with other resources. The resulting digital catalogue will allow historians, scholars and the general public - with their own memories of the BBC - to search for the first time ever this archive for a myriad of links between people, places and events, spanning decades of broadcast history.

By bringing together recent developments on the areas of data documentation, structuring, modelling and analysis, we are informing our technical development procedures with current research trends from the fields of Digital Humanities, (Audiovisual) Data Curation and Digital Cultural Heritage.

This paper will focus on three pivotal research and development areas of the project related to the digital oral history archive of the “Connected Histories of the BBC” project, by discussing mainly the challenges and the related decisions on these fronts:

**Data documentation**

The BBC oral history archive, currently containing a variety of files and formats, needs to be digitised to highest standards, documented and tagged in such a way that will allow its further processing. Issues of metadata standards for various data formats and qualities as well as preservation standards need to be addressed too.

**Data modeling**

A robust data modeling strategy will provide the project team with the conceptual tools in order to describe more accurately data structures, relationships and data semantics. By adopting a curation and research-driven data modeling strategy, a linked open data framework will be developed as well as a domain-specific ontology for oral history.

**Data analysis**

The project aims to develop and apply a set of sophisticated data analysis techniques for the oral history interviews, such as entity extraction, data mining, topic modelling etc. One of the biggest challenges for the project team is to move beyond the text-based analytical approaches to oral history collections, and to experiment with audio feature analysis and Music Information Retrieval (MIR) methodologies in order to extract semantic information from the oral history dataset.

MANUELA SOLDI<sup>1</sup>

<sup>1</sup>Festivaletteratura

## **L'archivio sonoro di Festivaletteratura di Mantova**

Festivaletteratura è una manifestazione all'insegna del divertimento culturale, nata nel 1997 dall'iniziativa di un comitato di otto volontari che chiesero la collaborazione della cittadinanza per realizzare una grande festa del libro in città, ancor oggi resa possibile ogni anno da circa ottocento volontari. Molti dei protagonisti di questa iniziativa avevano in precedenza partecipato a un percorso volto alla valorizzazione delle risorse culturali portato avanti dagli enti locali lombardi con la collaborazione dell'agenzia inglese Comoedia, che aveva visto l'istituzione di un tavolo pubblico-privato per l'individuazione dei punti di forza della città e di strategie di valorizzazione. Da questo confronto era emersa l'idea di fare di Mantova una città del libro. Il festival, ormai uno dei più longevi eventi letterari in Italia, ha mutuato ed adattato il modello dei festival letterari anglosassoni per creare una manifestazione diffusa nel piccolo centro padano finalizzata all'incontro informale tra autori e lettori che andasse oltre la mera presentazione di novità editoriali, stimolando la creazione di contenuti ad hoc. Agli incontri più classici si sono affiancati fin dalle prime edizioni laboratori, spettacoli, concerti, percorsi guidati e installazioni. Articolata su cinque giorni (all'inizio di settembre), la kermesse propone eventi che si tengono contemporaneamente in vari punti della città (una ventina di location ogni anno) stimolando il pubblico a una scelta consapevole e alla costruzione di percorsi personali all'interno di un'offerta molto varia.

A partire dal 2009 il Comitato organizzatore ha intrapreso un lungo percorso volto alla creazione di un archivio che raccogliesse la documentazione prodotta dalla prima edizione ad oggi dalla manifestazione e dall'Associazione Filofestival, che raduna i volontari della manifestazione. Questa iniziativa presentava caratteristiche di novità nel panorama dei festival letterari nazionali, e prevedeva nel tempo la creazione di 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 dove, anche al di fuori del festival, si potesse rivivere la stessa atmosfera e apprezzarne nuovamente l'offerta culturale lungo il corso dell'anno.

I primi passi di questo percorso vengono mossi nel 2009 con l'avvio del progetto "Un archivio per l'innovazione. I giovani e i documenti di Festivaletteratura" con il sostegno dell'ANCI, che l'aveva inserito nella Rete dei Festival aperti ai giovani, del Ministero della Gioventù e del Comune di Mantova. Durante i primi mesi del 2010 un corso di formazione ospitato presso l'Archivio

di Stato di Mantova ha coinvolto 13 giovani aspiranti archivisti provenienti da tutta Italia. Tra il corpo docente operatori attivi in alcune tra le maggiori realtà archivistiche italiane che ha accompagnato i ragazzi in un percorso tra le varie tipologie di archivio e di documentazione. A seguito di questo periodo di formazione, una serie di stages li ha coinvolti nell'allestimento dell'Archivio di Festivaletteratura.

Nel settembre 2010 è stata aperta la sede stabile dell'Archivio dove è confluita la documentazione raccolta nei primi mesi di lavoro e continua il lavoro di raccolta e catalogazione edizione dopo edizione, dato che si tratta di un archivio aperto, prodotto da un ente in piena attività. Il patrimonio dell'archivio è eterogeneo: alla documentazione amministrativa e relativa alla direzione creativa, prodotta dal comitato organizzatore e dalla segreteria organizzativa (ca 500 unità), si uniscono oltre centosettanta mila immagini fotografiche e circa duemilacinquecento video relativi alle varie edizioni, ai suoi ospiti, alla vita dei volontari per lo più prodotte dai volontari stessi, a una vasta rassegna stampa, a gadget, materiali promozionali, grafica, alla biblioteca di conservazione delle pubblicazioni curate dalla manifestazione, si affianca un importante patrimonio sonoro, ovvero la registrazione di oltre tremila eventi tenuti dal 1998 ad oggi dagli oltre quattromila ospiti del festival, tra i quali figurano i protagonisti della cultura internazionale degli ultimi vent'anni.

Quando nel 1998 il comitato organizzatore ha iniziato a effettuare le registrazioni, lo scopo era la documentazione interna e l'auto-memoria del proprio operato. Non era prevista la registrazione integrale di tutti gli eventi del festival, ma solo di una selezione, la cui entità è poi cresciuta negli anni. Solo in seguito è stato percepito il valore storico di questa operazione, tanto più che i contenuti espressi dagli ospiti durante la loro partecipazione alla manifestazione spesso non hanno visto altra forma di pubblicazione. 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 si rivelano oggi superate ma che possiedono un notevole valore storico per la ricostruzione del dibattito intorno alle più varie questioni, voci che parlano in un evento pubblico ma in un contesto spesso avvertito come protetto e informale, talvolta confrontandosi su problematiche che esulano dai campi abituali degli autori coinvolti. Per questi motivi nel 2014 il patrimonio dell'archivio di Festivaletteratura è stato dichiarato dalla Soprintendenza archivistica della Lombardia di notevole interesse culturale.

L'archivio sonoro riflette in pieno la multiformità del festival, che ha prodotto in ventidue edizioni quasi cinquemila e cinquecento eventi senza mai darsi temi precisi, ma esplorando i campi del sapere: dalla storia alle scienze, dalla musica alle nuove tecnologie, dalla narrativa all'ambiente, con un occhio di particolare riguardo per le giovani generazioni (sia per quanto riguarda il pubblico che gli autori emergenti) e ha sollecitato la riflessione sulla loro indicizzazione, necessaria per permettere all'utenza di esplorare una così vasta offerta. La sfida che si è

presentata agli operatori dell'archivio era di comporre una lista di tag contenutistici che rispecchiasse questa varietà, facilitando l'accesso alla documentazione sulla base degli interessi dell'utente, ma anche della comunità letteraria entro la quale si ascrive l'opera dell'autore trattato, e ancora del format dell'evento. Una risposta è stata fornita guardando alle catalogazioni per soggetto in uno in ambito biblioteconomico, dalle quali si è partiti per allestire un lemmario gerarchico.

Nel 2015, il ventennale del festival ha fornito l'occasione per avviare la pubblicazione online di una vasta selezione della sezione sonora dell'archivio. Sono online in streaming on demand attualmente oltre duemila file, che aumenteranno ulteriormente quando nei primi mesi del 2019 saranno pubblicati quelli relativi all'edizione 2018. Diverse le problematiche incontrate nel corso degli anni necessari per raggiungere questo obiettivo. La prima riguardava il diverso stato della documentazione. Il Festival è nato e si è sviluppato in un periodo relativamente breve, segnato però da una velocissima evoluzione tecnologica dei supporti e delle tecniche di produzione documentaria, quindi la documentazione e il materiale audiovisivo e iconografico si trovano in parte su supporti cartacei e analogici e in parte allo stato digitale. La repentina obsolescenza dei supporti e delle tecnologie digitali ha reso urgente non solo la digitalizzazione di ciò che si trovava ancora su supporto cartaceo e analogico, ma anche un'adeguata conservazione della documentazione nativa digitale, che scongiurasse il pericolo di perdite di dati o di metadati e facilitasse la fruizione da parte del pubblico. Sono perciò stati allestiti progetti di acquisizione digitale e un portale archivistico per la consultazione del materiale catalogato. La fruizione è tuttavia possibile anche attraverso il portale Soundcloud, dove vengono effettivamente caricati i files sonori. In questo modo a platea raggiunta si amplia, talvolta raggiungendo fasce di pubblico che non conosce il festival. Recentemente si è scelto di effettuare un esperimento di pod-cast, che prevede la pubblicazione periodica di registrazioni di singoli eventi conservate in archivio nel corso dell'anno.

L'archivio può essere fruito in molti modi diversi, ponendo l'accento sui contenuti oppure, per esempio, puntando sulla ricchezza linguistica, essendo presenti moltissime registrazioni in lingua straniera, e che riceve una buona percentuale di ascolti anche dall'estero. Gli eventi che prevedono ospiti stranieri sono da sempre provvisti di una traduzione consecutiva, e la loro registrazione permette dunque una fruizione bilingue<sup>1</sup>. Per questo motivo il festival cerca promuovere questo nucleo documentario tra gli insegnanti, come supporto d'apprendimento per la comprensione orale delle lingue straniere (la più rappresentata è l'inglese, seguita da francese e spagnolo).

---

<sup>1</sup> Solo in anni più recenti il festival ha iniziato a organizzare un limitatissimo numero di eventi in lingua originale senza traduzione, ma anche in questo caso la registrazione in lingua è perfettamente fruibile come l'evento originale.

Le sue modalità di fruizione e valorizzazione sono oggetto di continua riflessione all'interno dell'organizzazione, che tenta di fare di questo patrimonio e della sua riscoperta un punto di partenza per nuove elaborazioni culturali, e vi ha attinto per realizzare pubblicazioni, spettacoli, installazioni. Un percorso ormai quasi decennale che pone sfide continue per mantenere vivo un patrimonio sempre più vasto e significativo.

SONIA CENCESCHI<sup>1</sup>, ALESSANDRO TRIVILINI<sup>1</sup>, LICIA SBATTELLA <sup>2</sup>,  
ROBERTO TEDESCO<sup>2</sup>

<sup>1</sup>Scuola Universitaria Professionale della Svizzera Italiana (SUPSI), <sup>2</sup>Adaptable, Relational and Cognitive Software Environments Laboratory, Politecnico di Milano

## Collecting Italian spontaneous social media speech: the WAsp2 project

This work presents the WAsp2 project (WhatsApp SPontaneous SPeech), focused on collecting private social media speech thanks to the WhatsApp application. The main purpose is to collect a wide corpus of spontaneous vocal messages useful in forensic investigations but also in other research topics such as experimental linguistics, speech therapy, artificial intelligence, and ASR systems.

WhatsApp is a *social media* (Kaplan, 2015) app that allows people to connect with other members, exchanging text messages, photo, audio and video materials. Nowadays, social media constitute a fundamental part of human communications and they are widely used from people of all ages and professions, representing a very large digital pool from which to draw. We think WhatsApp audio messages contributed to introduce a new speech communication style (Cenceschi et al., 2018; Nencioni 1983), characterized by broken conversations and non-consecutive turns, often alternated with text, media, and long temporal breaks. The WAsp2 corpus will be also useful to investigate these communication behaviours and new speakers' expectations compared to the traditional speech.

Many works explore the textual social media language (Stark, 2016-2018; Al Khawaldeh *et al.*, 2016; Forsyth & Martell, 2017), but far less are those focused on the speech, mostly related to voicemail phone recordings (Chateau *et al.*, 2004; Koumpsis & Renals, 2005; Inanoglu & Caneel, 2005). A large variety of spontaneous Italian speech corpora have been collected among years (Cresti *et al.*, 2002; Cresti & Panunzi, 2013) but none of them focused on social media-style speech. The structure of this work is partially inspired by the *Common Voice project* (Mozilla common voice, 2017) by Mozilla and the *Corpus Compilation of Private Social Media Messages* (Verheijen & Stoop, 2016) but it differs for two main reasons: Wasp2 regards strictly vocal messages and collects spontaneous speech.

Italian native speakers are asked to donate WhatsApp private chats compiling an on-line questionnaire and then sending chat contents by e-mail. We are currently developing the WAsp2 website. A landing page contains a short presentation and the direct link to the authorization to process personal data for research purposes. Once accepted, the donor is redirected to a short

questionnaire. We prefer to insert few questions in order to lighten the participant, even if it means losing details. Its purpose is to allow to catalogue recordings according to the following classes.

- Age
- Gender
- Education level
- Geographical origin (where the donor spent most of her/his life)
- Phone model
- If headset has been used (“I don’t remember” option is provided)

The selection is not mandatory, but chats without these data will be separate from the others in the final dataset. Once the questionnaire is fulfilled, the screen displays the project e-mail and the summary with WhatsApp instructions to export complete chats or single audio messages through different devices (iPhone/Android/Win phones). This passage can be easily done using the app through the WhatsApp supplied functions. For privacy reason we’ll extrapolate only the donor’s chat components. Video and images are deleted while audio recordings are divided from the text. Texts are saved for further processing, while audio messages are filtered to keep only the ones containing human speech. ASR techniques are then used to extrapolate transcriptions.

Another webpage will be finally developed in order to validate randomized audio and their corresponding transcriptions. A narrow group of natural language experts will check part of materials; this part is however still being defined.

The “Protocol for the collection of forensics databases” (Morrison *et al.*, 2012) suggests collecting at least two noncontemporaneous recordings of each speaker using different styles, but we prefer to suggest people to send more than 1 chat without further requirements: being them sensitive private data, we prefer to collect as much material as possible, opting for the volunteer’s light commitment.

To our knowledge, WAsp2 is the first attempt to ask people to donate their WhatsApp speech so we cannot predict how they will react. An encouraging example is the multilingual large-scale corpus of “What’s up, Switzerland?” (Stark, 2016-2018), which collected about 967 textual chats (1 291 022 messages) in four years; it is difficult to make a comparison, however, because people could be more reluctant to donate their voice.

Moreover, we start focusing on the northern Italy and Switzerland varieties reducing the possible number of donors, but being able to rely on a powerful dissemination network. It must be considered that advertising will be more widespread with respect to the previous project, and our intention is to re-launch WAsp2 at fixed periods without a specific deadline.

A first survey, carried out among twenty people, collected only positive responses and returned an interesting suggestion: a crucial aspect is to create a relationship of trust with potential donors. The project will be then spread random, basing the disclosure on public interviews, newsletters to other

laboratories or affiliated companies, and exploiting the classic word of mouth among personal contacts. Another possibility is to pay donations, but despite it could be helpful to obtain more chats, we prefer to evaluate the option when the website will be finished.

The forensic research needs to deal with spontaneous speech, but corpora are very limited due to privacy reason. Many datasets are collected among phone services and media (TV, radio) but existing Italian spontaneous speech corpora (Cresti, 2000; Albano Leoni, 2006) cannot provide enough data to train recognition algorithms. The request for new spontaneous speech recordings is very high, but the only way to make up for its lack is currently to approximate reality with dedicated tasks (Baker & Hazan, 2011) or using recited/read speech. WAsp2 aims to give a contribute using a new collection methodology. Moreover, WAsp2 environmental recording contexts will be really variable, ensuring different background noise as well as emotional states, gender, age and speakers' speech styles. These characteristics are really useful in the forensic field because they give the opportunity to best approximate real cases. Finally, the WAsp2 corpus will be a possible reference dataset for speaker recognition investigation and Likelihood Ratio calculation, to analyse the influence of quality and audio compression on voice features variations or in investigations related to gender and speech rate.

The corpora will be freely available upon request for research and other non-commercial purposes and gradually updated with new incoming data.

### *Riferimenti bibliografici*

(2017) Mozilla common voice. [Online]. Available: <https://voice.mozilla.org/en>

ALBANO LEONI, F. (2006). *Il corpus clips. presentazione del progetto.*

AL-KHAWALDEH, N., BANI-KHAIR, B., MASHAQBA, B. & HUNEETY, A. (2016). A Corpus-Based Discourse Analysis Study of WhatsApp Messenger's Semantic Notifications. In *International Journal of Applied Linguistics and English Literature*, 5(6), 158-165.

BAKER, R., HAZAN, V. (2011). *Diapixuk: task materials for the elicitation of multiple spontaneous speech dialogs.*

CENCESCHI, S., SBATTELLA, L. & TEDESCO, R (2018). Verso il riconoscimento automatico della prosodia. In *STUDI AISV*, 433-440.

CHATEAU, N., MAFFIOLO, V., & BLOUIN, C. (2004). Analysis of emotional speech in voice mail messages: The influence of speakers' gender. In *Eighth International Conference on Spoken Language Processing.*

CRESTI, E. (2000). *Corpus di italiano parlato: Introduzione, volume 1.* Accademia della Crusca.

CRESTI, E., MONEGLIA, M., DO NASCIMENTO, F. B., MORENO-SANDOVAL, A., VÉRONIS, J., MARTIN, P., ... & BLUM, C. (2002). The C-ORAL-ROM Project. New methods for spoken language archives in a multilingual romance corpus. In *LREC.*

- CRESTI, E., PANUNZI, A. (2013). *Introduzione ai corpora italiani*. Bologna: Il Mulino.
- FORSYTH, E. N., MARTELL, C. H. (2007). Lexical and Discourse Analysis of Online Chat Dialog. In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, 19-26.
- INANOGLU, Z., CANEEL, R. (2005, January). Emotive alert: HMM-based emotion detection in voicemail messages. In *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, 251-253.
- KAPLAN, A. M. (2015). Social Media, the Digital Revolution, and the Business of Media. In *International Journal on Media Management*, 17(4), 197-199.
- KOUMPIS, K., RENALS, S. (2005). Automatic summarization of voicemail messages using lexical and prosodic features. In *ACM Transactions on Speech and Language Processing (TSLP)*, 2(1), 1.
- MORRISON, G. S., ROSE, P. & ZHANG, C. (2012). Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. In *Australian Journal of Forensic Sciences*, 44(2), 155-167.
- NENCIONI, G. (1983). *Di scritto e di parlato*. Zanichelli.
- STARK, E. (2016-2018). *SNSF project "What's up, Switzerland?"* (Sinergia: CRSII\_160714). University of Zurich. [www.whatsup-switzerland.ch](http://www.whatsup-switzerland.ch).
- VERHEIJEN, L., STOOP, W. (2016, September). Collecting facebook posts and whatsapp chats. In *International Conference on Text, Speech, and Dialogue*. Cham: Springer, 249-258.

PAOLO MAIRANO<sup>1</sup>, ENRICO ZOVATO<sup>2</sup>, VITO QUINCI<sup>2</sup>

<sup>1</sup>Université de Lille, <sup>2</sup>Università degli Studi di Torino

## **Do sentiment analysis scores correlate with acoustic features of emotional speech?**

### **Theoretic background and motivation**

Emotional and affective information can be conveyed at different levels (Reilly & Seibert, 2003): lexical (intensifiers, modals, hedges, etc.), syntactic (e.g. relative clauses to comment on actions and behaviors), paralinguistic (facial expressions, gestures), and by voice. The mainstream frameworks for the analysis of emotional and affective voice characteristics are based either on categories such as anger, excitement, disgust, fear, relief, sadness (e.g. Ekman, 2000), or on dimensions (e.g. Russel, 1980). The latter tend to define emotions as coordinates in a multidimensional space, each dimension representing a property of an emotional state. For instance, the framework proposed by (Russel, 1980) postulates two dimensions of emotions, namely valence (positive vs. negative) and arousal (high vs. low activation): a sad state can be defined as having negative valence and low activation, while a euphoric state can be defined as having positive valence and high activation. The classification of emotional states has proved difficult within categorical as well as dimensional frameworks, and the situation is further complicated by the interfering role of other linguistic levels (lexical, syntactic), as mentioned below: it is not yet clear whether lexical and syntactic features can be considered as ancillary or complementary to voice characteristics for the expression of emotions, and whether the relation between these two levels of analysis can change according to the style or the context of communication.

Despite such intrinsic problems, many studies have focused on emotional speech with the aim of finding acoustic correlates of affective states. For example, it has been found that limited pitch variations characterize affective states with negative valence, while higher pitch and higher pitch range generally characterize affective states with positive valence (Burkhardt & Sendlmeier, 2000). Beyond pitch, other prosodic features seem to play an important role in conveying emotions, and various other acoustic and perceptive correlates of emotions have been suggested in other studies (Banse & Scherer, 1996; Schröder *et al.*, 2001; Audiber *et al.*, 2005). However, one of the problems affecting such studies is the availability of reliable data. Given the difficulty of obtaining controlled emotional data elicited in an ecological context, most studies make use of acted speech, with the clear drawback that the resulting emotional speech tends to be prototypical or overacted (Scherer, 2003) and does not necessarily

correspond to natural realizations. In this contribution, we investigate the correlation between sentiment analysis metrics and acoustic characteristics of speech, as measured (Charfuelan & Schröder, 2012) on audiobooks. A similar analysis investigating the correlation between sentiment analysis scores and acoustic features has been attempted by (but only on data from one speaker and one audiobook. We extend the analysis to 251 audiobooks in the hope to serve a practical purpose (the investigation of whether sentiment analysis can be of help in gathering ecological emotional speech for analysis) and to contribute to the theoretic discussion about the roles of voice cues vs lexicon in expressing emotions.

### **Data and methodology**

In order to study the correlation of sentiment analysis scores and acoustic features of emotional speech, we used audiobooks, as per (Charfuelan & Schröder, 2012). Audiobook recordings came from the *LibriSpeech* ASR corpus, more precisely the train-clean-100 section, containing 100 hours of clean speech from 251 audiobooks read by different speakers. All the material was transcribed phonetically with a TTS front-end component, following transcription conventions of General American. Phonetic transcriptions were then forced-aligned to the acoustic signal enabling silence detection, and finally converted to *TextGrid* format for analysis with *Praat*.

Sentiment analysis scores were extracted from text for each sentence using *Vader* and *SentiWordNet*, both available within Python's NLTK library. We extracted acoustic features at sentence level and at word level in *Praat*: mean F0 (in semitones), pitch stdev (in semitones), pitch range (0.05 to 0.95 quartiles), pitch max (0.95 quartile), pitch min (0.05 quartile), shimmer, jitter, Hammarberg index (HAM, difference between max energy in the 0-2 kHz and 2-5kHz bands, Hammarberg *et al.*, 1980), Do1000 (drop off spectral energy above 1000 Hz), Pe1000 (relative energy in the frequencies above 1000 Hz versus energy below 1000 Hz, Scherer, 1989; Drioli *et al.*, 2003). For the sentence-level analysis, we also extracted the total duration in ms from first to last phoneme (DUR), speech rate (SR), articulation rate (AR), and pause/speech ratio (PSR). All the acoustic parameters were transformed to z-scores by speaker, in the attempt to normalize individual differences.

### **Results and discussion**

The data were entered in linear mixed-effects models evaluating the relation between sentiment analysis scores and acoustic parameters. We then built separate models for negative (*Vader* score < 0) vs positive (*Vader* score > 0) sentences. Models for sentences with positive *Vader* values showed that *Vader* positivity score was a significant predictor not only for pitch parameters, but also for rhythmic and spectral parameters. Similarly, models for sentences with negative *Vader* values showed that *Vader* negativity value was a significant predictor not only for pitch parameters, but also for AR and shimmer. These

results seem to suggest that, while pitch parameters show a (modest) linear correlation with valence, rhythmic and spectral parameters correlate with arousal.

The acoustic features have then been used to train a neural network classification model, whose targets were sentiment analysis categories. The goal was to verify whether non-linearities involving acoustic features can contribute to some extent to the prediction of sentiment classes. Two classifiers were trained, one based on sentence-level features, the other on word-level features extracted on the stressed vowel. We considered three-classes models (Neg, Neu, Pos), as well as binary classification with just Pos and Neg classes. Multilayer perceptrons were trained for this experiment on the whole set of *LibriSpeech* data. They were composed of three hidden layers with ReLU activation functions and an output softmax layer. An early stopping criterion was adopted during training, on the basis of loss values of the validation set (30%), and on 10 consecutive not-improving epochs. Adam optimizer 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 such a way to have a balanced number of occurrences among the three/two categories. A subset of data (10%) not used for training were then used for evaluation. Results in Table 1 show that accuracy is little above chance level in two and three classes models.

