UNIVERSITÀ DEL SALENTO
Dipartimento di Studi Umanistici

UNIVERSITÀ DEGLI STUDI DI ENNA "KORE"

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Booklet of abstracts

Progetti di Rilevante Interesse Nazionale - PRIN 2017JNKCYZ

IL PARLATO IN AMBITO MEDICO: Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD: Linguistic analysis, technological applications and clinical tools

Lecce, 15–17 February 2023
University of Salento, Lecce

AISV
Associazione Italiana Scienze della Voce

ASL Lecce
PugliaSalute

Autorizzazione ASL Lecce n. 1195/2023

IL PARLATO IN AMBITO MEDICO:
Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD:
Linguistic analysis, technological applications and clinical tools

WEDNESDAY 15 FEBRUARY
KEYNOTE SPEECHS AND ORAL PRESENTATIONS

Language impairment in neurodegenerative diseases

Isabella Laura Simone
Università degli Studi di Bari

Language and speech are one of the most complex human activities, involving cognitive-linguistic processes, motor speech planning, programming, control, phonetic integrit, and finally neuromuscular execution. “Language” refers to a core system that enables the user to assign or decode a symbol to an object or concept that one wishes to convey or comprehend, and also to apply appropriate grammatical rules to arrange or decipher phrases and sentences. In this definition, language is the “engine” of communication. When language is impaired as a result of injury to the brain, the disorder is termed “aphasia. On the other hand, “speech” is the physical process of speaking involving lungs, trachea, breathing muscles, vocal cords, mouth, tongue, facial muscles, and it is defined as the mechanism by which language is orally expressed and articulated. Some individuals may have a motor speech deficit, making difficult to produce words, e.g., “dysarthria” or to coordinate complex movements for articulation e.g., “apraxia of speech”. Two specific brain areas are related to understand and generate speech: Broca's area related to speech production and articulation, and Wernicke's area responsible for speech comprehension. However, the organization of language and the comprehension and production of speech are mediated by a much broader expanse of neural networks that covers a large number of cortical and subcortical regions and their interconnecting fiber pathway, e.g., inferior and middle frontal gyri, superior precentral gyrus of the insula, basal ganglia, posterior middle temporal gyrus, cerebellum, as demonstrated by new neuroimaging techniques.

Some disorders of nervous system may manifest with language/ speech changes.

In the last decade, language in neurodegenerative diseases has been recognized as a specific marker not only for distinguishing different language isolated syndromes, or primary progressive aphasia (PPA) and its variants, but also for diagnosing various neurodegenerative disorders characterized by language impairment amongst other cognitive deficit, and neurological signs and symptoms, as Parkinson, Alzheimer, Amyotrophic Lateral Sclerosis, Multiple Sclerosis. The study of language processing with a better knowledge of the language network, due to imaging techniques, may contribute to a greater understanding of the extension of neurodegenerative process that characterize such diseases.

Dysarthria in Parkinson's disease: towards a cross-language perspective to understand its pathophysiology

Serge Pinto
Laboratoire Parole et Langage, University of Aix-Marseille

Motor Speech Disorders refer to a set of signs affecting the control and production of speech consequent to neurological impairment. They are characterized by an approach which dichotomizes motor speech disorders in two modalities: apraxia of speech and dysarthria, which can be distinguished on at least two fundamental points: (1) dysarthria is the consequence of motor dysfunctions also involving the limbs (rigidity, akinesia, ataxia, dystonia, etc.) and of which a specific pathophysiology is determined; (2) dysarthric disorders are constant, predictable, whereas this is less the case for patients suffering from apraxia of speech. After presenting the classification and pathophysiology of dysarthrias associated with specific movement disorders, I will briefly introduce the relevance of targeting research on dysarthria, and mainly hypokinetic dysarthria in Parkinson's disease, as a model for a better understanding of the involvement of cortico-basal ganglia-cortical pathways in speech motor control. Some elements in favour of including a cross-language approach will also be discussed.

Pathological spoken Italian in adults and elderly

Francesca M. Dovetto
Università degli Studi di Napoli Federico II

In ancient cultures, the conception of 'voice' also includes its *margins*, i.e., vocal gestures and every form of buzz or noise or fragment of voice. The Western tradition, on the other hand, has focused on certain aspects only; in particular, on the meaningful word, a normative and fundamental instrument, ending up denying for a long time its multiplicity and variety. Yet in numerous diseases, the modulation and condition of the voice can constitute an important diagnostic index. Starting from a historiographical approach to the topic of voice, in this contribution some trends of the most recent linguistic reflection will be mentioned, with particular attention to interdisciplinary openings on the topic of language pathologies with reference to adults and older people.

Motor speech disorders and dysphagia in neurodegenerative diseases: clinical features compared

Sciancalepore Diletta, Lavermicocca Valentina, Illuzzi Teresanna, Lumaca Laura, Dipace Simone, Fiorella Maria Luisa
Università degli Studi di Bari

Neurodegenerative diseases are often characterised by motor speech disorders and dysphagia even though the relationship between these symptompsis not fully understood. The epidemiological studies describing motor speech disorders and its co-occurrence with dysphagia are still few and conflicting. Anyway, many studies assert that dysarthria should be considered as a major clinical sign of the presence of oropharyngeal dysphagia (Ko et al., 2018; Malandraki et al., 2011; Wang et al., 2020).