Table 1: *Classification accuracy (precision and recall) results of NN sentiment predictors. The sentence-level gain is 0.12 with respect to chance level (which is 0.33 for three classes, 0.5 for two classes)*

<b>Accuracy</b>	<b>Word</b>	<b>Sentence</b>
3-classes	0.400 (0.402, 0.400)	0.460 (0.463, 0.461)
2-classes	0.561 (0.555, 0.619)	0.638 (0.633, 0.690)

These results suggest that lexical and acoustic cues of emotion do not necessarily go hand-in-hand, so that speakers/listeners may rely on either in order to realize or perceive emotions in speech. This may be especially true for read speech where the reader is not personally involved in the content, as for audiobooks. If so, the aptness of sentiment analysis for the study of emotional speech may be limited, since relying on such metrics would mean disregarding any acoustic cues of emotion that do not co-occur with lexical cues. However, we acknowledge a certain number of limitations of this study. Firstly, despite the undeniable advantages of a fully automatic approach for the transcription, annotation and analysis, the procedure is likely to contribute to a certain amount of noise in the data. This potentially has the effect of reducing the observable relations among the variables studied. Secondly, the lack of punctuation in *LibriSpeech* corpus sentences makes it impossible to differentiate narrative chunks from reported speech by characters, where more pronounced emotional features can be expected. Finally, rule-based open-source tools for sentiment analysis

certainly provide a convenient solution, but more accurate sentiment analysis algorithms may yield different results.

### *Riferimenti bibliografici*

- AUDIBERT, N., AUBERGÉ, V. & RILLIARD, A. (2005). The prosodic dimensions of emotion in speech: the relative weights of parameters. In *Ninth European Conference on Speech Communication and Technology*, 4-8 September, Lisbon, Portugal.
- BANSE, R., SCHERER, K.R. (1996). Acoustic profiles in vocal emotion expression. In *Journal of personality and social psychology*, 70(3), 614–636.
- 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 *4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (ES3)*, 26 May 2012, Istanbul, Turkey, Proceedings, 99-103.
- DRIOLI, C., TISATO, G., COSI, P. & TESSER, F. (2003). Emotions and voice quality: experiments with sinusoidal modeling. In *Voice Quality: Functions Analysis and Synthesis (VOQUAL) Workshop*, 27-29 August, Geneva, Switzerland, Proceedings.
- EKMAN, P. (2000). Basic Emotions. IN T. DALGLEISH, T. POWER (eds.), *Handbook of Cognition and Emotion*, London: John Wiley & Sons, 45-60.
- 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.
- JOHNSTONE, T., SCHERER, K.R. (2000). Vocal communication of emotion. In M. LEWIS, J. HAVILAND (Eds), *Handbook of emotions 2*. London-New York: The Guildford Press, 220–235.
- REILLY, J., SEIBERT, L. (2003). Language and emotion. In R.J. DAVIDSON, K.R. SCHERER & H.H. GOLDSMITH (Eds), *Handbook of affective sciences*. OUP, 535–559.
- RUSSELL, J.A. (1980). A circumplex model of affect. In *Journal of personality and social psychology*, 39(6), 1161-1178.
- SCHERER, K.R. (1989). Vocal correlates of emotion. In A. MANSTEAD, H. WAGNER (Eds), *Handbook of psychophysiology: Emotion and social behavior*, London: Wiley, 165-197.
- SCHERER, K.R. (2003). Vocal communication of emotion: A review of research paradigms. In *Speech Communication*, 40, 227-256.
- SCHRÖDER, M., COWIE, R., DOUGLAS-COWIE, E., WESTERDIJK, M. & GIELEN, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *EUROSPEECH 2001 – Seventh European Conference on Speech Communication and Technology*, September 3-7, Aalborg, Denmark, Proceedings.

RICCARDO ORRICO<sup>1,2</sup>, RENATA SAVY<sup>2</sup>, MARIAPAOLA D'IMPERIO<sup>1</sup>

<sup>1</sup>Aix-Marseille Université, Laboratoire Parole et Langage, Centre National de la Recherche Scientifique, <sup>2</sup> Università degli Studi di Salerno

## **Individual variability in Salerno Italian intonation: evidence from read and spontaneous speech**

In the last two decades, research on Italian intonation, especially within the Autosegmental-Metrical (AM) framework (Pierrehumbert, 1980; Ladd, 2008), has been rather prolific, and has shown the presence of considerable variability, not only across the different regional varieties, but even within the same variety (Gili Fivela et al., 2015). However, not every variety has received the same degree of attention by the scientific community: extensive research has been carried out, for example, for Neapolitan (D'Imperio, 1997; 2000; 2002), Pisa (Gili Fivela, 2008), and Bari Italian (Grice, Savino, 1997; Savino, Grice, 2011), and much less attention has been given to other varieties. In addition, the studies mentioned above have been conducted with different speech styles (read or spontaneous speech) and often using the AM framework in very different ways, with the exception of only a few studies that have attempted to provide a unified account of several varieties (Grice et al., 2005; Gili Fivela et al., 2015).

The aim of this study is to add to the literature on Italian intonation by providing an account of the basic phonological inventory of the tonal categories of Salerno Italian (SI), a variety that has not received much investigation in the past, and to specifically address the issue of the huge intra-variety variability.

The basic inventory of pitch accents and boundary tones in SI has been established by analyzing the productions of 4 speakers (2 males and 2 females) performing a Discourse Completion Task. The annotations were made using a ToBI-like annotation system (cf. Grice et al., 2005, Gili Fivela et al., 2015). The types of utterance analyzed are: i) broad focus statements, ii) narrow focus statements, and iii) yes-no questions (having different dialogical sub-functions, *i.e.* information-seeking, confirmation-seeking, counter-expectational, and echo questions). The material used for this study is a selection of the corpus used in Gili Fivela et al. (2015). The prosodic labeling was performed by the first author and then a representative sample of the annotations was verified by the last author (an expert labeler).

The analysis of the speech material shows the presence of three pitch accents that can occur in nuclear position: H+L\*, typically found in broad focus statements; L\*+H, typically found in yes-no questions; and L+H\*, found in both narrow focus statements and yes-no questions. Previous investigations of the

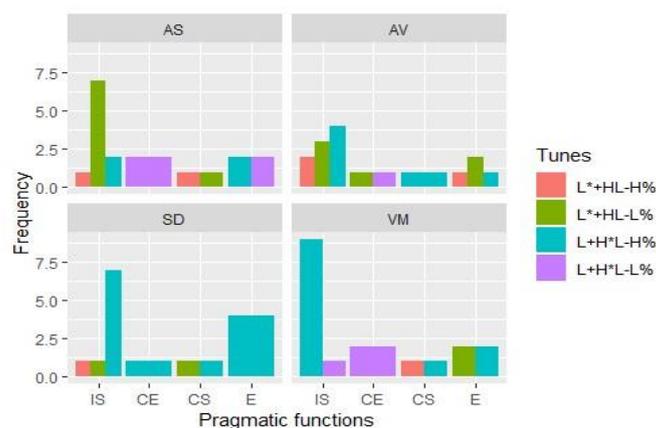
tonal system of SI propose the existence of a fourth nuclear pitch accent, *i.e.* H\*+L, which is claimed to be used to mark narrowly focused constituents in statements, as opposed to the L+H\* found in polar questions; the phonetic difference between the two accents is claimed to lie in the timing of the H\* target, which is earlier in H\*+L (Gili Fivela et al., 2015: 160); we argue that such phonological contrast does not exist and we use the label L+H\* for both accents. Also, an H\* pre-nuclear pitch accent has been found, which can occur in both statements and questions. As for the boundaries, both low (L%) and high (H%) tones are found in SI. In particular, statements – both broad and narrow focus – show systematically an L% at the end of the intonational phrase, while yes-no questions can end with both rising and falling tones.

While the tonal realization of statements appears highly consistent (H+L\* L-L% for broad focus statements and L+H\* L-L% for narrow focus statements), the analyses of polar questions show that their realization is much more variable: four different polar question tunes have been found, two of which are very frequent, *i.e.* L\*+H L-L% and L+H\* L-H%, which together amount to the 74% of the question patterns, and other two with lower frequency, *i.e.* L\*+H L-H% and L+H\* L-L%. Figure 1 shows the distribution of the different question tunes across the four speakers. As mentioned above, the polar questions analyzed include questions with different sub-functions; nevertheless, the different pragmatic uses of questions alone cannot explain the huge variability found in their distribution, which we argue to depend on a complex interplay between function and individual variability. In particular, we provide evidence of the fact that the choice of pitch accent is related to the interaction between question function and speaker, while the choice of boundary tone is mainly dependant on individual variability.

The issue of individual variability in yes-no questions is being currently investigated in a corpus of spontaneous speech: 5 pairs of SI native speakers were recorded while engaging in a ten-minute dialogue, using the ‘Spot the Difference’ dialogical game as elicitation technique. The analysis of polar questions in spontaneous speech, and its comparison with the tonal realization of read questions, provides an in-depth investigation of the use and frequency distribution of question tunes in SI, as well as allowing us to address some of the issues that still remain unresolved concerning question tunes distribution in Italian. In particular, it has been claimed that in some varieties the choice between a final fall (L%) or a rise (H%) in polar questions might depend on different speaking styles, *i.e.* read vs. spontaneous speech (cf. Gili Fivela et al., 2015); previous studies on Bari Italian, for example, argue that while the choice of pitch accent can be related to the different dialogical uses of polar questions, the difference between L% and H% is only dependant on stylistic choices, with the rising variant being typical of read-aloud speech (Refice et al., 1997; Grice, Savino, 1997). We argue that this is not the case in SI and that the distribution of falling vs. rising terminal contours is consistent across different speaking styles.

However, results concerning this part of the study are yet to be properly interpreted and will be discussed at the conference.

**Figure 1** – Distribution of the four question tunes across pragmatic functions, divided by speaker. The pragmatic functions are: Information-seeking (IS), Confirmation-seeking (CS), Counter-expectational (CE), and Echo questions (E)



### Riferimenti bibliografici

- D'IMPERIO, M. (1997). Narrow focus and focal accent in the Neapolitan variety of Italian. In A. BOTINIS ET AL. (eds), *Intonation: Theory, Models and Applications: Proceedings of the ESCA Workshop on Intonation*. Athens: ESCA/University of Athens, 87–90.
- D'IMPERIO, M. (2000). *The role of perception in defining tonal targets and their alignment*. Ph.D thesis, Ohio State University.
- D'IMPERIO, M. (2002). Italian intonation: an overview and some questions. In *Probus* 14(1): 37–69.
- GILI FIVELA, B. (2008). *Intonation in Production and Perception: The Case of Pisa Italian*. Alessandria: Edizioni dell'Orso.
- 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 *Intonation in Romance*. Oxford: Oxford University Press, 140-197.
- GRICE, M., D'IMPERIO, M., SAVINO, M., & AVESANI, C. (2005). Strategies for intonation labelling across varieties of Italian. In *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, 362-389.
- 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.

- LADD, D. R. (2008). *Intonational Phonology, 2nd edn.* Cambridge: Cambridge University Press.
- PIERREHUMBERT, J. (1980). *The phonetics and phonology of English intonation.* Ph.D thesis, Massachusetts Institute of Technology.
- REFICE, M., SAVINO, M., & GRICE, M. (1997). A contribution to the estimation of naturalness in the intonation of Italian spontaneous speech. In *Proceedings of Eurospeech 97 (Rhodos, 22-5 Sept. 1997)*, vol. 2, 783-6.
- SAVINO, M., GRICE, M. (2011). The perception of negative bias in Bari Italian questions. In S. FROTA, P. PRIETO & G. ELORDIETA (Eds.), *Prosodic categories: Production, perception and comprehension.* Dordrecht: Springer, 187-206.

ROSALBA NODARI<sup>1</sup>, CHIARA MELUZZI<sup>2</sup>

<sup>1</sup> Università degli Studi di Siena, <sup>2</sup> Università degli Studi di Pavia

## A sociophonetic investigation of Roman rhotic degemination

**Introduction** This paper aims at investigating rhotic degemination in the production of young Roman Italian speakers. We focused on rhotic degemination because it's considered a sociolinguistic marker of *romanesco* and it is overtly commented and associated with rudeness, low status speakers and informal communicative situations. Given the particular status of the R variable (Scobbie, 2006) in conveying socio-indexical information, we analyse if, and how, young Roman speakers produce degemination of the rhotic phoneme.

**Background** According to Rohlfs (1966), degemination of the intervocalic rhotic is attested in Northern Italy, but it appears to spread toward southern areas (Tuscany, Marche, Umbria) and “in misura particolarmente forte a Roma e nel Lazio meridionale” (Rohlfs, 1966: 313). The phenomenon is a strong sociolinguistic marker of Roman speaker (Stefinlongo, 1999, D’Achille et al., 2012), since it correlates both to speech style and to social groups, and it reaches the status of a stereotype (Johnstone et al., 2006). According to D’Achille (1995) the stigmatized value seems to be confined to degemination in post-tonic syllable (like *orrorè*), whereas degemination in pre-tonic position (es. *arivare*) can be found in informal speech style.

Even if it's worth mentioning Bernhard's (1998) empirical survey, no explicit sociophonetic study has been devoted to the investigation of rhotic degemination: in particular, we do not know if the phenomenon is still maintained in the speech habit of young Roman speakers, or if it is subject to variation according to different speech style. To our knowledge, the only study devoted to the problem of degemination is Marotta (2005), which shows that in a dialogic production from the corpus CLIPS, degemination does not appear in the production of unstressed syllable. Her results show that, at least for the two analysed subjects, young Roman speakers do not produce rhotic degemination, therefore confirming its low status and its possible disappearing in the production of highly educated speakers.

**Research questions** Inspired by Marotta's (2005) results, the current study has the aims of starting an investigation of rhotic realization in Roman Italian. In particular, we aim at answering at two specific questions: 1) given that rhotic degemination seems overtly stigmatized, do young roman speakers still realize it? Do they show it also in more controlled sociolinguistic speech events, like sentence reading task? 2) Assuming the maintenance of rhotic degemination in

the speech of young Roman speakers, how is this degemination realized? Namely, if we assume that the phenomenon is still maintained, do geminate rhotics show the same duration in ms. of the singleton ones? Furthermore, do other phonetic factors have a role in triggering degemination, as the anecdotal observations about the role of stress, with degemination being favoured in unstressed syllables? In this presentation we will show preliminary data from a small sample of speakers performing a sentence reading task.

**Methodology** We recorded 10 speakers from Rome, 5 males and 5 females: 7 out of 10 speakers in particular are from Eastern Rome, whereas the remaining 3 come from Southern Rome. We decided to focus our investigation of Eastern and Southern Rome, because they are considered the core of the true Roman working-class, but also the core of Roman alternative hipster artsy and musical scene, with many clubs and squatting spaces. Their age span from 23 to 30, and their education goes from high school diploma to PhD degree. We narrowed our investigation to this population to test if speakers from these areas, with these social characteristics, may use a low-status phenomenon in order to signal their affiliation to the working-class.

Speakers participated to a sentence-reading task. The list consisted of 70 sentences of equal length and controlled prosodic contour. Each sentence contained one token with a singleton and/or geminate /r/, in pre-tonic and/or post-tonic condition. Following Dmitrieva et al. (2018) some words were minimal pairs, to test if minimal pair neighbours of word can affect the acoustic realization of rhotic degemination. Each rhotic could be followed by any of the seven vowels. Sentences were constructed in order to resemble an informal, everyday communication: for this reason, we did not insert our word in a standard frame, like *Say X again*, but we constructed realistic, informal sentences, with swear words and regional terms too (es. *Non si sgarra co l'amici*). Thus, even if a sentence-reading task will normally elicit a more controlled speech, the design of the sentences stimulated the emergence of a more “vernacular-like” variety. The corpus consisted of 741 tokens, annotated in PRAAT following the protocol for rhotics developed by Celata et al. (2016).

Studies devoted to the investigation of Italian geminated consonants have demonstrated that different acoustic cues are involved in conveying gemination. In particular the most robust cue seems to be the consonantal lengthening, associated with a shortening of the preceding vowel: vowels preceding geminates are thus considerably shorter than when preceding a singleton consonant. When turning to rhotic gemination, in addition to durational parameters, qualitative dimension plays a role too: as Payne (2005) notes, contrast between singleton and geminate rhotic is usually a contrast between a tap and a trill. For the aforementioned reasons, we decided to investigate rhotics both qualitatively and quantitatively: we rely on durational measurements (preceding vowel and consonant duration) to test if the consonant is realized as a geminate or a singleton. In addition to this, the visual inspection of the spectrogram is used for

a qualitative investigation of rhotic realization, in order to analyse if rhotic degemination can be conveyed by a change, from trill to tap.

**Results and discussion** According to the bibliography, we analysed the durations of both rhotics and preceding vowels. For each variable we ran linear mixed models that included 4 fixed variables and their corresponding interactions: 1) speakers' sex; 2) gemination (tokens with lexical gemination, tokens with lexical singleton); 3) syllable structure (pre-tonic, post-tonic); 4) lexical competitor (the presence or not of a minimal pair neighbour, ex. *bara* vs *barra*). Results show that, for rhotic duration, Gemination and Lexical competitor were statistically significant, since geminated rhotics were longer than singleton rhotics, and words with a minimal pair neighbours were significantly longer than words without a corresponding one. Sex and Syllable were not significant: males and females show the same values for rhotic duration, and this is also true for pre-tonic and post-tonic rhotics. However, Sex and Syllable were found to interact: the difference between males and females was significant only for pre-tonic syllables, with males showing longer durations (males 43 ms, females 38 ms). The interaction between Gemination and Lexical Competitor showed that tokens with a minimal pair neighbour maximized the contrast between singleton and geminated rhotics, whereas tokens with no minimal pair neighbour tend to show a less sharp contrast between singleton and geminates. When looking at vowel duration, results show that only Syllable stress was statistically significant, with stressed syllable being longer than unstressed one (stressed = 113 ms, unstressed = 78 ms). Unexpectedly, no effect of Gemination was found. However, these results tested absolute duration: it will be interested to use relative measurements (e.g., singleton/geminate ratio).

Our results clearly show, at least for rhotic duration, the presence of gemination in the Roman Regional Italian; the unexpected results for vowel durations let, instead, hypothesize a different role of vowel duration in conveying gemination (see also Mairano, Di Iacovo, 2018 for similar inconsistencies). Qualitative analysis of the distribution of rhotics across the two conditions show a more complex picture. Intervocalic geminate rhotics seem to allow a greater range of possibilities: they can indeed be realized mainly as a trill, but approximant, fricatives, taps, and combined realizations (trill or tap with a fricative appendix) are found too. When looking at each speaker production, results show that some speaker seem to be more systematic in realising their geminates with lenited variants. These results show that relying only on quantitative analysis for investigating rhotic degemination can lead to misleading conclusions. The observation of lenited variants in the speech of some speakers show that is not only duration, but the use of tap variants that can have indexical purposes, since it is characterized by “statistical differences in [the] form’s distribution” (Foulkes, Scobbie & Watt, 2010). Rhotic degemination seems therefore to be, at least for some speakers, still present: from a sociophonetic perspective this could be interpreted as the maintenance of a marked feature to

show affiliation to a cultural-defined group, also given the fact that the recordings have been performed by a member of the same community. A further analysis of this phenomenon in casual speech in different communicative settings will help confirming the findings presented in this first sociophonetic analysis of rhotic degemination.

### *Riferimenti bibliografici*

- BERNHARD, G. (1998). *Das Romanesco des ausgehenden 20. Jahrhunderts: variationslinguistische Untersuchungen*. Tübingen: Niemeyer.
- 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.
- D'ACHILLE, P. (1995). L'italiano de Roma. In *Italiano & oltre*, 10, 38-43.
- D'ACHILLE, P., GIOVANARDI, C. (1995). Romanesco, neoromanesco o romanaccio? La lingua di Roma alle soglie del Duemila. In M. T. ROMANELLO, I. TEMPESTA (eds.), *Dialetti e Lingue nazionali. Atti del XXVII Congresso Internazionale di Studi della Società di Linguistica Italiana (Lecce, 28-30 ottobre 1993)*. Roma: Bulzoni, 397-412.
- D'ACHILLE, P., STEFINLONGO, A. & BOCCAFURNI, A. M. (2012). *Lasciatece parlà. Il romanesco nell'Italia di oggi*. Roma: Carocci.
- Foulkes, P., Scobbie, J.M., & Watt, D. (2010). Sociophonetics. In W.J. HARDCASTLE, J. LAVERAND & F. E. GIBBONS (eds.), *Handbook of Phonetic Sciences (2nd ed.)*. London: Blackwell, 703–716.
- JOHNSTONE, B., ANDRUS, J. & DANIELSON, A.E. (2006). Mobility, indexicality, and the enregisterment of “Pittsburghese”. In *Journal of English linguistics*, 34, 2, 77-104.
- DMITRIEVA, O., CELATA, C., MELUZZI, C. & CONCU, V. (2018). *The effects of lexical status and lexical competitors on the production of Italian stops*. Poster presented at the 16th Labphon conference, Lisbon (Portugal) 19-22 June 2018
- MAIRANO, P., DE IACOVO, V. (2018). La geminazione nell'italiano regionale del Piemonte. Presentation for *La variazione fonetica e fonologica dell'italiano: dati, metodi e modelli*. Bern (Switzerland), 7 September 2018.
- MAROTTA, G. (2005). Il consonantismo romano. Processi fonologici e aspetti acustici. In ALBANO LEONI, F., GIORDANO, R. (eds.), *Italiano parlato. Analisi di un dialogo*. Napoli: Liguori, 1-24.
- PAYNE, E. (2005). Phonetic variation in Italian consonant gemination. In *Journal of the International Phonetic Association* 35, 2, 153-181.
- ROHLFS, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino, Einaudi.
- SCOBIE, J. (2006). (R) as a variable. In K. BROWN (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed.). Oxford: Elsevier, 337–344.
- STEFINLONGO, A. (1985). Note sulla situazione sociolinguistica romana. In *Rivista italiana di dialettologia*, 9, 43-67.

STEFINLONGO, A. (1999). “Neoromanizzazione” del territorio. La lingua delle scritte murali nell'area metropolitana romana. In M. DARDANO ET AL. (a cura di), *Roma e il suo territorio. Lingua, dialetto, società*. Roma: Bulzoni, 267-285.

CHIARA MELUZZI<sup>1</sup>, NICHOLAS NESE<sup>1</sup>

<sup>1</sup> Università degli Studi di Pavia

## Dental affricates variation in Italian regional varieties: introducing the ZIta corpus

**Introduction:** Dental affricates are one of the most marked phonemes and Italian is one of the 30 languages considered by Ladefoged, Maddieson (1996) to present both the voiced and the voiceless variants in its phonological repertoire. Various phonetic descriptions of these phonemes have been provided by different scholars (Canepari, 1979, Celata, 2004, Loporcaro, 2009), without the support of a stratified sociolinguistic sample. For this reason, project ZIta will represent a sociophonetic corpus of dental affricates: in this poster, we present the full project and some preliminary results.

**Aims:** ZIta aims at fulfilling three major objectives: (1) document synchronic and diachronic variation of Italian regional varieties; (2) offer a fine-grained stratified survey of dental affricates; (3) prepare an instrument for studies on acquisition of Italian L2, and for the logopedic treatments of speech pathologies, in order to distinguish sociophonetic variations from “errors”. Thus, ZIta is not intended for corpus linguistic analysis but for sociophonetic research, and it will indeed represent a small but well balanced corpus of dental affricates, stratified for phonological contexts, frequency of the words in spoken Italian, for style and for speakers’ age, sex, place of origin and dialectal background.

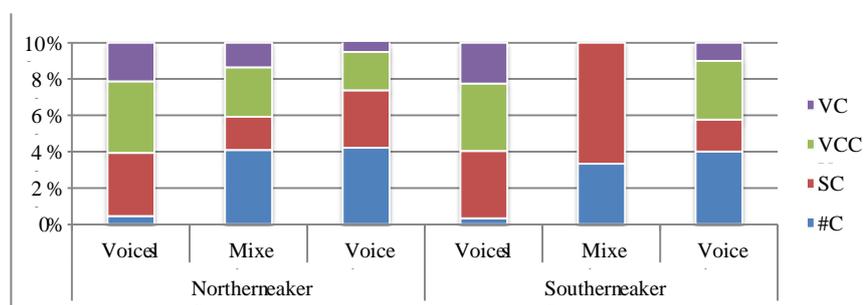
**Methods:** Controlled speech is elicited through sentence reading, whereas map-tasks are used for semispontaneous speech. Each task consists of 54 tokens of real words containing a dental affricate equally distributed in four phonological contexts, that is word-initial (#C-), post-sonorant (SCV), and intervocalic singleton (VCV) and geminate (VCCV). Tokens are annotated in PRAAT following the annotation protocol for dental affricates firstly introduced by Meluzzi (2014, 2016), who distinguishes among three sonority degrees (voiceless, voiced, and mixed), that is with voicing in the occlusion followed by a voiceless fricative segment (see also Nese, Meluzzi, 2018); the protocol annotates the two segments of the dental affricate (i.e., occlusive and fricative), also labelling the possible pause between the two parts (the so-called post-burst aperiodicity, cf. Foulkes et al., 2011).

**Data:** For this poster, we present the firsts results from the controlled speech of 4 male speakers, students of the University of Pavia, aged between 19 and 21, coming from two different linguistic areas, Sicily and Lombardy, whose dialects they know and speaks. We choose these groups as representatives of two very different Italian regional varieties, which present an opposite distribution of

sonority and place of articulation of dental affricates especially in the intervocalic and post-sonorant contexts. This first sample consists of 216 tokens, which have been annotated in PRAAT; two PRAAT scripts were then used to automatically extract the duration of the segments, including both the occlusive and the fricative parts, and the values of the center of gravity as measured on the central point of the fricative segment, with a FFT spectrum (pre-emphasis = 6 db) on a 20ms Hamming window (Jongmann et al., 2000). Other measures will be considered, but not yet included in this poster.

**First results:** The realization of dental affricates in the controlled speech of the 4 male subjects has shown differences with respect to the origin of the speaker but also of the phonotactic contexts. Data in Fig. 1 highlights how Sicilian speakers prefer mixed affricates in the post-sonorant context (SCV, 66,7%), whereas Northern speakers present the mixed variant in word-initial position (#C-, 40,9%); moreover, the geminate intervocalic realization is more likely to be realized as voiced by Southern speakers and as voiceless by Northern speakers.

Fig. 1 - Variation in the realization of the sonority of the dental affricate for phonological contexts and origin of the speaker (Northern s.:  $\chi^2(6)=24,146$ ,  $p<,01$ , Cramer's  $V=,336$ ; for Southern s.:  $\chi^2(6)=29,612$ ,  $p<,001$ , Cramer's  $V=,370$ )



A t-test independent analysis has been carried out on the mean values of affricate, stop and fricative duration, and on the mean values of the center of gravity. Because of the limited sample size, it was impossible to calculate these values for the mixed affricates. For voiced affricates, the means of duration and center of gravity were statically significant only in word-initial context, with the Northern speakers showing a more advanced place of articulation of the fricative segment (MD= 904,17 Hz,  $t(22)=2,398$ ,  $p=,025$ , Cohen's  $d=1,02$ ) and with slightly longer realization of the voiced dental affricate (MD= 36,137 msec,  $t(22)=3,210$ ,  $p=,004$ , Cohen's  $d=1,3$ ). Conversely, the center of gravity shows no variation for the voiceless affricates.

**Conclusions and further perspectives:** The preliminary results have illustrated the fine-grained variation in the realization of dental affricates, which

ZIIta aims at documenting on many other Italian regional varieties, by also taking into account different sociolinguistic variables.

### *Riferimenti bibliografici*

- CANEPARI, L. (1979). *Italiano standard e pronunce regionali*. Padova: Cleup.
- CELATA, C. (2004). *Acquisizione e mutamento di categorie fonologiche. Le affricate in italiano*. Pavia: Franco Angeli.
- FOULKES, P., DOCHERTY, G. & JONES, M. J. (2011). Analyzing stops. In M. Di Paolo, M. Yaeger-Dror (eds.), *Sociophonetics: A Student's guide*. London: Routledge, 58-71.
- JONGMANN, A., WAYLAND, R. & WONG, S. (2000). Acoustic characteristics of English fricatives. In *JASA* 108: 1252-1263.
- LADEFOGED, P., MADDIESON, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- LOPORCARO, M. (2009). *Profilo linguistico dei dialetti italiani*. Roma/Bari: Laterza.
- MELUZZI, C. (2014). *Le affricate dentali nell'italiano di Bolzano. Un approccio sociofonetico*. Università di Pavia / Libera Università di Bolzano, Tesi di dottorato in Linguistica.
- MELUZZI, C. (2016). A New Sonority Degree in the Realization of Dental Affricates /ts dz/ in Italian. In M. J. BALL, N. MÜLLER (eds.), *Challenging Sonority - Cross-linguistic Evidence*. London: Equinox Publishing, 252-275.
- NESE, N., MELUZZI, C. (2018), Accomodamento ed emergenza di varianti fonetiche: le affricate dentali a Pavia e Bolzano. In C. BERTINI, C. CELATA, G. LENOCI, C. MELUZZI & I. RICCI (a cura di), *Fattori sociali e biologici nella variazione fonetica / Social and Biological Factors in Speech Variation*. Milano: OfficinaVentuno, 69-86.

VINCENZO GALATÀ<sup>1,2</sup>, LORENZO SPREAFICO<sup>3</sup>, ALESSANDRO VIETTI<sup>3</sup>

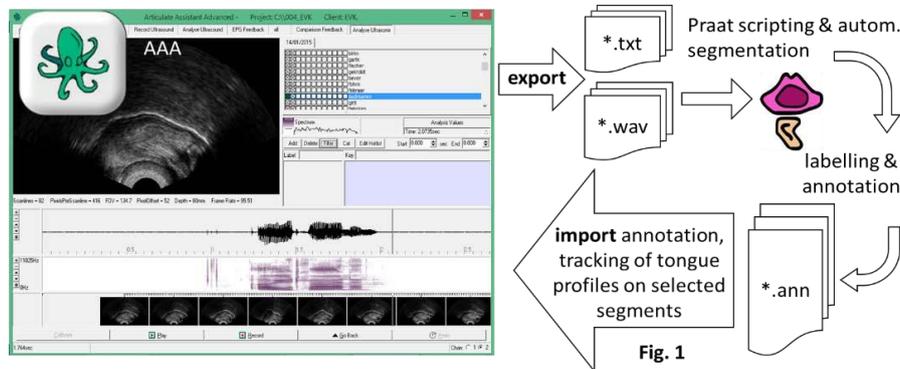
<sup>1</sup> Università degli Studi di Siena, <sup>2</sup> Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, <sup>3</sup> Libera Università di Bolzano

## **Costruzione e manutenzione di un archivio fonoarticolatorio**

Quale che sia l'intenzione per cui sono stati elaborati, gli archivi orali sono depositi e custodie di diversità e varietà linguistica. Fortunatamente, negli ultimi decenni la consapevolezza di quali siano le migliori pratiche per la costruzione di archivi orali è andata via via crescendo e ha riguardato tanto le tematiche della registrazione, organizzazione e conservazione dei dati, quanto quelle del loro trattamento per finalità analitiche. Tuttavia, va registrato come ancora oggi la quasi totalità delle iniziative sia dedicata agli archivi fonici, mentre poco spazio viene riservato - anche nella cerchia specialistica - alla costruzione di archivi fonoarticolatori che includano informazioni anche sui processi di produzione dei suoni operati dai parlanti (Marchal, Hardcastle, 1993; Westbury, 1994; Wrench, Hardcastle, 2000; Richmond et al., 2011; Narayanan et al., 2011; Canevari et al., 2015). Questa mancanza è tanto più grave quanto più si consideri che un numero crescente di studi ha messo ben in evidenza come i dati articolatori non solo permettano di fare dell'analisi sulla variazione linguistica socialmente strutturata, ma anche permettano di rilevare fenomeni linguisticamente e sociolinguisticamente rilevanti non ancora o non più rilevabili dall'osservazione del solo evento acustico (Lawson et al., 2008; Lawson et al., 2014). Per questo motivo nel nostro contributo presentiamo e discutiamo alcuni degli aspetti rilevanti del processo di costruzione di una base di dati fonoarticolatori che includano registrazioni acustiche e ultrasonografiche delle produzioni di parlanti bilingui, nonché la loro annotazione.

Per quanto riguarda la raccolta dei dati, le registrazioni acustiche (22050 Hz, 16 bit mono) sono state acquisite in cabina silente ricorrendo a un microfono Sennheiser ME2, collegato ad un registratore B1 Marantz PMD660. Le registrazioni articolatorie sono state effettuate ricorrendo a un ecografo Ultrasonix SonixTablet (Exam software v. 5.7.3) associato ad un trasduttore Ultrasonix C9-5/10. La sincronizzazione dei dati acustici e articolatori è stata attuata ricorrendo al software dedicato Articulate Assistant Advanced v. 2.15 (AAA) della Articulate Instruments Ltd (2015). La soluzione commerciale cui abbiamo fatto ricorso ha permesso di semplificare i problemi connessi con la registrazione dei dati acustici e articolatori, ma non ha reso disponibili - se non in parte - le procedure per la preparazione e l'elaborazione dei dati raccolti. Per tale motivo il gruppo di ricerca ha sviluppato un protocollo di lavoro basato sulla

esportazione indipendente dei dati acustici e articolatori, la loro annotazione indipendente e, infine, il loro riaccoppiamento (Fig. 1).



Per quanto riguarda l'annotazione dei dati acustici - che ha riguardato tanto produzioni in italiano quanto produzioni in tirolese (un dialetto bavarese meridionale) - è stata predisposta una procedura di allineamento forzato tramite lo strumento WebMAUS (Schiel, 1999; Kisler et al., 2012). Poiché il bavarese è una lingua minimamente documentata per cui non sono disponibili né raccolte orali sistematiche, né strumenti di trattamento linguistico automatico, per l'allineamento forzato si è fatto ricorso alla modalità dipendente dalla lingua tedesca. Questo passaggio di elaborazione ha prodotto un PRAAT TextGrid con una trascrizione ortografica a livello di parola e una trascrizione in SAMPA a livello di fonemi con il rispettivo allineamento al segnale per ogni parola di destinazione in ciascun file.

Le annotazioni acustiche così create sono poi state importate e accoppiate ai dati articolatori per mezzo dello stesso software AAA così da permettere l'analisi parallela o indipendente delle due dimensioni.

Per quanto riguarda l'annotazione dei dati articolatori, l'operazione è stata effettuata in maniera semiautomatica ricorrendo anzitutto ad una funzione di tracciatura automatica dei profili, seguita da una di correzione manuale dei profili. Anche in questo caso l'operazione ha comportato la risoluzione di problemi ricorrenti in ogni progetto di costruzione di banche dati di parlato quali anzitutto le dimensioni della coerenza (*consistency*) e la sensibilità (*delicacy*) degli annotatori coinvolti nel processo.

Dei problemi connessi con la stesura di un protocollo di preparazione dei dati e delle soluzioni adottate si discuterà criticamente, evidenziando come la definizione e l'utilizzo di procedure semi-automatiche siano in grado di ridurre in primis il carico di lavoro e non ultimo l'errore umano implicito in ciascuna operazione manuale.

### *Riferimenti bibliografici*

- ARTICULATE INSTRUMENTS LTD. (2015). *Articulate Assistant Advanced User Guide: Version 2.15*. Edinburgh, UK: Articulate Instruments Ltd.
- CANEVARI, C., BADINO, L. & FADIGA, L. (2015). A new Italian dataset of parallel acoustic and articulatory data. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (Vol. 2015–Janua). 2152–2156.
- KISLER, T., SCHIEL, F. & SLÖETJES, H. (2012). Signal processing via web services: the use case WebMAUS. In *Proceedings Digital Humanities 2012*. Hamburg, 30–34.
- LAWSON, E., STUART-SMITH, J. & SCOBBIIE, J. M. (2014). A mimicry study of adaptation towards socially-salient tongue shape variants. *University of Pennsylvania Working Papers in Linguistics*, 20(2), 101–110.
- MARCHAL, A. & HARDCASTLE, W. J. (1993). ACCOR: instrumentation and database for cross-language study of coarticulation. *Language and Speech*, 36(2, 3), 137–153.
- NARAYAN, S., BRESCH, E., GHOSH, P., GOLDSTEIN, L., KATSAMANIS, A., KIM, Y., ... ZHU, Y. (2011). *A Multimodal Real-Time MRI Articulatory Corpus for Speech Research*. Retrieved from <http://sail.usc.edu/span/>
- RICHMOND, K., HOOLE, P. & KING, S. (2011). *Announcing the Electromagnetic Articulography (Day 1) Subset of the mngu0 Articulatory Corpus*. Retrieved from <https://pdfs.semanticscholar.org/6406/34dd9cb18155e012a550270fd940627a2b2e.pdf>
- SCHIEL, F. (1999). Automatic phonetic transcription of nonprompted speech. In *Proceedings of ICPbS*. 607– 610.
- WESTBURY, J. (1994). *X-ray microbeam speech production database*.
- WRENCH, A. A., HARDCASTLE, W. J. (2000). A Multichannel Articulatory Database and its Application for Automatic Speech Recognition. In *Proceedings on the 5th Seminar of Speech Production*. 305–308.

LORENZO FILIPPONIO<sup>1</sup>, DAVIDE GARASSINO<sup>2</sup>, DALILA DIPINO<sup>2</sup>

<sup>1</sup> Humboldt-Universität zu Berlin, <sup>2</sup> Universität Zürich

## Tra fonetica e tipologia: la durata consonantica in due dialetti italo-romanzi settentrionali

È fatto noto che a nord della linea *Carrara-Senigallia* (più nota col nome di *La Spezia-Rimini*) il consonantismo sia caratterizzato principalmente dai fenomeni della lenizione e della degeminazione fonologiche, manifestatisi in quest'ordine cronologico.

In molte varietà, anche in termini meramente fonetici, la degeminazione è processo completo in protonia e in postonia. Le misurazioni di Uguzzoni e Busà (1995) per il frignanese, varietà emiliana con quantità vocalica distintiva (sulla cui insorgenza cfr. Filipponio, 2012a, Loporcaro, 2015), mostrano per le consonanti postoniche di bisillabi elicitati in contesto di frase cornice durate pressoché identiche tra scempie e geminate etimologiche. Altra è la situazione nei monosillabi (per apocope di vocale atona finale), dove si ravvisano delle lievi differenze: in virtù di questo, le Autrici mettono in guardia dal considerare soltanto la «media generale» delle misure di C postonica, che «appiattisce una realtà variegata», in cui «emerge una variabilità che dipende dai soggetti, dai tipi di vocale tonica e dalle strutture della parola» (Uguzzoni, Busà 1995: 14-15). Quanto mostrato rappresenta un esempio di come, in varietà a quantità vocalica distintiva, la realizzazione della durata consonantica in postonia, pur priva di rilevanza fonologica, possa essere soggetta a variazione.

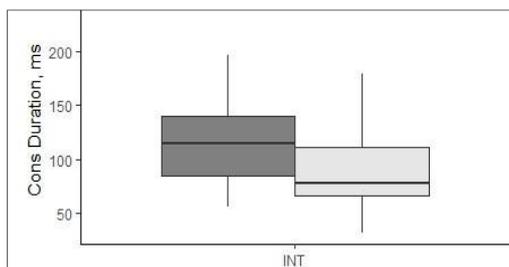
Prescindendo ora dal singolo soggetto parlante, di questa variazione sono individuabili per lo meno tre assi:

1. diatopico: la durata delle consonanti geminate etimologiche (o dopo vocale fonologicamente breve) può essere generalmente maggiore di quella delle scempie etimologiche (o dopo vocale lunga), come p.es. nella varietà emiliana appenninica di Lizzano in Belvedere (cfr. Filipponio, 2012a).
2. segmentale: la realizzazione fonetica della degeminazione può differire a seconda del tipo di consonante, come nei dialetti della Lunigiana e della Garfagnana, dove solo tra le ostruenti permangono differenze di durata (cfr. Filipponio, Pedrazzini, in preparazione).
3. (prosodico-)strutturale: la durata delle consonanti postoniche può dipendere dalla posizione all'interno di parola, come avviene anche nel caso del già citato frignanese, in cui nei monosillabi essa è maggiore che nei bisillabi – oltre a differire in base alla lunghezza della vocale precedente (v. punto 1).

Questi assi di variazione non si escludono mutualmente, ma si incrociano ortogonalmente, dando vita a una casistica la cui classificazione è lungi dall'essere compiuta e che dovrebbe essere inscritta nel quadro tipologico più generale del rapporto fra durata vocalica e consonantica (cfr., fra gli altri, Ridouane, 2010, Idemaru, Guion-Anderson, 2010 e Klingler et al., 2018).

Date tali premesse, intendiamo occuparci in questo contributo di due varietà, ovvero il dialetto emiliano orientale di Bologna e quello ligure occidentale di Porto Maurizio (portorino), entrambe con opposizioni di quantità vocalica in parossitonia e in ossitonia, in cui vi sono tracce di una diversa realizzazione delle consonanti postoniche dopo vocale breve rispetto a quelle dopo vocale lunga. Nel primo, è stata rilevata da più parti, a partire da Coco (1970, 1971), una tendenza alla geminazione allofonica delle consonanti che seguono vocale tonica breve (v. anche Loporcaro, 2015: 88), ravvisata da ultimo da Avesani et al. (2016). Per il secondo, le misurazioni di Filipponio e Garassino (in stampa) mostrano una notevole asimmetria nella durata consonantica dopo vocale lunga e dopo vocale breve:

**Figura 1.** Durata delle consonanti postoniche in portorino dopo vocale breve (sinistra) e lunga (destra) in posizione interna di frase.



Non può però passare inosservato che questa apparente corrispondenza tra bolognese e portorino è calata nelle due varietà in strutture di parola radicalmente differenti. Il bolognese è infatti, tra i dialetti gallo-italici, la punta avanzata delle varietà a *compensazione* (cfr. Bertinetto, Bertini, 2010, Filipponio, 2012b), ovvero quelle in cui lo squilibrio di peso metrico-prosodico tra sillaba tonica e sillabe atone all'interno della parola fonologica, ascrivibile in prima istanza all'intensità dell'accento, ha portato a una condizione di quasi totale caduta del vocalismo atono (si pensi a sviluppi come PÖRTĪCUM > /'po:rdg/, OSPĪTALEM > /'zbdε:l/ ecc.). In questo quadro, la lunghezza allofonica delle consonanti potrebbe essere considerata una conseguenza dello stretto legame che viene a instaurarsi tra una vocale tonica fonologicamente (e foneticamente) breve emessa con notevole intensità e il fono seguente, nei termini individuati da Martinet (1966) sotto l'etichetta di *close contact* (cfr. Uguzzoni et al., 2003).

Il portorino, invece, come tutte le varietà liguri, si contraddistingue per la conservazione del vocalismo atono, che, in termini ritmici, deve essere considerata conseguenza di una tendenza alla compensazione manifestatasi con intensità nettamente minore rispetto al bolognese (e infatti PÖRTĪCUM > /'pɔrtegu/, ecc.). Data questa situazione, è forse meno probabile che una geminazione allofonica vada ascritta a un *close contact* indotto dalla struttura ritmica; sarà piuttosto da considerare come mero tratto residuale di conservazione o effetto della pressione dell'italiano, peraltro difficile da valutare: infatti, almeno per questo livello strutturale di analisi, il fatto che certe caratteristiche possano proiettarsi sui corrispondenti italiani regionali lascia aperte molte questioni riguardanti i processi di convergenza verticale che stanno portando alla sostituzione di alcuni tratti di sistema dei dialetti italo-romanzi con quelli della lingua tetto (cfr. Cerruti et al., 2017: 6), tali da rendere la valutazione dei tratti intercorrenti tra i due codici estremamente delicata.

L'analisi qui proposta mira quindi a sottolineare la necessità di considerare i fenomeni segmentali all'interno del complesso strutturale che caratterizza un sistema. La verifica sperimentale sarà condotta sulla base di questionari composti da frasi di senso compiuto, in contesto pragmatico non marcato, con parola bersaglio in posizione interna. Le parole target comprendono coppie (semi)minime, distinte per la quantità vocalica, nelle quali la vocale tonica è seguita da diversi tipi di consonanti: occlusive (ad es., portorino ['brytu], 'brutto' ~ ['fry:tu], 'frutto'; bolognese ['tropa], 'truppa' ~ ['tro:pa], 'troppa'), fricative (port. ['mazu], 'maggio' ~ ['na:zu]; bol. ['masa], 'messa' ~ ['ma:sa] 'massa') e sonoranti (port. ['sene], 'cenere' ~ ['se:ne], 'cene'; bol. ['bala], 'bolla' ~ ['ba:la], 'palla').

I questionari saranno sottoposti a dieci informanti (cinque per ciascuna varietà), che tradurranno a voce in dialetto le frasi lette loro in italiano. Gli obiettivi della nostra indagine sono:

- i verificare la consistenza della geminazione postonica ed eventuali differenze determinate dai foni implicati o dalla struttura di parola. Verranno per questo misurate le lunghezze di V tonica e di C postonica, nonché il rapporto V/C, rilevante per il riconoscimento sintagmatico della quantità vocalica da parte dei parlanti, perché indipendente dalla velocità d'eloquio (cfr. Filipponio, 2012a);
- ii proporre alcune strategie sperimentali per verificare le ipotesi sopra formulate in merito alle cause soggiacenti all'analoga (in superficie) geminazione allofonica delle consonanti postoniche nei due dialetti, come l'analisi dei profili di intensità delle vocali toniche. L'ipotesi è che in una varietà in cui il vocalismo atono è stato sottoposto a pesanti riduzioni (bolognese) i valori di intensità delle vocali toniche siano in media più robusti rispetto ad un dialetto che conserva tendenzialmente le atone finali (portorino). Una differenza significativa in tal senso porterebbe prove

all'ipotesi di un effetto del *close contact* per spiegare la geminazione allofonica del bolognese.

I risultati dei test sperimentali sopra descritti ci permetteranno di riflettere, in termini tipologici, sulle differenti manifestazioni della quantità vocalica distintiva nei dialetti italo-romanzi settentrionali, sulla loro struttura di parola, nonché sulle possibili motivazioni che hanno portato alla persistenza o all'insorgenza di una parziale geminazione allofonica delle consonanti postoniche nelle due varietà.

### *Riferimenti bibliografici*

- 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 R. SAVY, I. ALFANO (eds.), *La fonetica nell'apprendimento delle lingue*. Milano: Officinaventuno, 391-410.
- BERTINETTO, P.M., BERTINI, C. (2010). Towards a unified predictive model of natural language rhythm. In M. RUSSO (ed.), *Prosodic Universals. Comparative studies in rhythmic modeling and rhythm typology*. Roma: Aracne, 43-77.
- CERRUTI, M., CROCCO, C. & MARZO, S. (2017). On the development of a new standard norm in Italian. In M. CERRUTI, C. CROCCO & S. MARZO (eds.), *Towards a New Standard. Theoretical and Empirical Studies on the Restandardization of Italian*. Berlin-New York: de Gruyter Mouton, 3-28.
- COCO, F. (1970). *Il dialetto di Bologna. Fonetica storica e analisi strutturale*. Bologna: Arnaldo Forni Editore.
- 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.
- FILIPPONIO, L. (2012a). *La struttura di parola dei dialetti della Valle del Reno. Profilo storico e analisi sperimentale*. Bologna: Arnaldo Forni Editore.
- FILIPPONIO, L. (2012b). Reflexionen über die Diachronie des Galloitalienischen Rhythmus. In E. SCHAFROTH, M. SELIG (eds.), *Testo e ritmi. Zum Rhythmus in der italienischen Sprache*. Frankfurt: Peter Lang, 61-87.
- FILIPPONIO, L., GARASSINO, D. (submitted). Center and Peripheries in Phonology. A stress-test for two Ligurian dialects. In *Italian Journal of Linguistics*.
- FILIPPONIO, L., PEDRAZZINI, L. (in prep.). *La progressione della degeminazione sul versante occidentale della linea Carrara-Senigallia*.
- IDEMARU, K., GUIONO-ANDERSON, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. In *Phonetica* 67(1), 25-46.
- KLINGER, N., MOOSMÜLLER, S. & SCHEUTZ, H. (2018). Vowel and Consonant Sequences in three Bavarian dialects of Austria. In *Interspeech 2017*, 2983-2987.
- LOPORCARO, M. (2015). *Vowel Length from Latin to Romance*. Oxford: Oxford University Press.
- MARTINET, A. (1966). Close Contact. In *Word* 22:1-6.

RIDOUANE, R. 2010, Gemimates at the junction of phonetics and phonology. In *Laboratory phonology* 10, 61-90.

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 M.J. SOLÉ, D. RECASENS & J. ROMERO (eds.), *Proceedings of the XVth International Congress of Phonetic Sciences, Barcelona, 3-9 August 2003*. Univ. Autònoma de Barcelona, III, 2717-2720.

DANIELE SANTINI<sup>1</sup>, GIOVANNA MAROTTA<sup>1</sup>, GIOVANNA LENOCI<sup>2</sup>, PIER MARCO BERTINETTO<sup>2</sup>

<sup>1</sup> Università di Pisa, <sup>2</sup> Scuola Normale Superiore

## Vocali ripetute in italiano. Uno studio esplorativo

Questo studio sperimentale prende in esame la produzione di sequenze di vocali identiche ripetute, che si possono trovare nella radice di lessemi (cf. *Sahara*, *coorte*), ovvero su confine morfologico (cf. *atrii*, *atee*), ovvero ancora in sequenze parimenti prodotte da processi morfologici che hanno per dato adito ad una piena lessicalizzazione (cf. *presistente*, *coordinare*).

L'effettiva pronuncia di queste sequenze è variabile, e certamente lo è anche in rapporto a fattori di natura strettamente lessicale. Si consideri infatti che mentre *coorte* si oppone a *corte*, *coordinare* non si oppone a *\*cordinare*. Uno stesso individuo può comunque assumere comportamenti diversi a seconda dello stile adottato, nonché della contingente velocità di elocuzione. Il nostro studio non intende tuttavia affrontare la questione da tale punto di vista, ma si propone piuttosto di avviare una ricognizione sulla dinamica articolatoria che sottende alla realizzazione accurata (benché non iperarticolata) di tali sequenze. L'ipotesi sperimentale, cui si è provato a dare risposta mediante l'uso della tecnica ecografica (UTI), concerne la possibile esistenza di un incipiente gesto di riarticolazione fra la prima e la seconda vocale. Merita sottolineare che si tratta del primo studio condotto con tecnica ecografica sull'argomento.