The aim of this study is to detect the presence of alterations in the acoustic parameters of the voice which can be correlated to swallowing difficulties and outline their peculiarities in various pathologies. It has been illustrated the swallowing and phonoarticulatory performance profiles in patients affected by Parkinson's Disease, Atypical Parkinsonism and Multiple Sclerosis, comparing the data obtained with those coming from a control group. For this purpose, the patients underwent clinical and instrumental evaluation of swallowing and acoustic analysis of the voice. The healty controls underwent only the phonoarticulatory evaluation.

The patients have been enrolled at Ear, Nose and Throat Unit of University Hospital in Bari: 14 patients were affected by Parkinson's Disease (PD), 5 patients were affected by Atypical Parkinsonism (AP) and 8 patients were affected by Multiple Sclerosis (MS). The control group included 15 healty people. Phonoarticulatory evaluation has been preceded by the collection of medical history information on voice and speech disorders, followed by the recording of speech samples. The articulatory abilities have been assessed through triangular voice space area, calculated using the formant frequencies of susteined vowels (tVSAsv) (Vizza et al., 2019) and the formant frequencies of vowels produced in connected speech (tVSAsp). Swallowing disorders have been investigated through clinical and instrumental evaluation (FEES).

Preliminary results indicate a statistically significant difference between the tVSAsv and the tVSAsp in the patients group ($p=0,04$). Analysis of covariance has been used to study the tVSAsp: a statistically significant difference between PD and Control group ($p=0,0002$), AP and control group ($p= 0,0082$) and MS and control group ($p=0,001$) has been found. The tVSA reduced values indicate a reduced intelligibility of speech, due to a reduction in the range of the articulatory movements.

Regarding the swallowing assessment, parkinsonian syndromes showed worse results than Multiple Sclerosis in terms of multiple swallows and bolus stagnation. From the dysphagia self-assessment questionnaires, a poor awareness of the swallowing disorder emerged in the group of Parkinsonian patients, with scores that do not reflect the actual swallowing ability. Conversely, people with Multiple Sclerosis have reported worse swallowing difficulties than the clinical objectivity.

In conclusion, we can state that in the patients group, speech and swallowing disorders are not similar either in terms of severity or onset. When both disorders coexist, dysarthria is more severe than dysphagia. Furthermore, the patients with dysphagia included in the study also have speech impairment, thus dysphagia did not occur isoleted in any case. This finding leads us to assume that motor speech disorders precedes dysphagia and corroborates the hypotesis that dysphagia should be investigated in neurodegenerative patients affected by motor speech disorders.

BIBLIOGRAFIA

Ko EJ, Chae M, Cho SR. Relationship Between Swallowing Function and Maximum Phonation Time in Patients With Parkinsonism. Ann Rehabil Med. 2018 Jun 27;42(3):425-432. doi: 10.5535/arm.2018.42.3.425. PMID: 29961740; PMCID: PMC6058589.

Malandraki GA, Hind JA,Gangnon R, Logemann JA, Robbins J. The utility of pitch elevation in the evaluation of oropharyngeal Dysphagia: preliminary findings. Am J Speech Lang Pathol. 2011 Nov;20(4):262-8. doi: 10.1044/1058-0360(2011/10-0097). Epub 2011 Aug 3. PMID: 21813823.

Vizza P, Tradigo G, Mirarchi D, Bossio RB, Lombardo N, Arabia G, Quattrone A, Veltri P. Methodologies of speech analysis for neurodegenerative diseases evaluation. Int J Med Inform. 2019 Feb;122:45-54. doi: 10.1016/j.ijmedinf.2018.11.008. Epub 2018 Nov 26. PMID: 30623783.

Wang BJ, Carter FL, Altman KW. Relationship between dysarthria and oral-oropharyngeal dysphagia: The current evidence. EarNoseThroat J. 2018 Mar;97(3):E1-E9. PMID: 29554404.

Acoustic biomarkers of Parkinson's Disease: a pilot study on the Italian Parkinson's voice and speech dataset

Alessandra Ferrera & Gloria Gagliardi
*Alma Mater Studiorum – Università di Bologna