L'analisi ha riguardato un corpus di 13 coppie semiminime, per un totale di 26 parole (registrate in isolamento o in opportune frasi) pronunciate da 4 parlanti settentrionali. La scelta dei locutori è stata fatta per evitare il più possibile che si creassero differenze timbriche negli incontri di vocali di altezza intermedia, le quali sarebbero *ipso facto* correlate a differenze articolatorie. A ciascuna parola del corpus corrispondeva analogo parola contenente una singola vocale nel punto critico (e dunque, in pronuncia accurata, una sillaba in meno): per es., *coorte* ~ *corte*.

Un primo aspetto preso in considerazione riguarda il confronto fra le durate delle correlative sequenze /VV/ ~ /V/. Da ciò è emerso che, nel corpus utilizzato, le vocali ripetute avevano una durata più che doppia rispetto alle corrispondenti semplici (indipendentemente dal timbro), a dimostrazione di una pronuncia effettivamente accurata.

Sul piano articolatorio, l'indice primario preso in considerazione concerne il confronto fra la relativa perifericità apicale, dorsale e radicale dei due profili linguali (vocali ripetute ~ vocali singole). Benché non sia emerso un comportamento rigorosamente costante, si è osservata una maggiore perifericità

delle vocali singole rispetto alle vocali ripetute, soprattutto nella fase iniziale dell'intervallo vocalico. Rispetto al contesto in cui i segmenti critici sono collocati, ed in particolare rispetto al confronto fra posizione interna ~ finale di parola, è emersa in generale una maggior stabilità articolatoria delle vocali ripetute, che (verosimilmente a causa della loro maggior durata complessiva) sembrano meno condizionate dagli effetti coarticolatori. Non sono invece emerse significative differenze in merito alla diversa posizione dell'accento sulle due vocali ripetute (cf. *dée* ~ *coórte* ~ *pátrii*), anche se l'esiguo numero dei casi confrontati induce a considerare con debita cautela tale risultato.

In sintesi, riteniamo che i dati raccolti, per quanto preliminari, indichino che, in caso di iato tra vocali dello stesso timbro, si attui (quanto meno in un parlato accurato) una riarticolazione del segmento, che comporta specifici profili articolatori ed adeguate lunghezze.

### *Riferimenti bibliografici*

- BERTINETTO, P.M. (1981). *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: 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.
- GARRAPA, L. (2007). Italian vowel deletion across word boundaries: phonology or morphology?. In A. LAHIRI, J. MEINSCHAEFER & C. SCHWARZE (eds.), *Documentation of the Workshop 'Formal and semantic constraints in morphology'*. Konstanz, 61-84.
- MADDIESON, I., EMMOREY, K. (1985). Relationship between semi-vowels and vowels, crosslinguistic investigations of acoustic difference and coarticulation. In *Phonetica*, 42(4), 163-174.
- 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, Classe di Lettere e Filosofia*, 17(3), serie iii, 847-887.
- MAROTTA, G. (1988). The Italian diphthongs and the autosegmental framework. In P.M. BERTINETTO, M. LOPORCARO (eds), *Certamen Phonologicum. Papers from the 1987 Cortona phonology meeting*. Torino: Rosenberg & Sellier, 389-420.
- MAROTTA, G. (2010a). Dittongo. In *Enciclopedia dell'italiano online*: [http://www.treccani.it/enciclopedia/dittongo\\_\(Enciclopedia-dell'Italiano\)\\_\\_\\_\\_\\_](http://www.treccani.it/enciclopedia/dittongo_(Enciclopedia-dell'Italiano)_____) (ultimo accesso: 14/05/18).
- MAROTTA, G. (2010b). Iato. In *Enciclopedia dell'italiano online*: [http://www.treccani.it/enciclopedia/iato\\_\(Enciclopedia-dell'Italiano\)/](http://www.treccani.it/enciclopedia/iato_(Enciclopedia-dell'Italiano)/) (ultimo accesso: 14/05/18).
- MAROTTA, G., SORIANELLO P. (1997). Vocali contigue a confine di parola. In P.M. BERTINETTO, L. CIONI (a cura di), *Unità fonetiche e fonologiche. Produzione e percezione, Atti delle VIII giornate di studio del Gruppo di fonetica sperimentale (ALA) (Pisa, 18-20 dicembre 1997)*. Pisa: Scuola Normale Superiore, 101-113.

- MIONI, A. M. (1986). Fonetica articolatoria: descrizione e trascrizione degli atteggiamenti articolatori. In L. CROATTO (a cura di), *Trattato di foniatría e logopedia. III. Aspetti fonetici della comunicazione*. Padova: La Garangola, 15-87.
- ROMEO, L. (1968). A phonemic inventory of the Italian bivocalic sequences. In *Forum Italicum*, 2, 117-143.
- 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. (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.
- SALZA, P.L., MAROTTA G. & RICCA D. (1987). Duration and formant of Italian bivocalic sequences. In *Proceedings of the Eleventh international congress of phonetic sciences (ICPhS) (August 1-7, 1987, Tallinn, Estonia)*. Tallinn: Academy of Sciences of the Estonian SSR, 6 voll., vol. 3°, 113-116.

PHILIPPE BOULA DE MAREÛIL<sup>1</sup>, VALENTINA DE IACOVO<sup>2</sup>, ANTONIO ROMANO<sup>2</sup>, IOANA VASILESCU<sup>1</sup>, FRÉDÉRIC VERNIER<sup>1</sup>

<sup>1</sup> Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur-Centre National de la Recherche Scientifique & Université Paris-Saclay, <sup>2</sup> Laboratorio di Fonetica Sperimentale "Arturo Genre" – Università degli Studi di Torino

## **Verso un atlante sonoro integrato delle lingue e dei dialetti di Francia e d'Italia**

Descriviamo qui un atlante linguistico sonoro che assume la forma di un sito web composto dalle mappe interattive della Francia e dell'Italia, dove è possibile cliccare su più di 300 punti di indagine per ascoltare diversi campioni di parlato e leggere una trascrizione di ciò che viene detto in lingue locali (o regionali). Benché in origine l'operazione sia nata separatamente nei due laboratori coinvolti e abbia condotto a versioni parziali pubblicate in sedi distinte (Boula de Mareuil, 2017; De Iacovo & Romano, 2017; Romano, 2016), un'intesa si è stabilita progressivamente in vista della fusione di tutti i campioni di parlato raccolti in un unico archivio di cui si dà qui una prima descrizione.

Abbiamo registrato la favola di Esopo "Il vento di tramontana e il sole" (utilizzata dall'Associazione Fonetica Internazionale per illustrare molte lingue del mondo) in una ventina di varietà gallo-romanze e una ventina di varietà italo-romanze, in catalano, in sardo, in friulano e in alcune varietà non-romanze come le seguenti: basco, bretone, fiammingo occidentale, francone e alsaziano in Francia, griko, arbëresh e walser in Italia. Le lingue o dialetti gallo-romanzi di cui abbiamo almeno una registrazione trascritta sono i seguenti: piccardo, normando, mainiot, gallo, angioino, pittavino-santongese, borgognone, francoconteese, lorenese, champenois, vallone, berricione-borbone, francoprovenzale, guascone, linguadociano, provenzale, limosino, alverniate e vivaro-alpino. Quelli italo-romanzi sono i seguenti: piemontese, ligure, lombardo, veneto, emiliano-romagnolo, toscano, marchigiano, molisano, pugliese, salentino, calabrese, siciliano, lucano, campano, romanesco, gallurese e còrso. Le trascrizioni ortografiche sono state fornite dai locutori stessi e omologate da linguisti. Questi dati formano così un'istantanea sonora che testimonia l'utilizzo attuale di alcune varietà linguistiche che, pur essendo minacciate di estinzione, costituiscono ancora oggi una preziosa realtà.

La base di dati è in continua espansione, si prevede quindi di aumentare il numero di locutori per ciascun punto d'inchiesta, mentre le informazioni sociolinguistiche sono disponibili nei riferimenti bibliografici dei singoli lavori di ricerca per i quali le registrazioni sono state effettuate. Parimenti, le caratteristiche

acustiche, nonché il protocollo seguito per la raccolta dei dati sonori, saranno consultabile in una sezione apposita del sito.

Oltre ai confini delle regioni amministrative, abbiamo incluso i confini tra i domini linguistici, ispirati dagli Atlas régionaux par région per la Francia, dalla classificazione di Pellegrini per l'Italia (Pellegrini, 1977), con una segnaletica speciale per le lingue non-romanze. In assenza di registrazioni in alcune aree — come in quella ladina o in quella dei dialetti italiani mediani, tra i quali l'umbro e il sabino meriterebbero più attenzioni —, non abbiamo (ancora) posto le etichette corrispondenti sulla mappa. Prevediamo di colmare le lacune eseguendo le registrazioni mancanti e lanciando nei prossimi mesi una campagna di sensibilizzazione verso gli utenti interessati alla variazione nello spazio romanzo, in modo da trovare in Italia (dove finora la pagina “tramontane” dell'università di Torino è stato esplorato da poche migliaia di visitatori) il successo incontrato (con oltre mezzo milione di visitatori) dall'Atlas sonore des langues régionales de France (Boula de Mareüil *et al.*, 2017). Questa piattaforma digitale permette quindi non soltanto la consultazione di dati sonori preziosi e non facilmente reperibili ma mira soprattutto alla formazione di un solido corpus sonoro con cui continuare ad analizzare fenomeni linguistici e dialettali dell'area romanza. Il sito per la Francia è accessibile all'indirizzo <https://atlas.limsi.fr>; quello per l'Italia su <https://atlas.limsi.fr/?tab=it>.

### *Riferimenti bibliografici*

BOULA DE MAREÜIL, P., VERNIER, F. & RILLIARD, A. (2017). Enregistrements et transcriptions pour un atlas sonore des langues régionales de France. In *Géolinguistique*, 17, 23–48 (disponibile all'indirizzo <https://atlas.limsi.fr>).

DE IACOVO, V., ROMANO, A. (2017). *Tramontane, Archivio on-line di più di 300 campioni sonori di parlato letto/narrativo in centinaia di lingue e dialetti* (disponibile all'indirizzo [http://www.lfsag.unito.it/ark/trm\\_index.html](http://www.lfsag.unito.it/ark/trm_index.html)).

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.

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

VIOLETTA CATALDO<sup>1</sup>, LOREDANA SCETTINO<sup>1</sup>, RENATA SAVY<sup>1</sup>,  
ISABELLA POGGI<sup>2</sup>, ANTONIO ORIGLIA<sup>3</sup>, ALESSANDRO ANSANI<sup>2</sup>, ISORA  
SESSA<sup>2</sup>, ALESSANDRA CHIERA<sup>2</sup>

<sup>1</sup> Università degli Studi di Salerno, <sup>2</sup> Università di RomaTre, <sup>3</sup> Università di Napoli  
“Federico II”

## **Tratti fonetici e funzionali delle pause, e gesti concomitanti, nel parlato delle guide turistiche**

### *Obiettivi e motivazioni del lavoro.*

Il contributo presenta una descrizione e analisi dei fenomeni di disfluenza nelle *performance* orali di guide turistiche durante l’accompagnamento delle visite museali.

Il tipo di parlato in oggetto costituisce un campo di indagine particolarmente interessante sotto l’aspetto delle disfluenze: accanto ai tipici fenomeni legati alla programmazione dell’enunciato, caratteristici dell’oralità *tout court*, infatti, lo stile d’eloquio della guida turistica presenta altri tratti legati ad una complessità di fattori, tra cui possiamo annoverare i seguenti:

- 1) funzione ‘retorica’ delle pause e delle ripetizioni: uso delle pause (parzialmente) consapevole e/o volontario, funzionale a catturare l’attenzione del pubblico, a sottolineare passaggi e termini salienti, a creare aspettativa;
- 2) carattere ‘recitativo’ del parlato: uso di una minima tecnica teatrale/recitativa in cui i fenomeni disfluenti si accompagnano e coordinano a gesti specifici, cinetica e mimica facciale;
- 3) tipologia ‘testuale’: testo descrittivo, parzialmente strutturato e ripetitivo, ricco di terminologia tecnica;
- 4) contesto di interazione: presenza di un pubblico eterogeneo, a volte distratto e ‘in movimento’;
- 5) contesto spaziale/situazionale: necessario ricorso alla deissi spaziale, sia verbale che gestuale, che di frequente interrompe il flusso dell’eloquio.

Per questi motivi, le ‘disfluenze’ della guida turistica assumono valori strutturali e funzionali, interessanti da indagare e da modellizzare.

Il lavoro si inserisce, infatti, nelle linee di ricerca del progetto CHROME - *Cultural Heritage Resources Orienting Multimodal Experience* (PRIN 2015) che ha come oggetto la definizione e sperimentazione di una metodologia di raccolta, analisi e modellizzazione di dati multimodali per la messa a punto di un “agente virtuale” in grado di presentare beni museali, attraverso un sistema di dialogo uomo-

macchina “antropomorfo”. La nostra ricerca ha, dunque, come scopo una migliore descrizione e modellizzazione di alcuni aspetti del comportamento verbale che contribuiscono alla naturalezza dell’eloquio e alla sua adeguatezza al contesto situazionale. In quest’ottica, fenomeni di ipospecificazione del significante fonico, di coarticolazione e di disfluenza vengono osservati sul versante ‘positivo’, alla ricerca di caratteristiche e regolarità implementabili per la sintesi “*text to speech*” (TTS) al fine di migliorarne le prestazioni.

All’analisi verbale si affianca l’osservazione puntuale degli elementi gestuali che cooccorrono con l’enunciazione. I gesti (in particolar modo quelli di tipo iconico e deittico) prodotti in concomitanza con le pause, infatti, oltre ad una funzione orientata alla produzione e all’interazione, possono svolgerne anche una orientata alla comprensione (Stam, Tellier, 2017), evidenziando e sottolineando, ad esempio, le parole/concetti più importanti all’interno del discorso, isolando parole chiave, guidando l’attenzione su elementi contestuali.

### **Corpus e metodologia**

Il materiale a disposizione consiste di registrazioni audio-visive di visite guidate condotte da guide esperte nel campo artistico-culturale. Il *corpus* completo è costituito da 3h30’ di parlato per ciascuna delle 3 guide (tutte donne), articolato in 6 porzioni che corrispondono a diversi ‘luoghi’ della visita (punti di interesse).

Per il presente lavoro, sono stati presi in considerazione 1 punto di interesse di tre guide (35’circa).

Sul materiale selezionato sono stati etichettati tutti i fenomeni di disfluenza, adottando un sistema di codifica che fa riferimento agli studi di modellizzazione di Shriberg (1994), Hiecke (1981) e Lickley (1998) adattato all’italiano parlato. Tale schema di annotazione si muove su quattro livelli (*Disfluency Type, Disfluency Function, Disfluency*

*Model, Disfluency Components*) che agiscono parallelamente su domini separati dell’occorrenza.

Nello studio qui proposto portiamo avanti un’analisi limitata alle pause, considerate secondo la seguente classificazione:

- pausa vuota (*silent pause*), distinta in pausa breve (*short pause*, <sp>) o pausa lunga (*long pause*, <lp>);
- pausa piena (*filled pause*, FP);
- pausa vocalizzata (*vocalized filled pause*, VFP), derivante da allungamento di parola.

I fenomeni disfluenti corrispondenti a queste categorie sono analizzati:

**A.** *sul piano fonetico acustico*, su cui sono state misurate:

- durata
- contenuto segmentale
- timbro vocalico
- profilo tonale (eventuale)
- livello di range (relativo rispetto alla sequenza adiacente)

**B.** *sul piano funzionale*, sul quale sono state individuate le seguenti funzioni primarie:

- *demarcativa*: distinguendo le pause fisiologiche da quelle di significato, legate alla scansione grammaticale/linguistica dell'enunciazione;
- *programmativa*: identificando le pause di pianificazione e/o recupero degli elementi lessicali operate dal parlante nel corso dell'enunciazione;
- *esitativa*: identificando le pause che manifestano un certo grado di incertezza nell'esposizione del parlante;
- *strategica*; identificando le pause utilizzate come strumento di controllo dell'interazione al fine di agevolare la produzione e allo stesso tempo la comprensione, permettendo ai parlanti di guadagnare tempo e di gestire l'attenzione dell'interlocutore.

C. *sul piano gestuale*, (specificamente i gesti delle mani) sul quale sono stati individuati:

- forma del gesto
- fasi di movimento (eventuale)
- ampiezza e direzione del gesto
- funzione del gesto
- sincronizzazione

Per l'analisi dei dati sul versante fonetico-acustico è stato utilizzato il software Praat (Boersma, Weenink, 2018), mentre l'etichettatura sia dei fenomeni di disfluenza, sia degli eventi gestuali è stata effettuata tramite ELAN (2018), che permette di visualizzare in sincrono le componenti audio e video del materiale raccolto per un'annotazione multilivello.

#### **Risultati (attesi)**

L'elaborazione dei dati, al momento della stesura di questo riassunto, è ancora in corso. Possiamo tuttavia riassumere i risultati preliminari più interessanti come segue:

- 1) la stragrande maggioranza (81%) delle pause di questo eloquio è costituito da pause piene; quasi la metà (45%) FP, il 36% VFP, da allungamento finale di parola; la loro incidenza è di circa il 10% sul totale delle parole (1 pausa ogni 20 parole circa);
- 2) la maggior parte delle pause (sia piene che vuote, 66%) è di tipo volontario, e realizza una funzione *demarcativa* o *strategica*, mentre solo il 34% si può addebitare a fenomeni 'esitativi' veri e propri o fisiologici
- 3) (più difficile è individuare le pause di programmazione e identificarne la volontarietà);
- 4) circa le loro posizioni, è sorprendente che ben il 73% occorra in posizione interna di unità intonativa e/o sintattica (con valori > 90% per le VFP), solo il 27% invece ad un confine (più spesso iniziale) di unità;
- 5) sul piano acustico, le pause piene (FP) durano in media il doppio delle altre pause, incluso le piene da allungamento di parola (48ms in media contro i 25ms delle VFP e i 23ms delle <sp>) e sono costituite in prevalenza da vocali nasalizzate;
- 6) Per quel che riguarda la relazione tra pause e gesti cooccorrenti (*co-speech gestures*), i dati si limitano al momento a osservazioni da approfondire:

- 7) in corrispondenza di FP sembrerebbe verificarsi una completa e momentanea interruzione del gesto in corso;
- 8) mentre, in corrispondenza delle VFP, si realizzano diverse forme di interazione tra il parlante (la guida) e il pubblico presente che sembrano coinvolgere, però, l'espressione facciale e i movimenti della testa (o dell'intero corpo) piuttosto che la gestualità delle mani.

### *Riferimenti bibliografici*

- BOERSMA, P., WEENINK, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.40
- CARLMEYER, B., SCHLANGEN, D. & WREDE, B. (2016). Exploring self-interruptions as a strategy for regaining the attention of distracted users. In *Proc. of the 1st Workshop on Embodied Interaction with Smart Environments*. ACM, 4.
- CHAFE, W. (1980). Some reasons for hesitating. In *Temporal variables in speech*, 169-180.
- ELAN (Version 5.2) (2018). [Computer software].
- GOLDMAN-EISLER, F. (1958). The predictability of words in context and the length of pauses in speech. In *Language and Speech*, 1(3), 226-231.
- GOLDMAN-EISLER, F. (1972). Pauses, clauses, sentences. In *Language and speech*, 15(2), 103-113.
- HIEKE, A. E. (1981). A content-processing view of hesitation phenomena. In *Language and Speech*.
- LICKLEY, R. J. (1998). *HCRC disfluency coding manual*. HCRC, University of Edinburgh.
- O'CONNELL, D. C., KOWAL, S. (1983). *Pausology*. In W. A. SEDELOW, S. Y. SEDELOW (eds.), *Computers in Language Research 2*. Berlin, New York, Amsterdam: Mouton Publishers, 221-301.
- REICH, S. S. (1980). Significance of pauses for speech perception. In *J. of Psycholinguistic Research*, 9(4), 379-389.
- SHRIBERG, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Doctoral dissertation, University of California, Berkeley.
- STAM, G., TELLIER, M. (2017). The sound of silence: the functions of gestures in pauses in native and nonnative interaction. In *Why Gesture? How the hands function in speaking, thinking and communicating*. John Benjamins 353-377.

ROSANNA SORNICOLA<sup>1</sup>, GIOVANNI ABETE<sup>1</sup>, ELISA D'ARGENIO<sup>1</sup>,  
CESARINA VECCHIA<sup>1</sup>

<sup>1</sup> Università degli Studi di Napoli "Federico II"

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

In questo contributo si discutono le esperienze maturate dal nostro gruppo di ricerca nell'ambito della costituzione e della diffusione di un archivio di parlato dialettale, con particolare riferimento ai risultati di un recente progetto sulla digitalizzazione e messa in rete dei materiali d'archivio. Con questo progetto si è cercato di definire e creare un nuovo modo di rendere vivo e accessibile un archivio in rete, valorizzando il materiale in archivio attraverso lo sviluppo di narrazioni culturali coinvolgenti, basate sugli stessi materiali di archivio e su altri materiali multimediali appositamente creati per il progetto. Tali "narrazioni" consentono di raccontare l'archivio a un pubblico di non specialisti, mettendo in luce l'interesse dei materiali raccolti tanto sul piano linguistico quanto sul piano storico-culturale. Questa modalità di fruizione "mediata" dell'archivio si affianca a una modalità di accesso più tipicamente archivistica riservata agli specialisti (dialettologi, etnografi, storici, etc.). In questo contributo ci concentreremo sugli aspetti divulgativi e di valorizzazione dell'archivio, cioè su quella che per gli archivisti italiani costituisce la terza fase di vita di un archivio, nella quale «l'archivio diventa prevalentemente, anche se non esclusivamente, oggetto di indagine storiografica e di valorizzazione» (Bonfiglio-Dosio, 2010: 30).

Nell'architettura del nostro archivio di parlato dialettale un ruolo fondamentale è svolto dagli etnotesti, che costituiscono anche l'unità documentaria primaria. Si definisce etnotesto un «testo orale che rappresenta la libera espressione di un settore o di un frammento della cultura di una comunità linguistica attraverso ricordi personali, testimonianze di usi, tradizioni, credenze, descrizione di oggetti e del loro funzionamento, ma anche indovinelli, filastrocche, leggende e storie» (Redazione ALEPO, 2003). Gli etnotesti, che possono avere anche estensione rilevante, «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).

Gli etnotesti hanno molteplici valenze: come oggetto di analisi linguistica e dialettologica; come fonte storica per la ricostruzione delle dinamiche sociali e culturali vissute da un territorio; come mezzo per la valorizzazione delle peculiarità linguistico-culturali del territorio. In questo contributo metteremo in

luce come tutte queste potenzialità possano essere sfruttate al meglio in un moderno archivio di parlato. Per farlo, prenderemo in esame, come caso esemplare, il trattamento nel nostro archivio degli etnotesti che si riferiscono alla pastorizia transumante, un'attività che ha avuto una grande importanza nella storia economica e culturale di molte aree d'Europa. I dati che verranno discussi riguardano un'area della Campania che va dalle pianure interne del vesuviano e del nolano alle montagne del Partenio e dei Monti Picentini, e che è stata a lungo e intensamente frequentata dai pastori transumanti irpini. Tali aspetti della storia della Campania sono poco noti ed emergono soprattutto dalle testimonianze orali dei pastori. Attraverso tali testimonianze si scopre l'esistenza di una rete di tratturi minori, ma storicamente molto importanti, che conducono non solo al tavoliere pugliese ma anche ai pascoli tirrenici, in particolare del nolano, dell'acerrano e dell'area vesuviana interna e, più a sud, della piana del Sele. La rilevanza di questi percorsi e della rete di relazioni create dalla transumanza si manifesta in molteplici aspetti della cultura locale: nelle tradizioni religiose, nelle abitudini alimentari, nel paesaggio umano, e, non da ultimo, nelle caratteristiche dei dialetti (Abete, 2016; Retaro, Abete, 2018; Vecchia, 2018).

Dopo aver presentato alcuni esempi degli etnotesti catalogati nel nostro archivio, si discuteranno le strategie che abbiamo utilizzato per raccontare questi materiali a un pubblico di non specialisti. In particolare, si mostrerà come a partire dai documenti di archivio relativi alla transumanza sia stata creata una narrazione multimediale ("Tratturi e antiche vie della terra di mezzo") che lega insieme i nostri documenti in una struttura narrativa coerente e li arricchisce di informazioni storiche, sociali ed economiche che aiutano a contestualizzare i documenti di archivio e ad apprezzarne la portata storica e culturale. Le narrazioni, che hanno una strutturazione gerarchica e sono fruibili su livelli diversi di approfondimento, consentono quindi di accedere ad una selezione dei materiali d'archivio, e di conoscerne la storia e il significato attraverso uno strumento comunicativo coinvolgente.

Questa modalità di presentazione dei documenti sonori è funzionale anche alla messa in opera di strategie volte alla valorizzazione delle specificità linguistiche e culturali del territorio e allo sviluppo turistico, in particolare dei luoghi "minori" rappresentati dai piccoli comuni. Nell'ultima parte della presentazione, ci soffermeremo dunque su questi aspetti, presentando un'altra peculiarità del nostro progetto, ossia la creazione di sedi decentrate dell'archivio situate in alcuni comuni dell'entroterra campano e concepite come piccoli musei della memoria o "mnemoteche". In particolare, in questo contributo verrà presentato il lavoro svolto dal nostro gruppo di ricerca per la progettazione e l'allestimento della "mnemoteca" di Ariano Irpino, un museo multimediale dedicato alla storia culturale, sociale e linguistica dell'Irpinia. All'interno della mnemoteca di Ariano Irpino il materiale audio-visivo raccolto nell'archivio è stato organizzato in tre diversi percorsi caratterizzati da differenti modalità di fruizione. Una prima sezione è dedicata al materiale fotografico con documenti

che riguardano i mestieri tradizionali, la vita familiare e le attività collettive delle comunità irpine. Una seconda sezione propone percorsi multimediali con postazioni tematiche interattive che offrono approfondimenti sulle attività produttive tradizionali, sul complesso quadro linguistico dell'Irpinia, sul patrimonio archeologico e artistico, sulle antiche forme di spiritualità e sulla vita religiosa del territorio. Il terzo percorso, infine, include filmati e brevi documentari su musei, siti archeologici, luoghi di culto, pratiche ed attività tradizionali.

Attraverso queste diverse strategie di divulgazione, il nostro progetto risponde alla riconosciuta esigenza di “restituzione” del dato al territorio, anche in considerazione delle possibilità offerte dalle nuove tecnologie (cfr. Matranga, in stampa), e offre dunque un modello per la valorizzazione delle fonti orali in quanto beni immateriali.

#### *Riferimenti bibliografici*

ABETE, G. (2016). La pratica della transumanza nella formazione dello spazio linguistico centro-meridionale: problemi e ipotesi di ricerca. In G. MARCATO (a cura di), *Il dialetto nel tempo e nella storia*. Padova: CLEUP, 379-386.

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

BONFIGLIO-DOSIO, G. (2010). Il sistema archivistico italiano. Archival organization in Italy. In *Documenta & Instrumenta*, 8, 29-45.

MATRANGA, V. (in stampa). Lessico dialettale e “dispositivi mobili”. In *Atti del Convegno “Lessicografia dialettale e etimologica: convegno in onore di F.D. Falucci”*, Corte, 28-30 ottobre 2015.

REDAZIONE ALEPO, (2003). *Le parole della geografia linguistica*. Scaricato da [www.alepo.it](http://www.alepo.it).

RETARO, V., ABETE, G. (2018). Sull'importanza delle aree intermedie: i dialetti del Vallo di Lauro. In ANTONELLI, R., GLESSGEN, M. & VIDESOTT, P. (a cura di), *Atti del XVIII Congresso Internazionale di Linguistica e Filologia Romanza*, Roma, 18-23 luglio 2016. Vol. 2. Strasbourg: ELiPhi, 957-968

VECCHIA, C. (2018). Elementi innovativi nel dialetto di Bagnoli Irpino (Av). In MARCATO, G. (a cura di), *Dialetto e Società*. Padova: CLEUP, 111-117.

CLAUDIO ZMARICH<sup>1</sup>, SABRINA BONICHINI<sup>2</sup>, SERENA BONIFACIO,  
VINCENZO GALATÀ<sup>1,3</sup>, YVAN ROSE<sup>4</sup>

<sup>1</sup> Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione,  
<sup>2</sup> Università degli Studi di Padova, <sup>3</sup> Università degli Studi di Siena, <sup>4</sup> Memorial University  
of Newfoundland

## **Bringing the phonological Development of Italian Children aged 30-47 months into PhonBank**

The present paper represents an important contribution towards the PhonBank Initiative (a computerized online database of recordings and transcripts documenting phonological development, e.g., Rose, MacWhinney, 2014), by filling the gap on the availability of Phon coded data on the phonological development of Italian children aged 30-47 months. PhonBank and Phon are recent outgrowths of earlier work on the Child Language Data Exchange System (CHILDES) Project.

In 1984, the primary emphasis of CHILDES was on the shift from handwritten notes to computerized files that could be subjected to automatic searching for developmental patterns, but it was not until 1995 that transcripts began to be linked to digitized audio. Prior to this, in 1993, an initial attempt was done to extend the CHILDES system to work on phonology. However, the lack of a consistent method for encoding International Phonetic Alphabet (IPA) characters on the computer at that time was a stumbling block. In recent years, the advent of Unicode, XML, improved digital audio and affordable file storage have removed the remaining technical stumbling blocks leading to the establishment of PhonBank (a computerized online database of recordings and transcripts documenting phonological development). The overall goals of PhonBank are fully compatible with those of TalkBank (a system for sharing and studying a wide range of conversational and communicative interactions), as well as with new developments in corpus phonology. The basic idea is that, by constructing a large database of accurately transcribed data on phonological development, we can test alternative theories of phonology and phonological development. For example, we can examine the role of universals in phonological markedness versus influences from distributional language specific patterns in the input to the children. We can then use these baseline patterns from normal development to understand the range of variation in learning and ways in which children with language disabilities diverge from developmental norms. The basic tools of PhonBank can also be used to study phonological systems beyond early child language; these tools can be applied to the acquisition of second languages,

the learning and mastery of dialects, and the study of phonological effects in aphasia and other communicative disorders.

Within PhonBank, phonological transcriptions and acoustic data can be elaborated thanks to Phon, an opensource software built to support corpus-based phonological analysis (Hedlund, Gregory & Rose, 2018. Phon 3.0.0 [Computer Software]. Retrieved from <https://phon.ca>, on 27/09/2018). The creation and development of Phon responds to the central need within PhonBank to have a mean to build corpora annotated by means of phonetic transcriptions and TextGrid annotations, and to analyse its data. PhonBank is unique because all of its corpora are fully transcribed in IPA, with most children documented between 10 and 36 months of age, which also includes relatively rare transcriptions of children's very early stages of production, including studies on babbling. Because of the young age of recorded infants who rarely produce spontaneous speech, PhonBank also includes some transcriptions of adults' inputs. In line with both CHILDES and the general TalkBank system (see below), another central feature of PhonBank lies in the idea of data sharing: linguists and researchers who utilize dataset or software developed within these systems are encouraged to contribute to its improvement and expansion. They can upload new data relevant to the PhonBank project using the guidelines provided from the Website <http://talkbank.org/share/contrib.html>. PhonBank is indeed part of a wider project called TalkBank, a large system that combines 17 different databases, each of which is specific of a particular branch (below called "banks") for the study of language. All these different data collections are central to current investigations in many different areas of language. PhonBank is the one specific for phonetic and phonological language development.

The general graphical interface of Phon is composed by a set of windows and each of them has a specific function. A typical operative display includes Media and Segmentation, Record Data, Syllabification, Alignment, and Waveform. Phon supports multimedia data linkage, unit segmentation, blind transcription, automatic syllabification of data (according to the rules of several languages, Italian included, thanks to ISTC-CNR, <http://www2.pd.istc.cnr.it/FESTIVAL/home/default.htm>), model (Target) and produced (Actual) phonological forms and their alignment. This alignment is useful to perform systematic, phone-by-phone comparisons between model and produced forms. An essential function of Phon is the Query browser for data extraction and analysis. According to one's research question, different types of queries can be created to fulfill specific analyses, many of which can be used to identify and study speech errors, thanks to the system's capacity to match the aligned model/produced phones present in corresponding IPA Target and Actual tiers. In order to exemplify the potential of PHON, we will present some of the analyses carried out on the speech productions of a group of 30 typically developing children from the Veneto region in Italy (aged from 30 to 47 months, equally divided by gender) which we coded into Phon. The children were

recruited at different times from 2011 to 2016 by graduating students of the logopedic course of the University of Padova, mainly in kindergarten or at the children's homes, in the Veneto region (I). The children's parents gave their informed consent, compiled the MacArthur CDI surveys for their children's productions (Caselli, Pasqualetti & Stefanini, 2007), from which we estimated the size of their vocabularies, and the parents also filled a questionnaire from which we inferred the children to be, among other things, of normal psycho-physical development and monolingual (Italian). The children were administered a new "Test Fonetico per la Prima Infanzia" (*Phonetic test for toddlers*" or TFPI, see XXX), which basically is a picture naming test. The children's speech productions were manually transcribed in IPA, by several skilled transcribers, which finally confronted with a second series of transcriptions made by the first author, until a final agreement was reached by consensus. Transcriptions were carried on with the help of PRAAT facilities, and textgrids were created with segmentations and labels which were later successfully exported into Phon, by means of a Praat script by XXX.

For the present proposal, we divided the children into 6 groups spanning three months each. Exploiting the functionalities available in Phon, we derived individual phonetic inventories as well as group inventories (based on age). A phone is considered to be acquired if it is realized at least 50% of the times it was proposed to the child within the target words (cutoff criterion). Beyond phonetic inventories, which is a product of an "independent" analysis of the children's corpus, since it does not assess the correctness of the productions, an "Accuracy" analysis was produced. This kind of analysis is qualified as "relational", because the phonetic segments in the words produced by children are systematically compared with the target words, and the target segments are classified as "accurate", "substituted", "deleted" (e.g. Rose, Stoel-Gammon, 2015; McAllister, Byun & Rose, 2016). The results of the above analysis will be compared and discussed with reference to both a previous study of the same authors (XXX) on phonetic inventories of Italian children aged 18 to 27 months, and to existing studies on Italian children of comparable ages (Zanobini, Viterbori & Saraceno, 2012; Tresoldi et al., 2018). In addition to this, another corpus consisting in the speech productions of 10 Italian children recorded every three months from 18 to 48 months of age, will be uploaded to PhonBank in the course of 2019, thus providing the scientific community with additional reference data on Italian.

### *Riferimenti bibliografici*

CASELLI M.C., PASQUALETTI P. & STEFANINI S. (2007). *Parole e frasi nel "Primo vocabolario del bambino"*. Milano: FrancoAngeli.

MCALLISTER BYUN, T., & ROSE, Y. (2016). Analyzing Clinical Phonological Data Using Phon. In *Seminars in Speech and Language*, 37(2), 85–105.

ROSE, Y., MACWHINNEY, B. (2014). The PhonBank Project: Data and Software-Assisted Methods for the Study of Phonology and Phonological Development. In J. DURAND, U. GUT, & G. KRISTOFFERSEN (Eds.), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press, 380–401.

ROSE, Y., & STOEL-GAMMON, C. (2015). Using Phonbank and Phon in Studies of Phonological Development and Disorders. In *Clinical Linguistics & Phonetics*, 29(8–10), 686–700. <https://doi.org/10.3109/02699206.2015.1041609>

TRESOLDI, M., BARILLARI, M. R., AMBROGI, F., SAI, E., BARILLARI, U., TOZZI, E., SCARPONI L. & SCHINDLER, A. (2018). Normative and validation data of an articulation test for Italian-speaking children. In *International Journal of Pediatric Otorhinolaryngology*, 110, 81-86.

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.

PIERO COSSU<sup>1</sup>

<sup>1</sup>Università di Pisa

## La metatesi di /r/ nel dialetto di Suelli (sardo campidanese). Una analisi descrittiva dei nessi aberranti

Il dialetto di Suelli, contestualmente a altri sub-dialetti del dominio sardo-campidanese, fa parte dell'area dialettale colpita dalla metatesi della consonante rotica /r/. Il *trigger* della metatesi sono le consonanti rotiche in coda sillabica e ha come esito un attacco complesso, perlopiù a inizio di parola, in cui /r/ è il secondo elemento (/ 'pərku/ > \*'proku+ “maiale”, /fər'miʝa/ > \*fro'miʝa] “formica”, /kar'bəne/ > \*kra'βəni+ “carbone”, Wagner, 1941, Virdis, 1978, Lai, 2013).

La metatesi di /r/ nei dialetti campidanesi opera da destra a sinistra, entro i confini lessicali, e per avere luogo devono essere presenti i seguenti vincoli fonotattici:

- a) la consonante rotica, nella forma soggiacente, deve trovarsi in posizione di coda;
- b) l'attacco successivo non può essere una sonorante, tranne la nasale bilabiale /m/;
- c) deve esserci un attacco precedente;
- d) l'attacco precedente non può essere una sonorante, tranne la nasale bilabiale /m/.

Se (b) non è rispettato, in luogo della metatesi occorre invece l'epentesi di una vocale la cui qualità è copia di quella che precede la consonante rotica /r/ (/ 'pərɫa/ > \*'pərɛɫa+ “perla”). Se (c) e (d) non sono rispettati, la direzione della metatesi è invertita da sinistra a destra (/ 'ɛrba/ > \*'ɛβra+ “erba”, /'nɛrβju/ > \*'nɛβrju+ “nervo” (Molinu, 1999)).

Oltre a una serie di TR assolutamente non marcati, interlinguisticamente molto frequenti ([pr], [tr], [kr], [fr], etc), la metatesi ha prodotto anche una serie di attacchi complessi marcati e rari tra le lingue del mondo: ['srɔʝa+ “suocera”, \*'mrekunizi] “mercoledì”, ['tsrupu] “cieco”, ['ʃreβu] “cervo”.

Gli ultimi due nessi consonantici, in particolare, sono, a nostra conoscenza, presenti solamente nel dialetto campidanese. I primi due invece, pur essendo attestati anche in altre lingue del mondo, dal punto di vista teorico sono un *monstrum* in quanto i segmenti che li compongono sono detti essere incompatibili per formare un attacco complesso, come è sostenuto in alcuni *framework*

fonologici (Fonologia della Reggenza (Kaye, Lowenstamm & Vergnaud, 1985, 1990), Modello CVCV (Scheer, 2004)).

Nel nostro studio intendiamo fornire una analisi descrittiva della fonetica di tali nessi aberranti e infrequenti, assente prima d'ora in letteratura. Le domande che ci poniamo sono:

- a) l'interazione tra i segmenti che compongono gli attacchi complessi risultanti dalla metatesi presenta della peculiarità acustico-articolatorie o si allinea a quella dei nessi non marcati?
- b) i *pattern* di articolazione della consonante rotica nei nessi aberranti è lo stesso di quello riscontrato quando si trova in posizione intervocalica o nei nessi non marcati?
- c) sono in atto delle strategie di riparazione per agevolare l'articolazione di questi nessi (soprattutto quelli composti da affricata e consonante rotica)?

Per provare a rispondere a queste domande abbiamo creato un *corpus* di registrazioni a cui hanno partecipato venti parlanti nativi del dialetto di Suelli, ai quali è stato chiesto di tradurre 26 frasi in italiano regionale di Sardegna nel proprio dialetto (*ergo*, il subdialetto di Suelli). Abbiamo scelto l'italiano regionale di Sardegna perché presenta una sintassi più vicina al sardo campidanese rispetto all'italiano standard, di modo da ridurre il carico di lavoro in fase di traduzione. Abbiamo scartato l'opzione di presentare le frasi in una varietà sovralocale sarda (come ad esempio quella del dialetto di Cagliari, in cui la metatesi non è attestata (Viridis, 1988)) per escludere l'influenza della grafia come fattore concorrente all'attivazione o meno della metatesi. Ogni frase conteneva una parola *target* nella quale attendevamo uno dei quattro nessi marcati discussi, una delle due affricate [tʃ] o [ts] non seguite da [r], le quali fungono da primo segmento nei nessi tipologicamente più marcati in assoluto, o la consonante rotica in posizione intervocalica a inizio o in corpo di parola. Per quanto riguarda i nessi marcati, oggetto principale del nostro studio, [tsr], [mr] e [sr] occorrono 51 volte ciascuno, mentre il nesso [tʃr] è presente 68 volte.

L'analisi fonetica è stata condotta tramite l'ausilio del *software* PRAAT. Abbiamo considerato principalmente la variabile durata della consonante rotica e il modo di articolazione della stessa. I risultati da noi ottenuti mostrano che il *range* dei modi di articolazione della consonante rotica non cambia se questa si trova in posizione intervocalica, è il secondo elemento di un attacco complesso marcato o comune. La consonante rotica nei dialetti sardi in posizione intervocalica è stata studiata da Contini (1987), il quale la descrive come polivibrante alveolare, se geminata, e monovibrante, se scempia. Nel nostro *corpus* essa si manifesta come approssimante alveolare [ɹ], fricativa alveolare non sibilante [ɹ̥], polivibrante alveolare [r] e monovibrante alveolare, o *tap*, [r̥]. Gli allofoni riscontrati sono quelli previsti dal modello di Ladefoged (1996) quando il gesto inteso dal parlante è quello di una polivibrante [r] e dipendono dal grado di accuratezza articolatoria (Solé, 2002). Per far sì che l'apice della lingua vibri in prossimità della cresta alveolare devono mantenersi per tutta la durata

dell'articolazione del fono una adeguata differenza di pressione orofaringale, una sostenuta rigidità del dorso della lingua e rilassamento dell'area apicale (*ibidem*). Anche l'articolazione e l'acustica dei primi elementi degli attacchi in esame è la stessa di quando gli stessi elementi si trovano in posizione intervocalica. La durata della consonante rotica è soggetta ad alta variabilità sia al livello intra- che interindividuale: si aggira tuttavia in un *range* che va dai 20 ms ai 50 ms, mediamente. L'alta variabilità nella durata e nel modo di articolazione della consonante rotica, la quale mostra tuttavia il medesimo *pattern* quando la consonante è il secondo elemento di un attacco complesso o è in posizione intervocalica, indurrebbe a pensare che essa dipenda solamente da fattori prettamente articolatori (oscillazioni nella configurazione oro-linguale durante la produzione della consonante rotica) e non sia condizionata dal contesto fonetico e soprattutto fonologico in cui essa occorre. In ultima analisi, riteniamo che i parlanti nel nostro studio non abbiano alcun problema a articolare sequenze di consonanti come quelle dei nessi marcati del dialetto campidanese di Suelli e non interpretano la consonante rotica in maniera differente dal punto di vista fonetico-fonologico rispetto a quando si trova in altri contesti (quello intervocalico o in posizione preconsonantica: per quest'ultimo contesto abbiamo tuttavia pochissimi *token* in quanto non ne era prevista un'analisi in fase di elaborazione degli stimoli).

Per quanto riguarda le strategie attuate dai parlanti in fase di produzione dei nessi *sub indice*, abbiamo notato che, quando il fono realizzato è una monovibrante [r], il gruppo consonantico è sempre intervallato una piccola porzione vocalica di circa 20 ms (la cui durata varia al livello interindividuale e intraindividuale). Tuttavia, la sua sistematica occorrenza è stata riscontrata in un gran numero di lingue quando il *tap* segue o precede un'altra consonante in nessi non marcati (Harms, 1976, in finlandese, Jannedi, 1994, nel dialetto tedesco di Amburgo, Ramirez, 2006, per lo spagnolo, Baltazani, 2009, in greco, Savu, 2012, per le lingue slave, Celata et al., 2017, per il toscano). La nostra analisi ci ha condotto a analizzare il vocoide come parte integrante dell'articolazione del *tap* quando questo è in posizione postconsonantica. Si tratterebbe dunque di un elemento che non ha la funzione di riparare una struttura sillabica mal formata, ossia non è fonologicamente rilevante, e la sua presenza non è associata a un gesto articolatorio, ma è un prodotto secondario dell'articolazione della monovibrante [r]. Per Savu (2012) la porzione vocalica precede e segue il *tap* anche quando questo è in posizione intervocalica, come dimostrerebbero le transizioni agli estremi della fase di occlusione di [r] con valori formantici simili a quelli del vocoide che appare invece chiaramente sullo spettrogramma nei contesti Cr e rC.

In conclusione, pare che i parlanti campidanesi percepiscano i nessi aberranti di cui discutiamo come assolutamente ordinari e nessuna strategia di riparazione è messa in opera. I *pattern* articolatori e acustici dei segmenti nei TR metatetici non cambiano al mutare del contesto di occorrenza degli stessi elementi (quando

in attacchi semplici o in attacchi complessi non marcati). Il vocoide appare solo quando la consonante rotica /r/ è articolata come *tap* [r+: è parte dell'articolazione dello stesso e non deriva da una operazione fonologica riparativa.

### *Riferimenti bibliografici*

- BALTAZANI, M. (2009). Acoustic characterization of the Greek rhotic in clusters. In *Selected papers from the 18th ISTAL*.
- CELATA, C., VIETTI, A. & SPREAFICO, L. (2017, in press). An articulatory account of rhotic variation in Tuscan Italian: synchronized UTI and EPG data. In J. Gil, M. Gibson (Eds.), *Romance Phonology and Phonetics*. Oxford: Oxford University Press.
- CONTINI, M. (1987). *Etude de geographie phonetique et de phonetique instrumental du sarde*. Torino: Edizioni dell'Orso.
- HARMS, R.T. (1976). The segmentalization of Finnish 'nonrules'. In *Texas Linguistic Forum*, 5, 73–88.
- JANNEDY, S. (1994). Rate effects on German unstressed syllables. In *OSU Working Papers in Linguistics*, 44, 105–124.
- KAYE, J., LOWENSTAMM, J. & VERGNAUD, J-R. (1985). The internal structure of phonological elements: A theory of charm and government. In *Phonology*, Yearbook 2, 305-328
- KAYE, J., LOWENSTAMM, J. & VERGNAUD, J-R. (1990). Constituent structure and government in phonology. In *Phonology* 7(2), 193-232.
- LAI, R. (2013) *Positional Effects in Sardinian Muta cum Liquida. Lenition, Metathesis, and Liquid Deletion*. Alessandria: ed. dell'Orso.
- LADEFOGED, P., MADDIESON, I. (1996). *The sounds of the world's languages*. Oxford: Blackwells.
- MOLINU, L. (1999). Métathèse et variation en sarde. In *Cahiers de grammaire*. Toulouse: Université de Toulouse-le-Mirail.
- RAMÍREZ, C. J. (2006). Acoustic and perceptual characterization of the epenthetic vowel between the clusters formed by consonant + liquid in Spanish. In M. DIAZ-CAMPO (ed.), *Selected proceedings of the 2nd conference on laboratory approaches to Spanish phonetics and phonology*. Somerville, MA: Cascadilla Proceedings Project, 48-61.
- SAVU, C-F. (2012). An acoustic-phonetic perspective on the phonological behaviour of the rhotic tap. In *Proceedings of ConSOLE XX*, 213-227.
- SOLÉ, M-J. (2002). Aerodynamics characteristics of trills and phonological patterning. In *Journal of Phonology*, 30, 655-688.
- VIRDIS, M. (1978). *Fonetica del dialetto sardo campidanese*. Edizioni della torre.
- WAGNER, M. L. (1941). *Historische Lautlehre des Sardischen*. Halle (Saale): Max Niemeyer.

EMMA MILANO<sup>1</sup>

<sup>1</sup> Università degli Studi di Napoli “Federico II”

## **L’Archivio Sociolinguistico ed Etnografico di Greci nel progetto *Voci, parole e testi della Campania*. La fonte orale come risorsa per lo studio linguistico e per la valorizzazione e la tutela dei valori linguistico-culturali di minoranza**

In questo contributo saranno discusse alcune questioni metodologiche emerse nell’ambito della progettazione e della messa a punto di un archivio sociolinguistico relativo all’unica comunità di minoranza della Campania, la comunità arbëreshe di Greci in provincia di Avellino. L’archivio di Greci si iscrive all’interno del progetto *Voci, parole e testi della Campania*, finalizzato all’archiviazione, alla tutela e alla valorizzazione di beni culturali del territorio campano, e dunque nasce con l’intento di documentare la lingua e la memoria dell’unica comunità di minoranza linguistica campana.

Il progetto “Voci Parole e Testi della Campania” si propone nello specifico di rendere accessibile un ampio archivio sociolinguistico ed etnografico della Campania con l’obiettivo di promuovere la conoscenza del territorio e della cultura linguistica e sociologica campana.

In questo intervento si intende soffermarsi innanzitutto sulle modalità attraverso cui l’archivio di Greci si incardina nell’economia generale del progetto *Voci, parole e testi della Campania* e sul contributo che esso ha dato al progetto nel suo insieme, rappresentando una sorta di valore aggiunto. In secondo luogo oggetto di discussione sarà l’importanza delle fonti orali per la valorizzazione culturale delle comunità di minoranza e i vantaggi che la storia orale può offrire alla ricerca linguistica delle varietà di minoranza.

Entrambe le questioni sono strettamente connesse innanzitutto alla specificità linguistica e culturale di Greci e contemporaneamente ai suoi peculiari legami culturali, antropologici e linguistici con il territorio circostante.

Greci è un piccolo comune di circa 800 abitanti situato nella zona nord-orientale dell’Irpinia, sede dell’unica minoranza linguistica campana di origine italo-albanese di lingua arbëreshe. L’origine dell’insediamento di Greci si iscrive nella storia delle comunità albanesi, che, in diverse ondate, a partire dal XV secolo, migrarono verso l’Italia meridionale. Nel quadro dai contorni poco noti di tale interessante storia, Greci si caratterizza in maniera peculiare per diversi aspetti, spesso messi in relazione alla sua posizione geograficamente isolata

rispetto ai più numerosi e compatti arcipelaghi arbëreshë calabresi e siciliani. La piccola comunità irpina è riuscita infatti a custodire per ben 5 secoli una lingua che in molti altri comuni arbëreshë dell'Italia meridionale dava da tempo segni di cedimento e che, tra l'altro, ha conservato caratteristiche linguistiche peculiari. Negli ultimi trenta/quarant'anni tuttavia, i nostri informatori raccontano di una progressiva contrazione degli usi e delle competenze arbëreshe. Tale processo di perdita di lingua può essere inquadrato nell'ambito di una "diaspora nella diaspora" di cui sono state protagoniste le comunità arbëreshe d'Italia. Si consideri a questo proposito che in questo arco temporale il numero degli abitanti di Greci si è ridotto di circa il 60%.