Background

Da un punto di vista sintomatologico, la malattia di Parkinson è definita come un disturbo ipocinetico del movimento, caratterizzato da segni cardine come rigidità, bradicinesia e tremori a riposo. Oltre al generale rallentamento delle funzioni motorie, la letteratura ne ha ampiamente descritto gli effetti anche sulla voce e sul linguaggio, inquadrando le caratteristiche del disturbo nella definizione di una disartria di tipo ipocinetico. In particolare, è stato osservato che: nei pazienti la qualità della voce è spesso percepita come soffiata, aspirata, talvolta rauca; una possibile spiegazione è rintracciabile in anomalie nel meccanismo di fonazione e alterazioni al livello delle strutture laringee (cfr. Hanson et al., 1984); la ridotta mobilità degli articolatori mobili si ripercuote sulla realizzazione dei gesti articolatori, determinando un parlato tendenzialmente ipoarticolato: per le vocali, tale tendenza è stata descritta in termini acustici da metriche come l'area dello spazio vocalico (VSA) o l'indice di articolazione vocalica (VAI), significativamente più bassi nei pazienti parkinsoniani rispetto ai soggetti di controllo; anche nel caso delle consonanti, analisi acustiche e percettive hanno rilevato alcuni pattern ricorrenti – principalmente fenomeni di spirantizzazione – sempre volti alla riduzione della forza articolatoria (cfr. Logemann et al., 1978); riguardo agli aspetti prosodici, il parlato parkinsoniano è caratterizzato da una riduzione di variabilità della frequenza e dell'intensità; più controversi sono i risultati riguardanti le misure di durata (velocità di eloquio, di fonazione, distribuzione delle pause); lo studio di aspetti ritmici, infine, dimostra una tendenza all'accentuazione dei caratteri di isosillabismo, a prescindere dalla tipologia ritmica delle lingue analizzate.

In Italia, la descrizione del parlato ipocinetico è stata oggetto di studio di due gruppi di ricerca: il primo, legato all'Università degli Studi di Napoli "L'Orientale", si è concentrato sull'osservazione delle caratteristiche ritmiche dell'eloquio, ed in particolare sulla variazione di metriche quali il V-to-V e la percentuale di tempo vocalico (cfr. Maffia et al., 2021); il secondo, attivo invece presso l'Università del Salento, ha analizzato - oltre ad alcune caratteristiche prosodiche - aspetti specifici come la labializzazione (Gili Fivela et al., 2020) o la produzione di bilabiali scempie (Gili Fivela et al., 2015), anche a partire dall'osservazione congiunta di dati acustici e gesti articolatori.

Razionale dello studio

Partendo da queste premesse, la ricerca si inserisce nel quadro degli studi italiani sul tema cercando di individuare biomarker fonetico-acustici specifici del disturbo. A tal fine, è stato analizzato il dataset Italian Parkinson Voice and Speech (cfr. Dimauro et al., 2017), un corpus composto da registrazioni di 65 parlanti, provenienti prevalentemente dall'area linguistica di Bari e divisi in tre gruppi: 28 pazienti (19 M e 9 F, età $\mu = 67,2$), tutti sotto trattamento farmacologico e in stato on; 22 controlli anziani (10 M e 12 F, età $\mu = 67,1$) e 15 controlli giovani (13 M e 2 F, età $\mu = 20,8$). Fra le registrazioni disponibili, sono stati selezionati i seguenti task: serie di vocalizzazioni prolungate di /a/, /e/, /i/, /o/, /u/, prima e seconda lettura di un testo foneticamente bilanciato, lettura di frasi (10, di cui 4 interrogative). L'analisi è stata suddivisa in tre fasi: uso della DLBs Pipeline, strumento progettato specificamente per lo studio del parlato patologico e in grado di misurare un'ampia gamma di indici, fra cui le misure di durata delle pause silenti e dei segmenti di parlato, la velocità di elocuzione e di fonazione, il pitch, il centroide spettrale e la dimensione frattale di Higuchi (cfr. Gagliardi et al., 2022); studio delle formanti e delle metriche vocaliche tVSA, VAI e FCR, calcolate a partire dalle vocalizzazioni e con l'ausilio del software Praat; analisi dei dati relativi alle misure jitter, shimmer e Harmonic-to-Noise r. (HNR), calcolate da Toye & Kompalli (2021).

La significatività statistica delle osservazioni è stata valutata applicando in R test statistici inferenziali non parametrici (test di Kruskall Wallis, Wilcoxon-Mann-Whitney, p-value > 0.05). Oltre alla distinzione fra pazienti e soggetti di controllo, sono state prese in considerazione altre variabili ricavabili dal corpus,

quali la distinzione dei parlanti per sesso biologico e, internamente al gruppo dei pazienti, il punteggio ottenuto nella valutazione motoria MDS-UPDRS (cfr. MDS Task Force on Rating Scales for Parkinson's Disease, 2003).

Risultati

L'indagine rileva in primo luogo una netta distinzione fra i controlli giovani e l'insieme dei soggetti geriatrici (pazienti e partecipanti di controllo più anziani), dimostrando come, prima ancora del manifestarsi della patologia, diverse caratteristiche del parlato dei pazienti siano da attribuire a un processo più generale di invecchiamento della voce: è il caso delle alterazioni nella F0 (che tende ad abbassarsi nelle donne e ad alzarsi negli uomini con l'avanzare dell'età) o della riduzione globale della velocità di elocuzione.