La comunità grecese è dunque depositaria di valori culturali linguistici antropologici unici, ma dall'altra è partecipe di una serie di dinamiche storiche, culturali e sociali che riguardano tutta l'Italia meridionale. Basti pensare da un punto di vista storico-sociale a processi come l'emigrazione e l'urbanizzazione cui sono stati sottoposti, come Greci e gli altri comuni arbëreshë, tanti centri dell'Italia meridionale, e dal punto di vista linguistico culturale al rapporto lingua e cultura locale e lingue e cultura dominante, o, ancora, al ruolo della scuola nei processi di italianizzazione nell'Italia post unitaria, fenomeni che accomunano Greci a gran parte del territorio italiano.

Nell'ambito del progetto "Voci Parole e Testi della Campania", Greci occupa pertanto una posizione peculiare. Per il suo peculiare statuto di comunità di minoranza, le dinamiche sopracitate sembrano infatti manifestarsi in maniera più evidente e forse potremmo dire 'drammatica' e la piccola comunità arbëreshe irpina ha rappresentato pertanto una sorta di laboratorio e in cui è stato possibile sperimentare una serie di strategie di intervento, misurandone, pesandone l'impatto in una maniera più immediata e manifesta.

In questa prospettiva Greci ha contribuito a fornire un'ulteriore tessera di quel puzzle ricco, complesso e altamente sfaccettato costituito dal territorio dell'Irpinia e della Campania e la chiave, per quanto riguarda la parte del progetto riguardante Greci, è stata proprio rappresentare le caratteristiche di unicità in termini storici e linguistico-culturali della minoranza senza trascurare i legami con il territorio di insediamento inteso in maniera ampia.

L'Archivio di parlato di Greci, attraverso l'ausilio di materiali multimediali e audio-visivi, ha consentito di approfondire aspetti della cultura, della storia e soprattutto della lingua di questa comunità caratterizzata da rilevanti specificità culturali e linguistiche.

Nonostante il difficile quadro appena tracciato, a Greci infatti oggi si continua a parlare *arbërisht*. L'arbëresh di Greci rimane inoltre ancora oggi una varietà che ha ben conservato alcune peculiarità linguistiche, in particolare alcuni aspetti della struttura fonologica, della morfologia nominale e quella verbale, mostrando una considerevole capacità di attivare in maniera sistematica processi di adeguamento dei prestiti, che riguardano soprattutto l'ordine lessicale.

Uno spazio centrale all'interno dell'Archivio è infatti occupato dagli etnotesti. Gli etnotesti dell'Archivio di Greci costituiscono infatti un luogo privilegiato di osservazione e di studio della variazione interna ai sistemi linguistici di lingue 'minacciate' e delle dinamiche di conservazione e perdita delle lingue minoritarie. Nel contempo essi rappresentano una fonte storica di inestimabile valore per la ricostruzione di dinamiche sociali, economiche, demografiche e culturali che, se da una parte caratterizzano la storia di Greci in maniera peculiare, dall'altra ne illuminano gli strettissimi legami con il resto dell'Arberia, con il territorio Irpino e con i tanti piccoli centri dell'Italia meridionale, e non solo. I racconti orali documentati negli etnotesti hanno rappresentato infine un insostituibile strumento di tutela e valorizzazione dei valori culturali e linguistici di minoranza in un processo virtuoso che ha visto la comunità acquisire una sempre maggiore consapevolezza di sé e della propria specificità identitaria culturale e linguistica, essendo nel contempo protagonista di un affascinante viaggio nel tempo e nella memoria.

#### *Riferimenti bibliografici*

- ALTIMARI, F., BOLOGNARI, M., CARROZZA, P. (a cura di) (1986). *L'esilio della parola. La minoranza linguistica albanese in Italia. Profili storico-letterari, antropologici e giuridico istituzionali*. Pisa: ETS Editrice.
- ALTIMARI, F. (2010). Sui prestiti dell'italiano e dei dialetti italo-romanzi nel sistema verbale dell'italo-albanese. In *Bollettino Linguistico Campano*, XVIII. Napoli: Liguori, 7-20.
- AUSTIN, P. K. (2010). Current issues in language documentation. In *Language documentation and description*, 12-33.
- CALAMAI, S. (2011). Ordinare archivi sonori: il progetto Gra. Fo. In *Rivista italiana di dialettologia*, 35, 135-164.
- CAMAJ, M. (1971). *La parlata albanese di Greci in provincia di Avellino*. Firenze: Leo S. Olschki editore.
- CONFORTI, G. (1922). *Appunti di Storia Cronologica di Greci*. Napoli: Tipografia Pontificia Degli Atigianelli.
- FISCHETTI, A. (2008). Creazione e gestione della fonte orale. In C. BERMANI, A. DE PALMA, (a cura di), *Fonti orali, istruzioni per l'uso*. Venezia: Società di mutuo soccorso Ernesto de martino.
- MACCANI, L., VIOLA, M. (2008). *Comunicare l'identità. Una strategia di valorizzazione delle minoranze linguistiche*. Franco Angeli.
- MATRANGA, V. (2010). Quattro appunti sulla consapevolezza linguistica in contesto italoalbanese. In T. KREFELD, E. PUSTKA (a cura di), *Perzeptive varietätenlinguistik*. Frankfurt: Peterlang, 457-72.
- MILANO, E., VALENTE, S. (2010). *Minoranze linguistiche e territorio. Il caso della comunità italo-albanese di Greci*, Numero Monografico del Bollettino Linguistico Campano, XVII. Napoli: Liguori.

ROMITO L., LIO, R. (2007). I corpora vocali come patrimonio culturale. Archiviazione, conservazione e restauro di materiale sonoro calabrese. In *Atti del 3° Convegno Nazionale AISV "Scienze Vocali e del Linguaggio Metodologie di Valutazione e Risorse Linguistiche"*, Trento, 29/11-1/12 2006. EDK Editore.

SAVOIA, L. M. (a cura di) (2008). *Studi sulle varietà arbëreshe*. Università della Calabria.

TELMON, T. (1999). Documenti sonori e ragioni della ricerca. In *Archivi Sonori Vercelli (22 gennaio 1993), Bologna (22-23 settembre 1994), Milano (7 marzo 1995)*, 28-34.

CLAUDIO ZMARICH<sup>1</sup>, ERICA BONATO, ROBERTO BOVO<sup>2</sup>, VINCENZO GALATÀ<sup>1,3</sup>, ALESSANDRA PINTON<sup>2</sup>

<sup>1</sup> Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, <sup>2</sup>Università degli Studi di Padova, <sup>3</sup>Università degli Studi di Siena

## **A new phonological discrimination test for children aged 48-72 months**

The proposed research is part of a wider project aiming at the creation of a phonological discrimination test for pre-school subjects. More specifically, the test we describe is designed to evaluate the ability to discriminate stimuli pairs differing with respect to the consonant's phonological category representing the output of each of the simplification processes (the so called "phonological processes") that accompany the speech production of typical developing children.

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); ii) speech disorders and language disorders with a phonetic-phonological component are an important, if not the main, portion of the caseloads of speech therapists who deal with voice-speech-language disorders in childhood (Law, Boyle, Harris, Harkness, Nye, 2000). Generally speaking, 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. The phonological discrimination 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 5 (Tamashige, Nishizawa, Itoda,

Kasai, Igawa, Fukuda, 2009). 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 are 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 language disorders (Rvachew, Brosseau-Lapr e, 2012; Freitas, Mezzomo, Marques, Vidor, 2014). Phonological discrimination tests may vary with respect to both the form and the content. 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 (Gerrits & Shouten, 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. 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, syllabic representation than a segmental one. 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 (minimum 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). 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). A more in-depth understanding of the relationships between perception and production in the course of development would help to 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. We started our endeavor in the creation of this test moving from the consideration that in Italy existing tests do not evaluate the phonological discrimination of late preschoolers (48-72 months-olds) in a satisfactory way. To be precise, there is a unique norm-referenced test which makes use of pairs of non-words in an AX paradigm. This test is proposed as a part of a battery (BVN 5-11, Tressoldi et al., 2005; also in Pinton & Zanettin, 1998), but it presents a number of problems, among them: it does not test all the Italian consonants nor all the phonetic features, some of the tested consonants are proposed only once, 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, it is not proposed in a playful way, the verbal stimuli have to be pronounced anew each time by the operator (i.e. they are not

prerecorded) and the list is written in orthography rather than in IPA. To begin with, we split the proposed test development in two sets: we first devised a set of stimuli based on contrasts involved in systemic phonological processes (mainly phones' substitution) and administered it to a group of children; we then devised a set of stimuli based on contrasts involving structural phonological processes (mainly phones' harmony or phones' addition/deletion). 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 pretest session precedes both test sets with the aim of familiarizing the children with the task. The first set consists of 39 non words 'VCV minimal pairs, testing all the 10 distinctive features of the Italian phonological system (as exposed by Schmid, 1999: 134), representing as many systemic processes (stopping, backing, gliding etc.). Each distinctive feature is represented in two different pairs for each of the chosen vocalic contexts ('aCa, 'iCi) with each pair opposing a consonant (+ feature) against a consonant (-feature). 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 example [+ continuous] VS [- 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 is called in this case "stopping": e.g. asa VS ata; isi VS iti; ava VS aba; ivi VS ibi. The second set consists of 68 minimal pairs, testing 12 structural processes (cluster reduction, consonant harmony, epenthesis etc.). Each distinctive process is represented in two different pairs for each of the chosen vocalic contexts ('aCa, 'iCi) with each pair opposing a target against the same target affected by a particular process. The structure of the test follows the AX paradigm and natural prerecorded stimuli are proposed after we verified their better perceptibility over synthetic stimuli (Gaiotto, 2016). The whole test administration procedure for both sets has been automated by means of a Praat script (XXX), and the randomized stimuli in each set were proposed to the children in a playful setting. The two sets of the test were administered by students graduating in speech therapy at the University of Padova to two groups of 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 (Bonato 2016); the second group (30 subjects) was assessed with the second set (Marchetti, 2017; Rossi 2017). According to their parents, who gave the informed consensus and compiled a questionnaire, all the children were normal under the psycho-physical profile. The children's audition, assessed through an audiometric screening, was within normal range. In a separate session all the children have also been tested with an articulation test (a modified version of TFPI, XXX, not published), in order to have the possibility to compare - in a separate study not addressed here - speech production and perception proficiency within each child. In the following analyses, we factorized ages (four groups spanning six months each)

and gender (male and female). Dependent variable is the percentages of discrimination. Among the main results relative to the first set (systemic contrasts) there are the following: no significant differences by age groups or by gender; the discrimination percentages improve as the number of distinctive features that distinguish pairs' stimuli increases; discrimination is better in /a/ context than /i/ context; the distinctive feature causing more difficulty is the "voice" feature. Among the main results relative to the second set (structural contrasts), there are the following: neither significant difference for vocalic context, nor for gender, but a difference almost significant emerges as to the age variable (between the first and the fourth age group); 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 the children.

GIOVANNA LENOCI<sup>1</sup>, IRENE RICCI<sup>1</sup>

<sup>1</sup>Scuola Normale Superiore

## **Curve-to-curve distances to assess lingual articulations in the speech of Italian children who stutter and control peers**

Stuttering is a speech motor disorder that may arise from an innate limitation in the speech motor control system, which fails to prepare and organize the movements required for speech (van Lieshout et al., 2004). In this view, Stuttering-Like Disfluencies (i.e., part-word repetitions, sound prolongations and broken words; Yairi, Ambrose, 2005), the hallmark characteristics of the disorder, reflect errors in the motor control system and they ultimately represent breakdowns in the precisely timed and coordinated articulatory movements required for fluent speech. Anticipatory coarticulation and stability of speech are two critical indexes of speech motor control maturity and the purpose of the study presented here is to assess if, even during fluent speech production, stuttering children use abnormal lingual articulations compared to control peers. A number of acoustic studies, in fact, revealed atypical coarticulatory patterns in the speech of children and adults who stutter compared to non-pathological population, but with quite equivocal results (Subramanian et al., 2003; Chang et al., 2002; Robb, Blomgren 1993; Sussman et al., 2010). A close understanding of the tongue movements during speech is provided by Ultrasound Tongue Imaging (henceforth, UTI) and so far, only few studies have been realized with stuttering children. In Lenoci, Ricci (2018) for example anticipatory coarticulation was investigated through the Locus Equations metric (henceforth, LE) while, for a subset of stimuli, stability of speech movements was estimated with the Nearest Neighbor Distances (henceforth, NND; Zharkova, Hewlett, 2009). Results show different articulatory patterns for the stuttering children, for both measures under investigation.

The purpose of the study presented here is twofold: a) to extend the analyses to a larger set of stimuli and b) to assess anticipatory coarticulation in the stuttering speech by using another method, the NND, which has been proposed as a good index of contextual coarticulatory variability (Rodríguez, Recasens, 2017).

UTI and acoustic data were collected at the SMART lab of Scuola Normale Superiore, from the speech of 4 school-aged (age range: 8-12 years) CWS (2 males and 2 females) and 4 age-matched CWNS (2 males and 2 females). In a child-friendly set up, participants were involved in a sentence repetition-after-listening

task in which they produced 12 repetitions of a dysyllabic pseudo-word, embedded in a carrier phrase of the type: 'La stellina C1V1ba salirà'.

Within the stressed first syllable (C1V1), C1 is /b/, /d/ or /g/ and V1 one of the three cardinal vowels, /a/, /i/ or /u/. Recordings were realized with an ultrasound system (Mindray UTI system-30 Hz) with a microconvex probe (Mindray probe 65EC10EA); the UTI images were synchronized with the audio through a synchronisation unit (Synch Bright-up unit).

A semi-automatic tongue contour splining was performed in the acoustic interval spanning from the beginning of the consonantal closure to the end of the following vowel. Each spline was then defined in terms of x,y coordinates, which were then used for comparing tongue curves. For the purposes of the study presented here, one frame within each target syllables was selected for measuring coarticulation and stability of speech: the midpoint of the consonantal closure.

Within each place of articulation, the goal of the analysis was to discover if the shape and location of the tongue curves was affected by the categories to which they belonged, for example /ba/ or /bi/ or /bu/. NND were used to estimate the degree of separation between all the curves in one category and those of the other by calculating the mean of all the Euclidean distances between each point on one curve and its nearest neighbor on the other.

The stability of speech movements is examined from the analysis of variability of multiple repetitions of the same C within the same vowel context (ex. /ba/); by contrast variability in the production across different vowel contexts (ex. /ba/ vs /bi/ vs /bu/) is taken as an index of anticipatory coarticulation (Zharkova, Hewlett, 2009). All distances were extracted by R scripts and Nearest Neighbor Distances were determined within each individual and then used in statistical analyses of differences according to the speaking groups. A preliminary analysis shows that CWS reach fluency in their speech with different lingual patterns (i.e., higher token-to-token variability and different coarticulatory patterns) compared to controls, corroborating the hypothesis of an impaired speech motor control system for the former group.

Results, in fact, suggest that CWS show tongue contours that differ in shape and location compared to controls.

### *Riferimenti bibliografici*

- CHANG, S.E., OHDE, R.N. & CONTURE, E.G. (2002). Coarticulation and formant transition rate in young children who stutter. In *Journal of Speech, Language and Hearing Research*, 45, 676-688.
- LENOCI, G., RICCI, I. (2018). An ultrasound investigation of the speech motor skills of stuttering Italian children. In *Clinical Linguistics & Phonetics*, 32:12, 1126-1144.
- ROBB M., M. BLOMGREN (1997). Analysis of F2 transitions in the speech of stutterers and nonstutterers. In *Journal of fluency Disorders*, 22, 1-16.

- RODRÍGUEZ, C., RECASENS, D. (2017). An evaluation of several methods for computing lingual coarticulatory resistance using ultrasound. In *Journal of Acoustical Society of America*, 142, 3783-3888.
- SUBRAMANIAN, A., YAIRI, E. & AMIR, O. (2003). Second formant transitions in fluent speech of persistent and recovered preschool children who stutter. In *Journal of Communication Disorders*, 36, 59-75.
- SUSSMAN, H. M., BYRD, C. T., & GUITAR, B. (2010). The integrity of anticipatory coarticulation in fluent and non-fluent tokens of adult who stutter. In *Clinical Linguistics & Phonetics*, 25 (3), 169-186.
- VAN LIESHOUT, P. H. H. M., HULSTIJN, W. & PETERS, H. F. M. (2004). Searching for the weak link in the speech production chain of people who stutters: A motor skill approach. In MAASSEN, B., KENT, R., PETERS, H. F. M., VAN LIESHOUT, P. H. H. M & HULSTIJN, W., *Speech motor control in normal and disordered speech*. Oxford: Oxford University Press, 313.
- YAIRI, E., AMBROSE, N. (2005). *Early Childhood Stuttering: for clinicians by clinicians*. Austin, TX: Pro-Ed.
- ZHARKOVA, N., HEWLETT, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *Journal of Phonetics*, 37, 248.
- ZHARKOVA, N., HEWLETT, N., & HARDCASTLE, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. In *Motor Control*, 15, 118.

STEFANO NEGRINELLI<sup>1</sup>, GIULIA DONZELLI<sup>1</sup>

<sup>1</sup>Universität Zürich

## Il progetto *AIS, reloaded*: un archivio sonoro per 36 varietà dialettali della Svizzera meridionale

‘AIS, reloaded’ è un progetto finanziato dal Fondo Nazionale Svizzero per la Ricerca Scientifica [100012\_162482<sup>1</sup>] in corso dal settembre 2016 presso l’Università di Zurigo e diretto dal Prof. Dr. Michele Loporcaro e dal Prof. Dr. Stephan Schmid. Avvalendosi di tecnologie avanzate per la raccolta, la digitalizzazione, la gestione e la presentazione di dati linguistici, il progetto ha come obiettivo quello di mettere a disposizione dei ricercatori due grandi serie di dati linguistici raccolti a quasi 100 anni di distanza; verrà così incentivata e facilitata l’analisi dell’evoluzione diacronica di alcuni dialetti della Svizzera italiana e del Canton Grigioni tramite la comparazione fra il materiale dello *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS) e i nuovi dati raccolti sul campo.

Il progetto, di durata triennale, si concentra sulle 36 località della Svizzera meridionale indagate nell’AIS originale: 18 dialetti lombardi e 18 varietà romance suddivisi tra i Cantoni Ticino e Grigioni. La raccolta dei dati linguistici è avvenuta direttamente sul campo: in ogni località si è resa necessaria una permanenza di circa tre giorni in quanto, a ogni singolo informatore, è stato somministrato l’intero questionario AIS (nella versione “normale”), che conta all’incirca 1800 entrate, per una media di circa 12 ore di registrazione audio-video a inchiesta. La trascrizione manuale immediata delle risposte ottenute dalla viva voce dei parlanti, caratterizzante l’AIS originario, è oggi sostituita da un metodo di raccolta dati che convoglia le registrazioni audio-video in una banca dati informatica. Il risultato di tale confluenza di dati è un vero e proprio archivio sonoro delle risposte date dagli informatori contemporanei al questionario AIS, archivio che sarà reso liberamente consultabile in rete e, soprattutto, interrogabile secondo differenti criteri di selezione pensati *ad hoc* per ricerche di carattere fonetico, sintattico, etnografico e storico-linguistico.

La registrazione delle entrate del questionario è attuata mediante un’interfaccia audio (USBPre 2) collegata a un microfono a cravatta Sennheiser MKE 2 (direttività onnidirezionale, gamma di frequenza di 20-20.000 Hz  $\pm$  23dB e coefficiente di trasmissione a vuoto di 10 mV/Pa  $\pm$  2.5 dB), così da garantire un’elevata qualità di registrazione. Le sessioni d’inchiesta sono state inoltre interamente registrate con una videocamera, così che resti traccia anche delle

---

<sup>1</sup> Una descrizione più approfondita del progetto è disponibile sul sito dell’università di Zurigo all’indirizzo web: <http://www.rose.uzh.ch/de/forschung/AIS-reloaded.html>.

frequenti spiegazioni delle sfumature semantiche e delle eventuali informazioni etnografiche visive fornite dagli informatori, per poter arricchire tale archivio sonoro con una eventuale sezione audio-visiva. La totalità delle registrazioni è stata raccolta grazie all'utilizzo del programma di registrazione *SpeechRecorder* (Draxler, Jänsch, 2004), versione 2.10.16; in una primissima fase progettuale è stato, infatti, scansionato l'intero questionario AIS cartaceo, così da esportare in un secondo momento un file in formato *.txt* utilizzato come base per la creazione di un 'progetto *SpeechRecorder*'. L'utilizzo di questo programma di registrazione permette il salvataggio di ciascuna sequenza sonora, sia essa una singola parola o un'intera frase, in un file sonoro a sé stante e l'assegnazione automatizzata di un codice numerico univoco per ciascuna registrazione, una sequenza numerica che presenta informazioni relative alla località, al parlante e alla specifica entrata del questionario, così da rendere più semplice e immediato l'inserimento delle stesse in una banca dati e il loro recupero da parte degli utenti. Il programma in questione ha inoltre permesso non solo di mostrare in video il questionario ai parlanti, abbattendo così i tempi di somministrazione del questionario stesso, ma anche l'uso di materiale visivo (immagini, schemi) a supporto della parola, fondamentale in luce soprattutto del divario culturale creatosi inevitabilmente nello scorrere del tempo.

Il risultato finale delle inchieste odierne sarà dunque un archivio sonoro che renderà disponibile agli utenti, calcolando una media di 1600-1700 file per varietà dialettale, quasi 60'000 registrazioni audio. La banca dati online sarà poi completata con le trascrizioni in alfabeto fonetico (IPA) delle registrazioni odierne: in questo modo sarà possibile una ricerca anche a partire dalla forma scritta, rendendo così il processo di estrapolazione del dato più immediato e diretto. I dati (trascrizioni fonetiche, file audio, spezzoni di file video, informazioni biografiche degli informatori) saranno organizzati in un archivio che verrà ospitato sui server dell'Università di Zurigo, mentre i filmati integrali delle inchieste saranno conservati presso l'Archivio Fonografico dell'Università di Zurigo.

Parallelamente alle inchieste sul campo procede l'attività di informatizzazione di parte dei dati raccolti tra il 1919 e il 1928. Il lavoro di digitalizzazione delle mappe AIS ha base e origine in *NavigAis* (Tisato, 2009<sup>1</sup>), una piattaforma multimediale a libero accesso attraverso la quale è possibile scorrere e visionare tutto il materiale AIS precedentemente scansionato. Al programma di consultazione si affianca un sistema di riconoscimento ottico dei caratteri (OCR), progettato all'uopo da Graziano Tisato, in grado, attraverso un processo semi-automatico, di trasformare in stringhe Unicode tutte le trascrizioni originariamente scritte a mano. In questo modo si intende digitalizzare all'interno del nostro progetto il 50% del corpus delle trascrizioni originarie, così da rendere

---

<sup>1</sup> Rimandiamo, per un confronto, direttamente al sito: <http://www3.pd.istc.cnr.it/navigais-web/>.

tali dati non solo consultabili visivamente ma anche, e soprattutto, interrogabili, sul modello di quanto già realizzato per il Veneto con l'Atlante Multimediale dei Dialetti Veneti (AMDV, Tisato et al., 2014).

La finalità del progetto è quindi duplice: da un lato abbiamo la via che porterà alla creazione di un archivio audio-visivo interrogabile e liberamente accessibile online, che si presenterà sotto forma di un atlante multimediale; dall'altro abbiamo invece la linea della digitalizzazione di quanto raccolto negli anni '20 del secolo scorso. Il ricongiungimento dei due binari renderà possibile un rapido confronto delle trascrizioni fonetiche delle inchieste degli anni '20 con quelle del triennio 2016-2019 relative ai 36 punti della Svizzera italiana e del Canton Grigioni. In merito dunque alle prospettive di ricerca offerte da archivi sonori di questo genere per le scienze della voce troviamo molto interessante la possibilità di accedere alle mappe *NavigAis* partendo dalle trascrizioni e soprattutto ai file audio (in formato .wav), cosa che permetterà agli studiosi, come in parte già avvenuto (si vedano a tal proposito ad esempio Pescarini, Donzelli, 2017; Negrinelli, 2018; in parte anche Bernardasci, Negrinelli, in stampa), di sfruttare agevolmente tali dati anche nei diversi ambiti d'indagine delle scienze del parlato.

#### *Riferimenti bibliografici:*

AIS = JABERG, K., JUD, J. (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 voll. Zofingen: Ringier & Co.

AMDV = TISATO, G. ET ALII (2014). *Atlante Multimediale dei Dialetti Veneti*. Fondazione Cassa di Risparmio di Padova e Rovigo - Università di Padova - ISTC-CNR.

BERNARDASCI, C., NEGRINELLI, S. (in stampa). Analisi fonetiche in due dialetti lombardo-alpini: parlato spontaneo e parlato controllato a confronto. Contributo accettato per gli *Atti del XIV Convegno Nazionale AISV, Bolzano 25-27 gennaio 2018*.

DRAXLER, C., JÄNSCH, K. (2004). *SpeechRecorder*.

<https://www.bas.uni-muenchen.de/forschung/Bas/software/speechrecorder/> (consultato il 29.09.2018).