Approfondendo il confronto fra il gruppo dei pazienti e quelli dei soggetti di controllo di età avanzata, i risultati della nostra ricerca si pongono per buona parte in continuità con quanto affermato in letteratura. In particolare, nel gruppo dei pazienti viene confermata la tendenza a ridurre la variabilità del pitch, e la variazione degli indici di articolazione vocalica segue la variazione prevista: tVSA e VAI aumentano rispetto ai controlli, FCR diminuisce. Meritano considerazioni a parte le diverse misure di durata, calcolate con l'ausilio della DLBs Pipeline, per le quali si rende particolarmente utile la suddivisione dei partecipanti in base al sesso biologico. Confrontati con i rispettivi controlli, pazienti di genere opposto sembrano infatti manifestare tendenze diverse per alcune delle metriche considerate: nelle donne si nota una riduzione delle pause e un aumento delle misure relative ai segmenti di parlato; negli uomini, al contrario, aumentano le pause e si riducono i segmenti parlati, a fronte di una tendenza a ridurre il tempo di fonazione. L'interpretazione di questi dati suggerisce dunque un parlato patologico femminile più disteso, con meno pause, e un parlato degli uomini più affannoso, il quale sembra procedere per blocchi rapidi intervallati da pause più lunghe.

Il confronto fra gruppi di pazienti in relazione al punteggio MDS-UPDRS sembra inoltre suggerire una correlazione positiva fra questa variabile e l'alterazione degli indici misurati dalla pipeline. Fra le metriche per cui il risultato è significativo, sottolineiamo in particolare la regolarità temporale dei segmenti sonori, che cresce all'aumentare del grado di compromissione motoria, rendendo significativa la differenza fra i pazienti ai primi e agli ultimi gradi UPDRS: un dato che potremmo interpretare come conforme ai risultati ottenuti in letteratura circa le misure ritmiche.

Infine, il presente studio ha messo in luce una serie di indici volti a descrivere la qualità del segnale, come la dimensione frattale di Higuchi, significativamente più bassa nel gruppo dei pazienti, o il centroide spettrale - correlato acustico della brillantezza della voce - per il quale si registra una maggiore variabilità rispetto ai controlli. Differenze altamente significative sono state rilevate, nel nostro campione, anche per le misure di jitter, shimmer e HNR; sorprendono, rispetto alle attese, le direzioni della variazione: i primi due risultano infatti maggiori nei soggetti di controllo, mentre il calcolo del rapporto HNR restituisce un valore superiore nei parkinsoniani; un risultato interessante, che stimola l'interesse ad approfondire in futuro lo studio di queste misure su campioni ancora più ampi e con un'attenzione sempre maggiore all'uniformità delle procedure di raccolta dei dati.

Bibliografia

Dimauro G. *et al.* (2017). Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System. *IEEE Access, 5*, 22199-22208.

Gagliardi G., Tamburini F. (2022). The Automatic Extraction of Linguistic Biomarkers as a Viable Solution for the Early Diagnosis of Mental Disorders. *Proc. of the 13th Language Resources and Evaluation Conference (LREC 2022)*, 5234–5242.

Gili Fivela B. *et al.* (2020). Controllo motorio e disartria nella malattia di Parkinson: uno studio pilota sulla labializzazione, *Studi AISV, 7*, 429-440.

Gili Fivela B. *et al.* (2015). "Consonanti scempie e geminate nel morbo di Parkinson: la produzione di bilabiali" in Vayra M. *et al.* (eds.) Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio, 289-312.

Hanson D.G. *et al.* (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. *The Laryngoscope, 94*(3), 348– 353.

Logemann J. A. *et al.* (1978). Frequency and co-occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *The Journal of speech and hearing disorders, 43*(1), 47–57.

Maffia M. *et al.* (2021). Speech Rhythm Variation in Early-Stage Parkinson's Disease: A Study on Different Speaking Tasks. *Frontiers in psychology*, *12*, 668291.

Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease (2003). The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. *Movement disorders: official journal of the Movement Disorder Society, 18*(7), 738–750.

Toye A.A., Kompalli S. (2021). Comparative Study of Speech Analysis Methods to Predict Parkinson's Disease. *ArXiv*.

Linguistic correlates of cognitive decline: collection and analysis of a spoken corpus of patients with dementia living in Basilicata

Elena Martinelli [1], Vita Garrammone [2], Francesca Mori [2], Incoronata Nolè [2], Franca Cameriero [2], Matilde Martino [2], Gaetano Di Bello [2] & Gloria Gagliardi [1]

[1]Alma Mater Studiorum – Università di Bologna, [2] Universo Salute – Opera Don Uva (Potenza)

La demenza è una sindrome clinica che si osserva in risposta all'insorgenza di varie patologie e che si caratterizza per un progressivo e multiplo deterioramento delle funzioni cognitive, a cui inevitabilmente consegue l'alterazione e la successiva perdita della personale autonomia funzionale (APA, 2013).

A causa delle allarmanti stime epidemiologiche sulla sua incidenza (e mortalità) nella popolazione anziana, nonché per il gravoso dispendio di risorse sanitarie, socioeconomiche e assistenziali che comporta (Gauthier et al., 2021), è stata definita una "priorità mondiale di salute pubblica" dall'Organizzazione Mondiale della Sanità (WHO, 2012).