NEGRINELLI, S. (2018). Le ostruenti palatali nell'arco alpino: primi dati dal progetto *AIS Reloaded*. In: MARCATO, G. (Ed.). *Dialetto e società. Presentazione di lavori in corso*. Padova: CLEUP, 33-40.

PESCARINI, D., DONZELLI, G. (2017). La negazione nei dialetti della Svizzera italiana. *Vox Romanica*, 76, 74-96.

<http://www.rose.uzh.ch/de/forschung/AIS-reloaded.html> (consultato il 29.09.2018).  
<http://www3.pd.istc.cnr.it/navigais-web/> (consultato il 29.09.2018).

MARIA DI MARO<sup>1</sup>, GIUSY GAUDINO<sup>1</sup>, FRANCESCO CUTUGNO<sup>1</sup>

<sup>1</sup>Università degli Studi di Napoli “Federico II”

## Conversing on Artistic Representation Topics (CART): an Android Audio Guide with Dialogic Skills

With the present work we intend to describe the architecture of an interactive audio guide (CART)<sup>1</sup> 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.

The starting point of the development was collecting materials for textual and spoken contents to be added to the application. To fulfil this requirement, the audio-visual corpus of guided tours collected for the project was used (Self citation, 2018). Precisely, the corpus contains audio-visual material collected from guided tours at San Martino Charterhouse in Naples. For this purpose, three art history experts were recruited to give onehour-long tours to four groups of four persons. From each tour, spoken and paralinguistic materials were recorded. Furthermore, each visitor was provided with a questionnaire with Likert scale evaluations and open questions (Self citation, 2018). The spoken data and the suggestions coming from the questionnaires were used to model the contents to be inserted in the application.

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. As a matter of fact, the content can be further explored by asking for information. Differently from other dialogue systems, which make use of different separated modules to be trained – specifically Automatic Speech Recognition, Natural Language Understanding, Dialogue Manager, Natural Language Generation and Speech Synthesis – the presented system can be classified as an end-to-end system. This kind of systems does not consider the separation into different components. Conversely, it indeed considers them as a single system to be trained as a whole (Vinyals et al., 2015, Serban et al., 2018). In particular, *DialogFlow*<sup>2</sup>, which uses Machine Learning algorithms to enable the system to communicate with users based on specific modelled *intents*, was here used for training. Each *intent* is made up of training sentences given to the system to learn to recognize the intention

---

<sup>1</sup> The acronym refers to the Chartusian logo.

<sup>2</sup> DialogFlow: <https://dialogflow.com/>

of the user, and textual answers to be generated. 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*, *artist*, *sculptor*, *architect*, *author*, *creator*, *designer*, *inventor*, *master*. The use of annotation is advantageous to correctly classify an intent uttered with variable terms, and to improve disambiguation processes.

The dialogue was structured on the basis of the information added to the audio-guide, making use of different strategies. On 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, some details were left pending with rhetorical or provocative questions, and highlighting pieces 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 exclamation points represent the missing details that can be asked to the system, and specifically *Which other frescos are situated in the Parlor? What's the name of the fresco representing St. Ugo? What does it mean? What do the seven stars 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?*

The application was developed using Android Studio<sup>1</sup>. The developed interface for the Museum was then enriched with the dialogue modelled in DialogFlow using Android SDK for DialogFlow<sup>2</sup>, to which Text-toSpeech technologies were integrated to give answers to user queries.

Besides the audio and vocal channels, the multimodal nature of the application also made use of the visual one, as comparisons with other similar environments, work of arts, or Charterhouses were visually shown, when requested or implied in the offered content. The future application of this system is intended to be further developed to comprise users' position knowledge. This will enable the system to use spatial references and deitics (i.e. spatial, temporal and personal pronouns) focusing the listener's attention to a specific object of the specific environment, thus enriching the linguistic potential of the system itself.

---

<sup>1</sup> Android Studio: <https://developer.android.com/studio/>

<sup>2</sup> Android SDK for DialogFlow: <https://github.com/dialogflow/dialogflow-android-client>

The application was tested on 10 users, who made use of the interactive audio guide in the modelled areas of the Charterhouse, and precisely the Pronaos, the Great Cloister, the Parlor, the Chapter Hall, the Choir, and the Treasure Hall. After the visit, users were asked to fill a questionnaire 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 et al. 2010). With the additional use of open questions we were able to collect opinions and impressions to get the usefulness of such systems and how to improve them. As a matter of fact, the generically positive results were paired up with useful suggestions, such as returning vivid and significant contents which are not too long-winded, as the attention tends to decrease when listening at someone talking without having that someone in the same room.

### *Riferimenti bibliografici*

FIEDLER M., HOSSFELD T. & TRAN-GIA P. (2010). "A generic quantitative relationship between quality of experience and quality of service". *IEEE Network*, 24(2).

SELF CITATION (2018).

SERBAN I. V., LOWE R., HENDERSON P., CHARLIN L. & PINEAU J. (2018). "A Survey of Available Corpora For Building Data-Driven Dialogue Systems: The Journal Version". In *Dialogue & Discourse*, 9(1), 1-49.

TODD S. (2013). *Learning desire: Perspectives on pedagogy, culture, and the unsaid*. Routledge.

VINYALS O., LE. Q (2015). *A neural conversational model*. arXiv preprint arXiv:1506.05869.

SILVIA CALAMAI<sup>1</sup>, GRETA BIANCHINI<sup>1</sup>, LUCILLA GIGLI<sup>1</sup>, CATERINA PESCE<sup>2</sup>, MARICA SETARO<sup>1</sup>, PETRA VANGELISTI<sup>1</sup>

<sup>1</sup> Università degli Studi di Siena, <sup>2</sup> Università degli Studi di Padova

## **Ritrovare voci. Il lavoro intorno all'archivio sonoro di Anna Maria Bruzzone**

Questo intervento prende le mosse 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). La vita pressoché secolare di questa cittadella manicomiale toscana ha edificato, nel tempo, un vero e proprio palinsesto fatto di spazi fisici, fonti, documenti, memorie, testimonianze, ricordi che hanno innervato il tessuto urbano circostante, il mondo dentro diviso dal mondo fuori. Un palinsesto che, in una logica di apertura trasformativa e grazie a politiche oculate, si è saputo convertire in una cittadella universitaria. Una consolidata abitudine – legata ad un'iperproduzione di natura scritta – fa pensare agli archivi come a un luogo fisico pieno di carte. Anche gli archivi degli ex ospedali psichiatrici non si sottraggono a questa descrizione. Ma sarebbe 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 registrazioni audio, filmati, testimonianze dirette rese da chi ha ricordato o ricorda gli ultimi anni di funzionamento del manicomio prima della totale dismissione. Lo scarto significativo è prodotto dall'esperienza pirelliana: il manicomio abbatte i suoi muri e chiama il mondo fuori a vedere l'orrore chiuso dentro. Arrivano intellettuali, artisti, giornalisti, fotografi. E ad Arezzo arriva Anna Maria Bruzzone (1925-2015), autrice del volume *Ci chiamavano matti. Voci da un ospedale psichiatrico*, pubblicato da Einaudi nel 1979. 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.

L'Archivio sonoro raccolto da Anna Maria Bruzzone, dato per disperso, è stato ritrovato a Torino nel 2016 e donato all'Università di Siena nel 2017. Esso è stato digitalizzato grazie al contributo della Soprintendenza Archivistica e Bibliografica della Toscana ed è ora in fase di catalogazione e di studio da varie

prospettive disciplinari, all'interno di un progetto di ricerca che vede uniti linguisti, storici orali, filosofi, archivisti.

L'intervento presenta lo stato di avanzamento dei lavori ed è suddiviso in quattro parti. Una prima parte descrive la consistenza dell'archivio: esso si compone di 36 audiocassette; 9 quaderni "Trascrizioni letterali"; 2 fascicoli di trascrizioni dattiloscritte (suddivise in "Donne", n. 33, "Uomini" n. 32). Segue una sezione dedicata alle questioni etiche e legali associate all'eventuale riuso di fonti documentali prodotte da soggetti vulnerabili, alla luce del Regolamento europeo in materia di protezione dei dati personali, entrato in vigore a partire dal 25 maggio 2018.

La terza parte affronta le fonti orali raccolte incrociando una prospettiva storica con quella sincronica e muove da un lavoro svolto nell'ambito di una tesi dottorale. Le voci recuperate dai nastri ritornano dal passato per incontrare il presente dei soggetti che le hanno pronunciate. Attraverso la rete di ex operatori dell'ONP (infermieri, assistenti sociali, medici) è stata avviata una campagna di ricerca di questi soggetti (o loro eredi) che ha consentito di rintracciare alcuni ex pazienti presenti nelle interviste del '77 di Bruzzone. Le interviste condotte oggi con gli ex pazienti seguono un questionario costruito precedentemente e volto a raccogliere dati relativi alla vita dell'intervistato e alla sua esperienza manicomiale. Successivamente, gli intervistati hanno ascoltato la loro intervista resa a Bruzzone nel '77. Analoghe interviste sono state fatte a medici, infermieri e assistenti sociali per ricostruire le personalità e le esperienze di chi prese parte al processo antiistituzionale guidato da Agostino Pirella.

L'ultima parte dell'intervento è dedicata all'analisi di alcune delle registrazioni recuperate. In primo luogo sono commentati esempi di schede catalografiche. Oltre a contenere le informazioni sulla fonte sonora dell'intervista e i dati biografici del parlante, le schede presentano una sintesi del dialogo, suddivisa in aree tematiche, una lista di parole chiave e un breve commento linguistico-contenutistico, che fornisce informazioni anche sul repertorio linguistico usato nell'intervista e sulla modalità di scambio dialogico documentato nella registrazione. In secondo luogo si offrono i risultati di una analisi comparativa tra il testo a stampa e il documento sonoro originale dell'intervista, a partire dal caso del paziente F. In particolare, vengono commentati gli interventi di carattere più macroscopico compiuti dall'autrice come per esempio i tagli di contenuto (soprattutto, quelli che si riferiscono all'esperienza pirelliana), gli interventi più minuziosi come quelli relativi al trattamento dei dati sensibili (sistematica espunzione dei dati personali, o loro profonda modifica). Infine, vengono mostrate le soluzioni adottate in merito alla rappresentazione di fatti tipici del parlato come false partenze, ripetizioni e sequenze di autocorrezione, e vengono ispezionate le varie scelte adottate da Bruzzone in merito alla resa (o alla "non resa") delle caratteristiche dialettali, sia genericamente toscane che più specificamente del dialetto aretino.

### *Riferimenti bibliografici*

BRUZZONE, A. M.(1979). *Ci chiamavano matti. Storie 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 M. BAIONI, M. SETARO (a c.di), *Asili della follia. Storie e pratiche di liberazione nei manicomi toscani*. Pisa: Pacini, 22-34.

CALAMAI, S., KOLLETZEK, C. & KELLI, A. (2018). Towards a protocol for the curation and dissemination of vulnerable people archives. In *CLARIN Conference 2018* <https://www.clarin.eu/clarin-annual-conference-2018abstracts#T>

FOOT, J. (2014). *La "Repubblica dei matti". Franco Basaglia e la psichiatria radicale in Italia, 1961-1978*. Milano: Feltrinelli.

[GDPR] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119, 4.5.2016, 1-88.

<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1515793631105&curi=CELEX:32016R0679> (29.09.2018)

MIKI MORI<sup>1</sup>

<sup>1</sup>Centre Universitaire de Mayotte

## **The CoPLACo Project: Creating a Multi-use, Digital Platform of Oral Language for the Comorian Islands**

This poster will present an ongoing project, CoPLACo (Corpus of Linguistic Heritage of the Comorian Archipelago), which aims to create a multi-use digital platform of oral language for research, teaching and archival purposes. The inhabitants of the Comorian islands, situated between Mozambique and Madagascar, speak a variety of Shikomori. A mainly oral language, little documentation exists on dialectal variation and few linguistic archives are available for Shikomori. The CoPLACo proposes to be an open site where researchers, educators and community members can go to access a variety of oral language observed in the archipelago.

Through inter-island collaboration using the CoPLACo, community members can better understand their language, manage their heritage archives and reflect upon language planning and teaching issues. Given the oral preference of this language, the corpus also entails transcription of the archive through a unified orthographic system, which has yet to be seen. Finally, the corpus will encourage linguists to conduct various types of research on the language, including issues related to phonology, morphology and syntax.

This poster will present the background of the project while elaborating on data collection and field methods, including types of oral data to be gathered and solicitation methods. In addition, the poster will discuss the structure of the corpus, including the interface, creation of an app, and the various components of the website, such as sections on teacher resources, transcription practices, and field methods. Future collaborations with community members through workshops will also be discussed.

SONIA I. D'APOLITO<sup>1</sup>, BARBARA GILI FIVELA<sup>1</sup>

<sup>1</sup>Università del Salento

## **Interazione contesto-accuratezza del parlato L2: studio pilota sulle produzioni di geminate in italiano L2 da parte di parlanti francofoni**

Un parlato accurato, chiaro e con uno stile di eloquio adeguato (senza troppe esitazioni, ripetizioni e/o lunghe pause) permette sia all'apprendente di esprimere il proprio messaggio senza sforzo (Zhang, 2009) e quindi di farsi comprendere, che ai percipienti di comprendere il messaggio senza sforzo e ambiguità. Da un lato, lo studio dell'accuratezza nella pronuncia dei suoni non-nativi di una L2 è, quindi, molto importante, in quanto consente di osservare l'esito della produzione del parlante nell'ambito della comunicazione; peraltro, la produzione del parlante rende solitamente evidenti le caratteristiche che derivano dall'interazione dei sistemi fonetici-fonologici della L1 e della L2 (Flege & Hillenbrand, 1984; Flege *et al.*, 1999). D'altro canto, in letteratura si riporta che gli errori fonetici e fonologici hanno un impatto solo del 22% sull'intelligibilità del parlato L2 (Zhang, 2009; Smith & Nelson, 2006), benché il contesto in cui avviene la comunicazione possa fortemente influenzare le attese dei percipienti, come anche l'accuratezza in produzione (Lindblom, 1990).

Questo studio è parte di un progetto sull'italiano L2 parlato dagli apprendenti francofoni, volto a comprendere il ruolo dell'accuratezza all'interno della comunicazione. In particolare, in questo contributo l'obiettivo è quello di osservare: 1) l'interazione tra le caratteristiche dei sistemi foneticofonologici L1 e L2 nella produzione di geminate; 2) l'interazione tra contesto (contesto più/meno ricco di informazioni) e stile di eloquio (parlato controllato e semispontaneo) rispetto a: 1) l'accuratezza nella produzione di geminate; 2) lo sforzo degli apprendenti nel produrre suoni nonnativi e il suo impatto sulla loro fluency verbale.

L'accuratezza e lo sforzo nella produzione di geminate sono state osservate in diversi compiti di produzione, in cui le consonanti target sono state realizzate da apprendenti francofoni di italiano L2 con due livelli di competenza (principiante e avanzato), variando il contesto in base alla quantità di informazioni disponibili ed elicitando sia un parlato controllato che un parlato semispontaneo. Si ipotizza che la realizzazione delle geminate possa variare in base: 1) al livello di competenza degli apprendenti, in quanto ci si aspetta una maggiore accuratezza da parte dei parlanti di livello avanzato (Flege & Hillenbrand, 1984) al contesto e allo stile del parlato, perché ci si aspetta un

maggior numero di errori segmentali in un contesto più ricco di informazioni e meno controllato rispetto al contesto con minori informazioni e più controllato (Lindblom, 1990). In particolare, ci si aspetta una minor accuratezza nella realizzazione della geminata (minor durata), un maggior numero di pause/esitazioni in corrispondenza della parola target che include la geminata e uno sforzo maggiore in caso di contesto povero e più controllato rispetto al contesto con maggiori informazioni e meno controllato. Quattro apprendenti francesi di Italiano L2, due principianti e due di livello avanzato (e 2 italofoeni per controllo) hanno partecipato all'esperimento. Il corpus comprende quattro coppie minime per ciascun suono target (21 coppie minime) (Tab. 1). Per il parlato semispontaneo, gli apprendenti hanno effettuato due compiti: 1) naming di oggetti presentati attraverso le sole immagini (isolamento spontaneo; contesto povero; 20 parole – Tab.2 B1); e 2) descrizione di una sequenza di immagini che includono coppie minime (contesto ricco; 3 coppie minime – Tab.2 B2). Dopo quest'ultimo compito, ai parlanti è stato chiesto di giudicare il proprio sforzo nel distinguere le coppie minime (scala Likert a 9 punti). Per il parlato controllato, le parole con i suoni target (40 parole x 4 ripetizioni) sono state realizzate a inizio e fine frase in coppia minima nella lettura di un testo corrispondente a una interazione in contesto che: a) non facilita la disambiguazione (carrier sentence; contesto povero – Tab.2 C1); b) facilita la disambiguazione (coppia minima inserita in frase con contesto naturale; contesto ricco – Tab.2 C2).

I dati acustici sono stati segmentati in PRAAT (Boersma & Weenink, 2008) per: a) osservare la realizzazione delle geminate (intero fono per tutti i target) e il numero di esitazioni/pause; b) effettuare misurazioni acustiche per osservare: 1) durata e durata normalizzata (ratio tra durata segmento e durata enunciato/numero di sillabe) dei segmenti target; 2) durata vocale precedente; e 3) la fluenza verbale (es. velocità di eloquio/articolazione; durata delle pause/esitazioni).

Risultati preliminari riguardano sia dati acustici (durata segmento) che i giudizi degli apprendenti. Dal punto di vista acustico, gli apprendenti e soprattutto i principianti mostrano una riduzione delle geminate nella corrispondente consonante scempia. Quanto al contesto e allo stile di eloquio, una minore accuratezza (es. maggior riduzione) si riscontra nel contesto ricco di informazioni e meno controllato. Rispetto alla competenza, nei principianti si osserva un numero maggiore di casi di riduzione consonantica e di esitazioni in corrispondenza della parola target che include la geminata. Per quanto riguarda le autovalutazioni, la maggior parte degli apprendenti giudicano difficile la produzione delle geminate con uno sforzo maggiore per la vibrante.

Tabella 1 - I suoni dell'italiano e la possibile realizzazione in Italiano L2 da parte di apprendenti francofoni

Suoni italiano L2		Possibile realizzazione in francese
Geminate	/tt/, /dd/, /ss/, /nn/, /ll/, /rr/	Riduzione a scempia – ad eccezione di /rr/ poiché in francese l'opposizione /r/-/rr/ esiste in alcune coniugazioni verbali (es. imperfetto <i>vs</i> condizionale)

Tabella 2 - Esempio di corpus e dei compiti di produzione

<b>A.</b>		
<b>Suono target</b>	<b>Coppia</b>	<b>Parole</b>
/rr/	/rr/-/r/	Sera/serra - Bara/barra
<b>Parlato semi-spontaneo</b>		
<b>B.</b> 1. Naming	 <p>Sete sette canne cane palla pala</p>	
2. Descrizione sequenza	Immaginare di ordinare gli oggetti in una scatola	
<b>C. Parlato controllato</b>		
1. Interazione (contesto povero)	<b>I</b>	Cosa hai detto? – Dico <b>sera/serra</b> di nuovo a Maria
	<b>F</b>	Cosa hai detto? – Maria ha detto <b>sera/serra</b> di nuovo
2. Interazione (contesto ricco)	<b>I</b>	Cosa ti va di fare? – Questa <b>sera</b> vorrei andare al cinema
		A cosa serve? – Questa <b>serra</b> serve per le piante nuove
	<b>F</b>	Ti sei vista con Luca martedì? – Ci siamo visti la <b>sera</b> prima
		Dove le mettiamo le piante? – Le mettiamo nella <b>serra</b> verde

### Riferimenti bibliografici

BOERSMA, P., & WEENINK, D. (2008). *PRAAT: doing phonetics by computer*.

- FLEGE J. E., BOHN O. S., & MEADOR D. (1999). Native Italian speakers' production and perception of English vowels. In *Journal of Acoustical Society of America*, 106, 2973-2987.
- FLEGE, J. E., HILLENBRAND, J. (1984). Limits on pronunciation accuracy in foreign language speech production. In *Journal of the Acoustical Society of America*, 76, 708-721.
- LINDBLOM, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In HARDCASTLE, MARCHAL (Eds.), *Speech Production & Speech Modeling*, Dordrecht, 403-439.
- SMITH L. E., NELSON C., L. (2006). World Englishes and issues of intelligibility. In B. KACHRU, Y. KACHRU & C. NELSON (eds.), *The Handbook of World Englishes*. Blackwell Publishing Ltd, 428-445.
- ZHANG, S. (2009). The role of Input, Interaction and Output in the development of oral fluency. In *English Language Teaching*, Vol. 2 (4), 91-100.

DUCCIO PICCARDI<sup>1</sup>

<sup>1</sup> Università di Pisa

## **Bilingual Frequency in a Favorable Context (BFFC) in the Italian dialectal area. A protocol for the analysis of geminate lateral retroflexion in Antona (MS)**

From the end of the past century, usage-based approaches to the analysis of language variation have increasingly become widespread in experimental linguistic research. In particular, exemplar theory posits the storage of detailed eventive traces in the mind of the individual, comprehensive of linguistic, social and environmental features. In this framework, the structure of the mental clouds (e.g. Pierrehumbert, 2001) of exemplars is of central importance in determining the individual choices, revolving around the properties of salience, recency and density, i.e. frequency of exposure to similar events. One of the main question scholars had to answer in order to test the effectiveness of this model is what kind of frequency exemplar theory is concerned with. Bybee (2002) pointed out that researchers should not consider the raw frequency of a word as the main predictor of its role in a variation phenomenon; instead, since the linguistic components of an exemplar are purely phonetic, the frequency of a word in across-boundaries contexts of segmental variation is of greater concern for usage-based models. This idea remained largely untested until Raymond and Brown's proposal for a quantification method for this parameter, named Frequency in a Favorable Context (FFC) and calculated as the frequency of a word in a varying context divided by its overall frequency (Raymond & Brown, 2012). Another related issue refers to the role of individual frequency of use in shaping exemplar clouds. From the moment that each linguistic lifespan is indeed a «personal odyssey» (Foulkes, 2010: 19) of events, the effects of individual patterns should be proved at least comparable to those derived from a corpus. Drager (2011) helped clarifying this point, integrating months of ethnographic fieldwork with experimental phonetic inquiries. In her data, individual frequencies of occurrence of /l/, shaped by its context and grammatical function, had a significant role in triggering sociophonetic variation.

In 2015, Brown (2015) expanded her line of studies to bilingual speakers. The scholar examined a corpus of oral interviews in a bilingual (Spanish-English) community in New Mexico, searching for the effects of word frequency on Spanish /d/ weakening, mainly occurring in post-vocalic contexts. She thus revised the FFC formula proposing the Bilingual Frequency in a Favorable

Context (BFFC), apt to verify the weight of existing cognate (i.e. phonologically and semantically similar) English forms on the Spanish variation phenomenon at hand. In the BFFC, the corpus frequency of the English cognate is integrated as a non-variating context and thus added to the overall Spanish frequency of the word. The experiment proved that the English cognates had a strong impact in inhibiting the inquired variation, theoretically confirming that bilingual speakers cluster all their communicative events in a single exemplar cloud.

In this paper I argue that this last point is crucial for the adjustment of sociophonetic methodologies to the Italian linguistic landscape. Italian dialects are in fact «sisters' of Italian, locally divergent developments of the Latin originally spoken in Italy» (Maiden & Parry, 1977: 2); from a sociolinguistic point of view, it has been noted that the structural distance between the dialects and the standard language «justify treating the Italian situation as bilingual rather than merely bidialectal» (Berruto, 1997: 394). The probability of the emergence of local traits in the dialectal practice would thus be conditioned by the BFFC, with a negative influence exerted by Standard Italian cognates. The main hurdle to overcome in order to test the BFFC in an Italian setting is the impossibility to extract dialectal frequencies from an adequate corpus. For this reason, I suggest using previously retrieved subjective frequency estimates for the dialectal component of the formula, in line with the above discussed compatibility between individual and corpus-grounded values. Subjective estimates have been proved to be closely correlated to objective, corpus-based frequencies, both from the theoretical (McGee, 2008) and experimental (Thompson & Desrochers, 2009) points of view; moreover, reliable measures can be built from a reasonably small percentage (Thompson & Desrochers, 2009; Balota *et al.*, 2001) of the represented population. The standard protocol for collecting subjective estimates has been established by Balota and colleagues (Balota *et al.*, 2001): interviewees are simply asked to rate on a Likert scale the frequency of their encounters with specific words, ranging from 1=never to 7=several times a day.