In linea con tale urgenza, negli ultimi anni la comunità scientifica ha dedicato notevole attenzione all'identificazione di "marcatori" in grado di supportare la diagnosi, come biomarker liquorali, demografici e cardiovascolari. Più di recente, ha destato crescente interesse l'individuazione di indici linguistici della patologia: analizzare la produzione verbale di una coorte di pazienti e individuare le disabilità linguistico-comunicative associabili alla condizione morbosa sottesa rappresentano procedure estremamente vantaggiose (perché meno invasive e più economiche) per la definizione di strumenti in supporto alle pratiche diagnostiche, di prevenzione e riabilitazione clinica, nonché per una migliore comprensione del funzionamento della competenza linguistica normofasica (Gagliardi, 2019).

Costruire campioni di linguaggio rappresentativi di una determinata popolazione clinica costituisce un essenziale prerequisito per la conduzione di tali studi: grazie ad una notevole mole di lavori condotti fin dagli anni '70 del Novecento (Boschi et al., 2017), molte lingue europee dispongono di adeguati corpora di eloquio dementigeno (es. TalkBank – DementiaBank, per la lingua inglese; Backer et al., 2004), ma le risorse per la lingua italiana sono estremamente limitate.

Lo studio si propone di confrontare (a livello fonetico-acustico, morfosintattico, semantico, sintattico e pragmatico) l'eloquio di soggetti italofoni affetti da demenza e quello di soggetti coetanei cognitivamente integri, al fine di evidenziare divergenze quantitative/qualitative funzionali alla profilazione delle caratteristiche linguistico-comunicative del deterioramento cognitivo per la lingua italiana.

Per raggiungere tale obiettivo sono stati raccolti campioni di eloquio semi-spontaneo patologico e normotipico: disporre di verbalizzazioni autentiche è infatti imprescindibile per l'identificazione dei biomarker language-specific del declino cognitivo, poiché la costruzione del profilo linguistico-comunicativo associabile ad un disturbo non può aprioristicamente basarsi sulla generalizzazione dei risultati ottenuti indagando lingue diverse.

Il reclutamento di individui nati e residenti in una regione non ancora coinvolta in studi di tale natura consente di disporre di materiale prezioso per futuri confronti sociolinguistici in merito alla potenziale incidenza di tratti linguistici regionali sui profili linguistici di pazienti appartenenti ad aree differenti della Penisola. In virtù della notevole variazione diatopica che caratterizza la lingua italiana (Berruto, 2021), infatti, è necessario raccogliere corpora di linguaggio parlato prodotto da soggetti residenti in zone diverse dell'Italia: solo in questo modo si può garantire una più affidabile generalizzabilità ai risultati ottenuti.

Il protocollo di ricerca ha previsto una preliminare somministrazione di prove psicometriche volte alla valutazione dello stato cognitivo dei partecipanti: MMSE (Magni et al., 1996), MoCA (Conti et al., 2015) e Test di Fluenza verbale, sia fonemica (Carlesimo et al., 1995; 1996) che semantica (Spinnler & Tognoni, 1987).

Sono stati reclutati 40 individui con medesima origine e residenza regionale (la Basilicata), di cui 24 di sesso femminile e 16 di sesso maschile, tutti con età superiore ai 60 anni. Il campione è bilanciato per sesso ed età (81± 6,6) ed è così suddiviso: 1) gruppo dei pazienti: 20 soggetti (12 F, 8 M), tutti con diagnosi conclamata di demenza e in cura presso la Residenza Sanitaria Assistenziale Opera Salute – Don Uva (sede

di Potenza, PZ). Nello specifico, sono stati reclutati 9 soggetti affetti da malattia di Alzheimer, 2 pazienti affetti da demenza mista, 5 pazienti affetti da demenza non ulteriormente specificata, 3 pazienti affetti da demenza vascolare, 1 paziente affetto da demenza frontotemporale; 2) gruppo di controllo: 20 soggetti (12 F, 8 M) cognitivamente integri (MMSE $\geq$ 22, MoCA > 19.26, F. fonemica $\geq$ 17.35, F. semantica $\geq$ 7.25).

L'eloquio dei partecipanti è stato elicitato mediante la somministrazione di tre task linguistici: due compiti narrativi (1. racconto di una gita / un viaggio; 2. racconto delle tradizioni legate alla festività del Natale); un compito descrittivo, per il quale è stato impiegato come stimolo visivo una figura inclusa nel test neuropsicologico "Esame del Linguaggio II" (Ciurli et al., 1996).

Il corpus di parlato semispontaneo raccolto consta in totale di circa 9 ore di sonoro, di cui 8 ore acquisite per i gruppi reclutati (ca. 4 ore per ogni gruppo) e ca. 45 min attribuibili agli interventi dell'intervistatrice.

Manualmente, mediante il software ELAN (2021), sono state realizzate la trascrizione ortografica, la segmentazione in enunciati (unità di riferimento dell'analisi linguistica; Austin, 1962), l'annotazione del parsing prosodico e di alcuni fenomeni paralinguistici (framework teorico: L-AcT; Cresti et al., 2018). Tramite la pipeline progettata da Gagliardi & Tamburini (2022), sono state automaticamente eseguite le procedure di tokenizzazione, lemmatizzazione, PoS-tagging e parsing a dipendenze, funzionali all'estrazione dei Digital Linguistic Biomarkers (DLBs), una serie di indici linguistici (in totale: 151) che consentono di indagare in ottica quantitativa il livello fonetico-acustico, lessicale, semantico, sintattico e di leggibilità dei testi orali.

La significatività statistica dei tratti linguistici indagati è stata valutata applicando test di inferenza statistica parametrici e non parametrici (es. T-test di Student, Test di Wilcoxon-Mann-Whitney, Test di Kruskal-Wallis) in R (R Core Team, 2014).

Il progetto di ricerca è stato approvato dal Comitato di Bioetica dell'Alma Mater Studiorum - Università di Bologna (n. 0072032/2022). Tutti i soggetti reclutati (oppure i loro rappresentati legali) hanno espresso il loro consenso informato alla partecipazione allo studio e al trattamento dei propri dati personali.

L'analisi complessiva dei DLBs estratti ha consentito di evidenziare 59/151 caratteristiche linguistiche significativamente differenti (p - value < 0.05) tra eloquio normotipico e patologico.

Di grande utilità a fini diagnostici/di screening è l'evidenziazione di una forte compromissione delle caratteristiche spettro- acustiche della voce (9/20 tratti significativamente diversi): l'analisi automatica dell'eloquio permette di individuare marker di malattia di ardua identificazione mediante le classiche procedure cliniche di valutazione dei casi di demenza. In particolare, i pazienti producono enunciati significativamente più brevi (la cui durata dipende crucialmente dal tipo di disturbo diagnosticato), costituiti da un maggior numero di disfluenze e un minore numero di parole, aspetti che concorrono a rendere le loro verbalizzazioni meno fluenti. I segmenti di speech prodotti si caratterizzano per alterazioni della dimensione frattale di Higuchi, del Centroide Spettrale e della Root Mean Square Energy (López-de-Ipiña et al., 2013).

I dati quantitativi estratti, tuttavia, riflettono la compromissione di tutti i livelli linguistici indagati. Dal punto di vista morfosintattico (15/34 tratti significativamente diversi), la registrazione della ridotta occorrenza di aggettivi, avverbi, preposizioni semplici e articolate, determinanti e articoli, al netto di una maggiore produzione di interiezioni, concorre a spiegare la povertà di densità contenutistica dell'eloquio dei pazienti. È osservabile, inoltre, una minore realizzazione di specifiche sottocategorie lessicali (es. verbi d'azione, pronomi relativi, ausiliari). Sotto il profilo semantico (29/84 tratti significativamente diversi), gli enunciati prodotti dai pazienti evidenziano aspetti psicopatologici classicamente associati alle sindromi demenziali (es. generale appiattimento emotivo, ridotta attività cognitiva). Viene confermato il complessivo mantenimento delle competenze sintattiche (4/9 tratti significativamente diversi), nonostante le strutture sintagmatiche prodotte risultino generalmente semplificate, si possa osservare un più alto margine di variabilità circa la profondità massima dell'embedding e si registri una riduzione della lunghezza media degli enunciati. La maggiore leggibilità di base e lessicale (2/4 tratti significativamente diversi) osservata per l'eloquio dei pazienti supporta quanto evidenziato analizzando i singoli livelli linguistici.

La considerazione dei tratti linguistici risultati significativamente differenti soltanto in alcuni task

permette di ampliare la caratterizzazione delle disabilità linguistico-comunicative associabili alla demenza (es. minore complessità del segnale acustico, ridotta occorrenza di congiunzioni subordinanti, maggiore espressione di inibizione). È possibile formulare anche importanti considerazioni circa l'adeguatezza dei task linguistici somministrabili in tali ricerche: la narrazione si configura come compito più funzionale all'evidenziazione di indici linguistici di deterioramento cognitivo.

Nel complesso, i DLBs caratterizzanti l'eloquio patologico sono interpretabili come i correlati linguistici "più forti" del declino cognitivo, poiché sono emersi nonostante le eterogenee forme di demenza diagnosticate ai pazienti. D'altro canto, la stessa composizione mista del gruppo può aver oscurato le peculiarità linguistiche di ogni quadro dementigeno: anche se i risultati sono interpretabili solo con cautela, le analisi mettono in luce per ognuno di essi una determinata costellazione di marker linguistici.

Infine, l'indagine qualitativa del corpus concorre a dettagliare il profilo delle disabilità linguistico-comunicative associabili alla demenza, permettendo di evidenziare deficit pragmatico-discorsivi (es. tangenzialità, deragliamento ideativo, coesione referenziale deficitaria, ripetizione/iterazione argomentativa), nonché ulteriori peculiarità, tra cui notevole presenza di parafasie e false partenze, attuazione di strategie in compensazione della severa anomia e contestualizzazione spazio- temporale degli eventi deficitaria.

Riferimenti Bibliografici

Alzheimer's Disease International, Gauthier, S., Rosa-Neto, P., Morais, J.A., & Webster, C. (2021). World Alzheimer Report 2021: Journey through the diagnosis of dementia. London, England: Alzheimer's Disease International.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Austin, J.L., (1962). How to do things with words. Clarendon Press, Oxford.