After a review of the relevant literature, this paper focuses on the steps of a protocol from a research in progress on geminate lateral retroflexion (traditionally transcribed as -ll- > -**ḏḏ**-) in Antona (MS). This feature, first described by Bottiglioni (Bottiglioni, 1911), covers the area of the Apuan Alps, not without internal variability (Savoia, 1980). In Antona, a small mountain village in the Frigido valley above the chief-town of Massa, lateral retroflexion is today the flagship feature of local linguistic identity, in a general picture of gradual loss of dialectal competence and consequent advergence (Berruto, 2005: 82) of the dialect to Standard Italian. For the sake of simplicity, in this explorative study the phenomenon will be observed only in word-internal positions, excluding syntactic doubling and equating the dialectal FFC value to the overall frequency of the word. The protocol consists in: a) extraction of singular nouns, ordered by frequency of occurrence and containing geminate laterals, from an Italian corpus; b) submission of the list to a small number of native informants, in order to

exclude the Italian nouns that do not share the cognate status with their dialectal counterparts; c) submission of the derived list of nouns to a socially distributed group of native informants to assess subjective frequency estimates, revising Balota and colleagues' instruction with the indication of the locality (i.e. Antona) of the encounters with the noun; d) calculation of the BFFC as the mean of the subjective estimate ratings divided by the same value plus the log. objective frequency of the Italian cognate; e) selection of a subset of nouns, equally distributed by their BFFC ratings; f) pairing of these nouns with graphic representation of their referents, using a corpus of images built for psycholinguistic testing (e.g. Szekely *et al.*, 2004); g) picture naming task submitted to a socially distributed group of native interviewees, specifying the informal communicative context of the simulation. If BFFC has an impact on the emergence of dialectal variants, a correlation between its values and the number of elicitations of lateral retroflexion in the task is expected. At present time, a list of 185 singular nouns containing a geminate lateral has been extracted from ItWac (Baroni *et al.*, 2009) through a case-insensitive query. The list has been presented to three informants that have cancelled the non-cognate nouns, writing down their dialectal equivalents (e.g. It. pipistrello  $\neq$  Ant. parpagon; see Mosti & Nancesi, 2005: 378). Given the young age of these informants, this phase of the research still needs more feedbacks before moving to the subjective frequency estimates. In fact, we could expect that older speakers retain a wider pool of dialectal lexemes, adding depth to our description of the interaction between the two linguistic systems in Antona.

#### *Riferimenti bibliografici*

- BALOTA, D. A., PILOTTI, M., & CORTESE, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. In *Memory & Cognition*, 29, 639-647.
- BERRUTO G. (1997). Code-switching and code-mixing. In M. MAIDEN, M. PARRY (eds.), *The Dialects of Italy*. London and New York: Routledge, 394-400.
- BERRUTO, G. (2005). Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In P. AUER, F. HINSKENS & P. KERSWILL (eds), *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: University Press, 81-95.
- BOTTIGLIONI, G. (1911). Dalla Magra al Frigido. Saggio fonetico. In *Revue de Dialectologie Romane*, 3, 77-143.
- 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.
- 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, 261-90.
- DRAGER, K. (2011). Sociophonetic variation and the lemma. In *Journal of Phonetics*, 39, 694-707.

- FOULKES, P. (2010). Exploring social-indexical knowledge: A long past but a short history. In *Laboratory Phonology*, 1(1), 5-39.
- M. BARONI, S. BERNARDINI, A. FERRARESI & E. ZANCHETTA (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In *Language Resources and Evaluation*, 43(3), 209-226.
- MAIDEN, M., PARRY, M. (eds, 1997). *The Dialects of Italy*. London and New York: Routledge.
- MCGEE, I. (2008). Word Frequency Estimates Revisited—A Response to Alderson (2007). In *Applied Linguistics*, 29(3), 509–514.
- MOSTI, E., NANCESI, M. (2005). *Dizionario enciclopedico del vero dialetto massese*. Massa: CRD.
- PIERREHUMBERT, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. BYBEE, P. HOPPER (eds), *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins, 137-157.
- 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 S. TH. GRIES, D. S. DIVJAK (Eds.), *Frequency Effects in Language. Vol 2: Learning and Processing*. The Hague: Mouton de Gruyter, 35-52.
- SAVOIA, L. (1980). Fonologia delle varietà apuane e garfagnine: consonantismo. In *Studi Urbinati di Storia, Filosofia e Letteratura*, II, 233-293.
- SZEKELY, A., JACOBSEN, T., D'AMICO, S. et al. (2004), A new on-line resource for psycholinguistic studies. In *Journal of Memory and Language*, 51(2), 247-250.
- 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.

CARLO BIANCHI<sup>1</sup>

<sup>1</sup> Associazione Italiana di Storia Orale (AISO)

## Interviste orali su musica e storia. Un'ipotesi di archivio consultabile fra problemi scientifici ed etici

Lungo la mia attività di musicologo scrittore, la forma dell'intervista si ritrova sia in parecchi scritti divulgativi, sia in alcuni saggi comparsi su riviste accademiche (o 'scientifiche'). Dopo oltre vent'anni, conservo una certa quantità di registrazioni (la maggior parte digitalizzate) delle conversazioni svolte con vari personaggi del mondo della musica classica e leggera che hanno parlato di musica, ma in veri modi anche di storia. Realizzate in varie circostanze, alcune preparate e ufficiali, altre fortunate o confidenziali, le interviste si inseriscono in quel filone della storia orale che indaga il rapporto fra intervistatore e intervistato e fra il dialogo orale e la sua trascrizione<sup>1</sup>. Intendo concentrarmi sul secondo aspetto, ossia sulle problematiche che insorgono nelle operazioni di trascrizione-redazione-pubblicazione-analisi di una conversazione (o *performance dialogica* come sarebbe meglio definirla, specie nel caso di un dibattito pubblico, conferenza etc.) e che conducono infine a una sua rappresentazione.

Onde rendere giustizia dei significati e della dimensione estetica della *performance*, la lingua della trascrizione non è mai del tutto coincidente con quella del discorso orale, per necessità di una interpunzione che deve relazionarsi alle pause e alle intonazioni,<sup>2</sup> ma sovente per ragioni di macro-forma e di ricostruzione del senso complessivo di certi enunciati, che richiedono una certa libertà quando non addirittura inventiva da parte dell'intervistatore-trascrivente. In ambito linguistico, le due forme, orale e scritta, vengono considerate due varietà distinte della lingua. In una prospettiva storiografica e filologica, di interpretazione delle fonti, è auspicabile che le trascrizioni-pubblicazioni siano accostate alla rispettiva fonte orale. Quali conseguenze potrebbe avere dunque l'apertura al pubblico di questo mio archivio digitale, diventando consultabile da altri studiosi o utenti generici? Indico qui tre problematiche, riguardo a [1] medesimi passaggi del dialogo interpretabili in modi diversi [2] passaggi omessi

---

<sup>1</sup> PORTELLI, A. (2007). *Storie orali. Racconto, immaginazione, dialogo*. Roma: Donzelli, 75-94.

<sup>2</sup> AVESANI, C., VAYRA, M. (1992). *Discorso, segmenti di discorso e un'ipotesi sull'intonazione*. In *Storia e teoria dell'interpunzione*, Atti del convegno internazionale di studi, Firenze 19-21 maggio 1988. Roma: Bulzoni, 355-408.

dalla trascrizione [3] etica legata al *placet* degli intervistati necessario per rendere l'intervista consultabile

[1] Se inevitabilmente si apre una zona di parole trascrivibili in modi differenti, e quindi l'ascolto di quelle conversazioni potrebbe portare a versioni differenti da quelle già pubblicate, questo potrebbe portare interpretazioni meno efficaci, oppure implicare nuove acquisizioni: dati, concetti, o anche solo sfumature uditive della *performance* che potrebbero essermi sfuggite. Se una frase è stata pronunciata dall'intervistato con un tono enfatico che non si coglie dalla trascrizione, va persa una parte del significato che l'ascolto della registrazione originale può invece restituire in modo completo. Cito l'esempio di una intervista che ho realizzato con Vittorio Nocenzi nel marzo 2017 in occasione di un concerto bresciano del Banco del Mutuo Soccorso. Intervista poi pubblicata sul Bimestrale «BresciaMusica»<sup>1</sup>. L'ultima mia domanda che si può leggere nella penultima colonna, “c'è qualcosa di quegli anni [Settanta] che allora non percepivi e di cui ti sei reso conto in seguito?”, è seguita dalla risposta: “In realtà eravamo all'interno di un movimento con delle caratteristiche di cui a volte non ci accorgevamo nemmeno noi...” Tuttavia, durante la nostra conversazione, la mia domanda un po' informale era stata: “a distanza di quarant'anni c'è qualcosa che... potresti dire: 'allora non me ne rendevo conto, e adesso invece sì, se ci ripenso'...?” A quel punto, Nocenzi ha risposto “... sì... una magia strana ... che ha continuato ad affascinare migliaia di giovani anche dopo quei fatti”. Ma nella trascrizione, questa risposta non giunge subito dopo la mia domanda, bensì dopo quell'altra frase di Nocenzi, che ho anteposto per conferire ancor più coerenza a ciò che lui voleva dire. Il punto è che di fronte alla mia domanda, quella risposta, a caldo, “... sì.. una strana magia...” è stata pronunciata da Nocenzi con un tono riflessivo e nostalgico che non si coglie dallo scritto – per quanto l'importanza di quella frase risalti infine, o prima di tutto, dal titolo che ho dato al servizio. L'ascolto della registrazione (che intenderei proporre dal vivo al convegno) può in tal senso illuminare un significato che rimane parzialmente nascosto. Il tono con cui Nocenzi ha risposto alla mia domanda lascia capire che in quel momento ho toccato una sua corda emotiva profonda – il mio input ha sollecitato l'attivazione di una memoria sensoriale, come direbbe Bruno Bonomo<sup>2</sup>. Bisognerebbe riprodurre l'espressione del volto di Nocenzi, nel momento in cui risponde, nonché l'attimo prima, in cui pensa al significato della domanda e al concetto che ho portato alla sua attenzione. Solo quel punto, si otterrebbe un'impressione del tutto rispondente – ma questo non è possibile perché la registrazione era solo audio.

---

<sup>1</sup> BIANCHI, C. (2017). Quella strana magia degli anni Settanta. Vittorio Nocenzi a Brescia con il “Banco”. In *BresciaMusica*, 153, 10-11.

[https://www.academia.edu/35022436/Quella\\_strana\\_magia\\_degli\\_anni\\_Settanta\\_Vittorio\\_Nocenzi\\_a\\_Brescia\\_con\\_il\\_Banco](https://www.academia.edu/35022436/Quella_strana_magia_degli_anni_Settanta_Vittorio_Nocenzi_a_Brescia_con_il_Banco)

<sup>2</sup> BONOMO, B. (2015). *Voci della memoria. L'uso delle fonti orali nella storia*. Roma: Carocci, 94.

[2] Nel mio caso, le conversazioni originali contengono talvolta interi passaggi che non sono stati trascritti, non rientrando nella necessità della pubblicazione del momento. Sempre durante quella conversazione con Nocenzi abbiamo poi discusso di tante altre cose, musicali e umane, di canzoni e poesia, fino al Nobel conferito a Bob Dylan. Quest'ultimo argomento si trova parzialmente in un'altra intervista sempre per «BresciaMusica»<sup>1</sup>, ma altri discorsi rimangono solo nella registrazione. Riflessioni di Nocenzi che appunto non ritenni fondamentali nell'economia della pubblicazione, ma potrebbero essere oggetto di riflessioni future. Nel campo dei repertori classici, una intervista di cui non ho riportato tutta la conversazione per i medesimi motivi di cui sopra, è quella con Luigi Pestalozza, nel ricordare il compositore Karl Amadeus Hartmann<sup>2</sup>, e quella di Furio Colombo ospite alla manifestazione *Le Dieci giornate di Brescia* nel 2008 nella giornata dedicata al tema “la musica e l'odio”<sup>3</sup>.

[3] La perizia dell'intervistatore risiede nel trascrivere senza travisare il significato della narrazione anche quando interviene con la propria inventiva. Onde assicurarmi che ciò non avvenisse, la mia preoccupazione è stata quella di sottoporre sempre all'intervistato di turno la trascrizione, quando possibile, e così ottenere il *placet* per la pubblicazione. Talvolta gli intervistati hanno assenti incondizionati, talvolta hanno richiesto piccole modifiche, talvolta modifiche più sostanziali. Eppure: certe cose sostanziali che ho dovuto eliminare erano, a mio modo di vedere, fra le più interessanti. Ad esempio, un noto musicista classico italiano (di cui non posso fare qui il nome) mi disse a proposito del rapporto fra il livello di istruzione italiana e quello europeo e internazionale: “ai concorsi internazionali, i nostri ragazzi arrivano che in genere sono meno preparati degli altri. Non vorrei che questo suonasse come un'accusa nei confronti del sistema didattico italiano, ma qualcuno deve avere il coraggio di dirlo”. Salvo poi chiedermi di eliminare questa asserzione – quando gli sottoposi la trascrizione – onde evitare polemiche.

Il problema in generale è che se l'intervistato chiede di modificare od omettere qualcosa, il rispetto della sua volontà può allo stesso tempo generare lacune o distorsioni rispetto al significato della fonte orale. Errori ermeneutici rispetto all'orizzonte personale dell'intervistato o della comunità sociale intorno a lui. Il problema non è solo scientifico, ma anche morale e giuridico legato all'art. 8 del Codice di deontologia e di buona condotta per i trattamenti di dati personali con

---

<sup>1</sup> BIANCHI, C. Letteratura senza dubbio. Vittorio Nocenzi sul Nobel a Bob Dylan. In *BresciaMusica*, 154, 17. [https://www.academia.edu/35022438/Letteratura\\_senza\\_dubbio.\\_Vittorio\\_Nocenzi\\_sul\\_Nobel\\_a\\_Bob\\_Dylan](https://www.academia.edu/35022438/Letteratura_senza_dubbio._Vittorio_Nocenzi_sul_Nobel_a_Bob_Dylan)

<sup>2</sup> ID. *Musica contro la dittatura. Luigi Pestalozza ricorda Karl Amadeus Hartmann* [Intervista inedita 17.11.2011 – unpublished draft] [https://www.academia.edu/36664495/Luigi\\_Pestalozza\\_ricorda\\_K.A.\\_Hartmann](https://www.academia.edu/36664495/Luigi_Pestalozza_ricorda_K.A._Hartmann)

<sup>3</sup> ID. (2006). La musica e l'odio. All'interno de “Le Dieci giornate di Brescia” un intervento di Furio Colombo. In *BresciaMusica*, 101, 15.

finalità storiografiche. Nel momento in cui si realizza un'intervista da pubblicare, o si costruisce un archivio orale consultabile, è necessario ottenere l'approvazione dell'intervistato (oltre che dell'intervistatore) "che naturalmente bisognerà rispettare" – chiosa Bonomo<sup>1</sup>. A quali accorgimenti dunque si potrebbe/dovrebbe ricorrere, in un caso come quello che ho citato, per superare l'attrito fra deontologia e autenticità della fonte?

---

<sup>1</sup> Bonomo, *Voci della memoria*, cit., p. 110.

NORAH KARROUCHE<sup>1</sup>, LILIANA MELGAR<sup>2</sup>, ROELAND ORDELMAN<sup>3</sup>

<sup>1</sup>Vrije Universiteit Amsterdam/Erasmus Universiteit Rotterdam, <sup>2</sup>Universiteit van Amsterdam/Netherlands Institute for Sound and Vision, <sup>3</sup>Universiteit Twente/Netherlands Institute for Sound and Vision

## **Never the twain shall meet? Collaborations between speech technologists and oral historians in the Netherlands**

In the past decade, speech technologists and oral historians in the Netherlands have collaborated on a number of projects, seeking to increase the usability of automatic speech recognition (ASR) for historians working with interview data. Of particular interest is CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities)<sup>1</sup>, the latest example of such a collaboration. In this paper, we detail the nature of collaboration in CLARIAH and identify the ways in which oral historians can profit from closer collaboration with speech recognition technologists.

CLARIAH aims to increase access to cultural heritage and research data and develop innovative digital tools for humanities and social science scholars in the Netherlands. In order to accomplish these goals, universities and scholarly research centres in the Netherlands collaborate extensively with museums, archives and libraries<sup>2</sup>. One of the sub-projects in CLARIAH focuses on the accessibility of audio(visual) data, via the so-called MediaSuite<sup>3</sup>. Research environments like the CLARIAH MediaSuite allow scholars to conduct cross-medial analyses on a plethora of subjects. The CLARIAH MediaSuite is primarily designed for media scholars and oral historians, and aims at opening up archival collections owned by several institutions in the Netherlands that would otherwise remain difficult to access, with a primary interest in audio(visual) archives, including some 70 oral history collections. The MediaSuite also provides scholars with tools, enabling them to search through vast amounts of audio, construct corpora and conduct source criticism and analysis in the same digital environment. Speech recognition is but one of these tools or services, as the MediaSuite also allows scholars to, for instance, perform explorative and comparative searches, bookmark their search results, add manual annotations and save these in a personal workspace.

---

<sup>1</sup> <https://www.clariah.nl>

<sup>2</sup> See CLARIAH, <https://www.clariah.nl/over/geschiedenis/voorstel#samenvatting> , consulted October 7 2018.

<sup>3</sup> See <http://mediasuite.clariah.nl/> , and Ordelman et al. (2018).

In this paper, we seek to analyze the results of this most recent collaboration between speech technologists and oral historians, with an emphasis on ASR-output in the MediaSuite. We will do so by briefly demonstrating how the MediaSuite is currently used by oral historians. From a speech recognition perspective, ASR caters to the two scholarly primitives (Unsworth, 2000) of *discovery* and *annotation* (Ordelman, van Hessen, 2018). The ASR-output available in the MediaSuite is viewed as an annotation that can additionally facilitate the discovery of oral history data. This has not been fully recognized by oral historians, who view transcription as a means of data preparation, not annotation. Hence, as also recently stated by Ordelman, van Hessen (2018), the usability of ASR is contested by oral historians. In this paper we therefore identify and explain marked differences of opinion between speech technologists and oral historians on what exactly constitutes quality in a transcript and more importantly, to what end transcripts should be used in scholarly research (compare Scagliola, de Jong, 2014). Issues addressed in the paper are: oral historians' continued preference of manually created (or corrected) transcripts over plain, possibly erroneous ASR-output, and verbatim transcripts over indexed ones.

We henceforth argue that continued collaboration with speech technologists can significantly alter the oral historian's digital workflow, specifically with regard to the oral historian's consideration of AV-material as primary source and ASR-output as an annotation and as a means by which they can discover primary sources. User documentation on the quality of ASR-output can accordingly encourage oral historians to reconsider the use of ASR in the research process. On the other hand, oral historians' recognition of ASR-output as a means to data discovery may change ruling opinions about the usability of ASR tools. This is particularly relevant in view of recent trends in historical research methodology and source criticism, where AV-material is considered more authentic than a verbatim transcript of that material.

ORDELMAN, R., VAN HESSEN, A. (2018). Speech Recognition and Scholarly Research: Usability and Sustainability. In *CLARIN Annual Conference 2018*. Pisa.

ORDELMAN, R. ET AL. (2018). Challenges in enabling mixed media scholarly research with multi-media data in a sustainable infrastructure. In *Digital Humanities 2018 (DH2018)*. Mexico City, Mexico.

SCAGLIOLA, S., DE JONG, F. (2014). Clio's Talkative Daughter Goes Digital. The Interplay between Technology and Oral Accounts as Historical Data. In R. BOD, J. MAAT & T. WESTSTEIJN (eds.), *The Making of the Humanities, Volume III: The Modern Humanities*. Amsterdam, 511-526.

UNSWORTH, J. (2000). Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? In *Humanities Computing: formal methods, experimental practice*, May 13, 2000. London, UK.

GIOVANNA MAROTTA<sup>1</sup>, MARTINA TURCONI<sup>1</sup>

<sup>1</sup>Università di Pisa

## **Ritmo linguistico e ritmo musicale. Un'analisi pilota sul ritmo della lingua cantata**

Dagli studi pioneristici di A.D. Patel e J.R. Daniele (2003) sul rapporto tra ritmo linguistico e ritmo musicale è nato un filone di ricerca che intende applicare all'analisi delle composizioni musicali i metodi di misurazione quantitativa elaborate in ambito fonetico, quindi riferite alla lingua parlata. La domanda di ricerca può essere riassunta come segue: “è possibile che il ritmo della lingua madre dei musicisti influenzi il ritmo musicale delle loro composizioni?” Sulla scia dell'intuizione di Patel e Daniele (2003) sono quindi stati analizzati brani strumentali della musica classica europea. I risultati empirici sono sorprendenti: i valori computati per il ritmo musicale rispecchiano quelli ottenuti in ambito linguistico sulla base della tradizionale classificazione ritmica delle lingue naturali in *stresstimed* e *syllable-timed*.

Negli ultimi anni, la ricerca in questo ambito ha ampliato il repertorio musicale, dando spazio a nuove aree linguistiche (Jekiel, 2014) ed a nuovi generi musicali, come il blues (Lorroque, 2012) e la musica popolare (Sadakata *et al.*, 2004), nonché la musica vocale. Inoltre, sono stati condotti anche esperimenti percettivi per confermare la classificazione ritmica emersa dalle misurazioni acustiche (Hannon, 2009).

Nella nostra comunicazione vorremmo presentare un'analisi pilota svolta sul ritmo nella lingua cantata in italiano ed inglese. I risultati che abbiamo raccolto finora, relativi a canzoni di musica leggera degli anni Sessanta e Settanta, sono promettenti, poiché la lingua cantata sembra rispecchiare la categorizzazione ritmica della lingua parlata di riferimento del cantante. Riteniamo infatti che le tendenze prosodiche, in questa sede segnatamente ritmiche, rilevabili nel parlato siano solidali con quelle rilevabili anche nel canto e nella poesia, dal momento che rispecchiano principi generali attivi su vari livelli nel comportamento motorio umano (cfr. Marotta, 2012).

Sul piano empirico, abbiamo analizzato sia l'indice di variazione (*Pairwise Variability Index*) della linea melodica dal punto di vista musicale (riprendendo cioè il metodo usato da Patel & Daniele, 2003) che l'indice di variazione della lingua cantata (cfr. Grabe & Low 2002). Abbiamo selezionato pertanto un *corpus* di canzoni di musica leggera inglesi e italiane unitamente alle loro corrispettive

traduzioni in italiano e inglese<sup>1</sup>. Dopo aver computato per ciascuna canzone l'indice di variazione del ritmo musicale della linea melodica, abbiamo calcolato la media del *PVI* normalizzato sia per le canzoni inglesi (51,7) che per le canzoni italiane (57,5). I valori ottenuti hanno confermato che sia a livello linguistico che a livello musicale operano le stesse categorie ritmiche. In altri termini, l'analisi mostra che il ritmo della lingua madre dei compositori è solidale con il ritmo musicale delle composizioni stesse.

Per quanto riguarda l'analisi della lingua cantata, l'applicazione del metodo *PVI* (Grabe & Low, 2002) non ha dato risultati soddisfacenti, dal momento che il valore *nPVI* intervocalico è risultato maggiore per le canzoni italiane (66,7) che per quelle inglesi (61,7), con la conseguenza di collocare le prime nella classe delle lingue *stress-timed* e le seconde in quella delle lingue *syllable-timed*. Abbiamo pertanto pensato di ricorrere al metodo *varco* (Dellwo & Wagner, 2003), che consente di normalizzare la velocità di eloquio nel parlato, e quindi anche il tempo nel cantato. A seguito della normalizzazione della velocità d'esecuzione, i risultati ottenuti concordano con le nostre aspettative: sia i valori *varco $\Delta V$*  (74) che quelli *varco $\Delta C$*  (58,4) inglesi sono maggiori di quelli italiani (*varco $\Delta V$*

58,3; *varco $\Delta C$*  49,9), confermando l'appartenenza della lingua cantata inglese alla classe *stress-timed* e della lingua cantata italiana alla classe *syllable-timed*.

La condivisione di tendenze prosodiche e vincoli strutturali tanto nella lingua parlata quanto nella lingua cantata non soltanto conferma la prossimità di lingua e canto nelle culture umane, ma pongono anche la questione teorica relativa alla presenza di alcuni principi ritmici generali che informano il comportamento motorio degli esseri umani a livelli multipli.

## Bibliografia

DELLWO, V., WAGNER, P. (2003). Relations between language rhythm and speech rate, in SOLÉ, M.J., RECASENS, D., ROMERO, J. (2003). Proceedings of 15th International Congress of Phonetic Sciences, pp. 471-474.

GRABE, E., LOW, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis, in Papers in Laboratory Phonology, 7, Cambridge University Press, Cambridge, pp. 515-546.

HANNON, E.E. (2009). Perceiving speech rhythm in music: Listener classify instrumental songs according to language of origin, in Cognition, 111, pp. 403-409.

---

<sup>1</sup> Si tratta delle canzoni seguenti: a. originali italiane Azzurro di Adriano Celentano (1968); Mi ritorni in mente di Lucio Battisti (1969); Balla Linda di Lucio Battisti (1968); b. originali inglesi: Stand by me di Ben E. King (1961); I'm a believer di Niel Diamond (1966); Elenore dei The Turtles (1968).

- 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, pp. 55-71.
- LARROQUE, P. (2012). The Rhythm of English and Blues Music, in *American International Journal of Contemporary Research*, 2, n. 5, pp. 123-136.
- MAROTTA, G. (2012). Piedi metrici e sillabe orfane nella prosodia dell'italiano, in E. SCHAFROTH e M. SELIG (eds.), *Testo e ritmi. Zum Rhythmus in der italienischen Sprache*. *Studia Romanica et linguistica* 35. Frankfurt a.M., Lang, pp. 89-103.
- PATEL, A.D., DANIELE, J.R. (2003a). An empirical comparison of rhythm in language and music, in *Cognition*, 87, pp. B35-B45.
- SADAKATA, M., DESIAN, P., HONING, H., PATEL, A.D. e 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, pp. 41-44