Becker, J. T., et al. (1994), The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. Arch Neurol, 51(6): 585-594.

Beltrami, D. et al., (2016). Automatic identification of Mild Cognitive Impairment through the analysis of Italian spontaneous speech productions. In Calzolari, N. et al. (ed), Proceedings of the Tenth International Conference on Language Resources and Evaluation – LREC 2016, ELRA – European Language Resources Association, 2086-2093.

Berruto, G. (2021), Sociolinguistica dell'italiano contemporaneo (Nuova edizione). Carocci Editore, Roma.

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., Cappa, S.F. (2017). Connected speech in neurodegenerative language disorders: a review. Front Psychol 8, 256.

Carlesimo, G. et al. (1995). Batteria per la valutazione del deterioramento mentale: standardizzazione ed affidabilità diagnostica nell'identificazione di pazienti affetti da sindrome demenziale. Archivio di Psicologia Neurologia e Psichiatria. 56: 471–488.

Ciurli, P., Marangolo, P. & Basso, A. (1996), Esame del Linguaggio II. Manuale e materiale d'esame. Giunti, Firenze.

Conti, S., Bonazzi, S., Laiacona, M., Masina, M., Vanelli Coralli, M., (2015). Montreal Cognitive Assessment (MoCA) - italian version: regression based norms and and equivalent scores. Neurol. Sci. 26: 209–214.

Cresti, E. & Moneglia, M. (2018). Chapter 13. The illocutionary basis of Information Structure: The Language into Act Theory (L-AcT), in Adamou, E. et al. (eds.), Information Structure in Lesser-described Languages: Studies in prosody and syntax. Amstredam: John Benjamins Publishing Company, pp. 360–402.

ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Scaricabile da https://archive.mpi.nl/tla/elan. Gagliardi, G. (2019), Linguistica per le professioni sanitarie, Pàtron, Bologna.

Gagliardi, G., Tamburini, F. (2022), The Automatic Extraction of Linguistic Biomarkers as a Viable Solution for the Early Diagnosis of Mental Disorders. In Proc. of the 13thLanguage Resources and Evaluation Conference (LREC 2022), Marseille, France, 21-23 June 2022, 5234–5242.

Lopez-de-Ipina, K. et al. (2013). On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis. Sensors, 13:6730– 6745. Magni, E., Binetti, G., Bianchetti, A., Rozzini, R., & Trabucchi, M. (1996). Mini-Mental State Examination: a normative study in Italian elderly population. European journal of neurology, 3(3), 198–202. https://doi.org/10.1111/j.1468-1331.1996.tb00423.x

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical.

Spinnler, H. & Tognoni, G. (1987). Standardizzazione e taratura italiana di test neuropsicologici: gruppo italiano per lo studio neuropsicologico dell'invecchiamento. Milano: Masson Italia periodici (Supplementum 8 - Italian journal of neurological sciences).

A computational analysis of speech patterns in dementia

Francesco Sigona[1], Barbara Gili Fivela[1], Pietro Vigorelli[2], Andrea Bolioli[3]
[1]Università del Salento – [2]Associazione Gruppo Anchise - [3] Maze.

La malattia di Alzheimer (Alzheimer's Disease, AD) è la forma più diffusa di disturbo neurocognitivo degenerativo ed è caratterizzata da un insieme di sintomi quali: difficoltà di memoria, di linguaggio, di risoluzione di problemi e compromissione di altre abilità cognitive che influenzano la capacità di una persona di eseguire le normali attività quotidiane [1]. Sebbene il deficit di memoria sia il primo e più caratteristico sintomo della malattia, la ricerca scientifica guarda con sempre maggiore interesse ai disturbi del linguaggio.

Sottili cambiamenti nel parlato possono essere osservati anni prima della comparsa dei sintomi caratteristici della malattia di Alzheimer [2, 3] e vengono rilevati anche nelle prime fasi del lieve deterioramento cognitivo [4]. È stato dimostrato, infatti, che sia il lieve deterioramento cognitivo che la malattia di Alzheimer influiscono sulla fluidità verbale, in particolare sulle esitazioni del paziente nel parlare e sulla velocità d'eloquio nel suo complesso; peraltro, la tipica difficoltà nel trovare le parole adeguate induce il paziente a ricorrere a circonlocuzioni o all'uso frequente di pause piene (ad es. ehm). Sono, inoltre, frequenti gli errori di tipo semantico, l'uso di termini indefiniti, la rianalisi della produzione, le ripetizioni, i neologismi, la semplificazione lessicale e grammaticale [5]. Il parlato nei malati di Alzheimer è caratterizzato da una ridotta coerenza, con dettagli non plausibili e irrilevanti [6]. Queste caratteristiche hanno il potenziale per diventare marcatori per la diagnosi precoce e il monitoraggio degli stadi di evoluzione della demenza e dell'Alzheimer [7]. D'altra parte, le moderne tecniche di trattamento automatico del linguaggio (TAL, ovvero Natural Language Processing, NLP) consentono oggi di elaborare ingenti quantità di materiali linguistici, sotto forma di registrazioni sonore e, più spesso, in forma scritta o trascritta, benché si rilevi una generale penuria di dataset che permettano indagini approfondite (spesso non contengono sia informazioni cliniche che materiali linguistici adatti per indagini statistiche, o basate su apprendimento automatico, in quanto includono materiali raccolti eterogenei, tipicamente monologhi e dialoghi [8]).

Il Corpus Anchise [9] è in continua crescita e rappresenta un significativo esempio di corpus in lingua italiana, in cui sono raccolte le trascrizioni di dialoghi tra operatori sanitari e pazienti affetti da demenza di Alzheimer, accompagnate in buona parte dalla rilevazione del Mini-Mental State Examination (MMSE, [10]), uno dei test neuropsicologici più utilizzati per la valutazione dello stato neurocognitivo e funzionale del paziente. Il Corpus è attualmente composto da 600 conversazioni raccolte in condizioni ecologiche nell'arco di 15 anni (17/1/2007-7/6/2022) da operatori sanitari con diversa formazione professionale (es. operatori sociosanitari, educatori, psicologi) e con diversi gradi di addestramento all'Approccio Capacitante [11]. L'Approccio Capacitante assegna all'operatore il compito di aprire "la conversazione in modo tale da favorire la continuazione del parlare e il benessere del paziente, possibilmente senza fare domande e restituendo il senso percepito delle sue parole (motivo narrativo)", concludendo "la conversazione quando il paziente dà segni di stanchezza o di disagio".

In questo contributo, si presenta un'analisi delle 600 conversazioni attualmente disponibili, incrementando quindi i materiali discussi in [9]: si tratta di 238 sedute-colloquio (7596 turni, oltre 116.000 parole), ottenute escludendo le conversazioni con più di due conversanti e quelle in lingua straniera (415 conversazioni denominate "Corpus Anchise 2022", corrispondente a 26900 "turni"), e selezionando i turni di parola dei pazienti per i quali è disponibile lo score MMSE. Sul corpus è stata svolta un'analisi di tipo linguistico-computazionale, basata su tecniche di NLP ed una successiva analisi di tipo statistico-descrittivo e inferenziale, allo scopo di verificare se le caratteristiche linguistiche dei turni dei malati di Alzheimer possano essere correlate con il grado di avanzamento della malattia espresso dallo score MMSE. In linea con i lavori individuati in letteratura, che, nella quasi totalità dei casi, prendono in considerazione task di produzione del parlato differenti da quello dialogico (come ad esempio, task di descrizione di immagini, di lettura e richiamo, di narrazione spontanea ecc.) l'ipotesi alla base dello studio è che la semplificazione lessicale e grammaticale osservata con l'avanzare della malattia [5] possa corrispondere ad una correlazione positiva tra l'MMSE e parametri linguistici come il numero di parole o di parti del discorso come gli aggettivi.

Data la trascrizione dei materiali, fatta individuando i turni di parola, l'analisi linguistico-computazionale, attraverso il tool "Stanza" [12], ha permesso di individuare parametri quali il numero di

turni di parola e di tokens (parole), indici di varietà lessicale (TTR, Types-Token Ratio), di frequenza di occorrenza delle parti del discorso (POS, Part-of-speech), di frequenza di forme, modi e tempi verbali. L'analisi statistico-descrittiva ha poi permesso di correlare la variazione dei parametri suddetti con i punteggi di MMSE suddivisi in tre gruppi: severo (MMSE <6), moderato (MMSE tra 6 e 10), lieve (MMSE tra 10 e 24) [13]. I risultati preliminari sembrano evidenziare differenze statisticamente significative tra il grado severo e gli altri gradi di decadimento cognitivo, relativamente al numero medio di parole per turno ed al tasso di occorrenza di alcune POS, come gli aggettivi. La quantità di parole media utilizzata dai pazienti con grado severo appare significativamente inferiore rispetto ai pazienti con grado moderato (Wilcoxon rank sum test: W = 320.50, p = 0.003; r (rank biserial) = -0.43, 95% CI [-0.63, -0.17]) e lieve (W = 1213.00, p = 0.003; r (rank biserial) = -0.38, 95% CI [-0.57, -0.15]). Cfr. Fig. 1. Il TTR dai pazienti con grado severo appare significativamente maggiore rispetto ai pazienti con grado moderato (t-test: diff. = 0.08, 95% CI [0.02, 0.15], t(50.39) = 2.83, p = 0.007; Cohen's d = 0.70, 95% CI [0.19, 1.20]) e lieve (W = 2584.00, p = 0.009; r (rank biserial) = 0.33, 95% CI [0.09, 0.53]). Cfr. Fig. 2. Analogamente per il tasso degli aggettivi (upos: ADJ) tra grado severo e moderato (W = 729.00, p = 0.029; r (rank biserial) = 0.32, 95% CI [0.05, 0.55]) e tra grado severo e lieve ((W = 2473.50, p = 0.011; r (rank biserial) = 0.32, 95% CI [0.08, 0.52])).

Saranno discusse anche analisi più approfondite, attualmente in corso, che consentiranno di caratterizzare con maggiore dettaglio ulteriori aspetti del parlato dialogico dei pazienti del corpus.

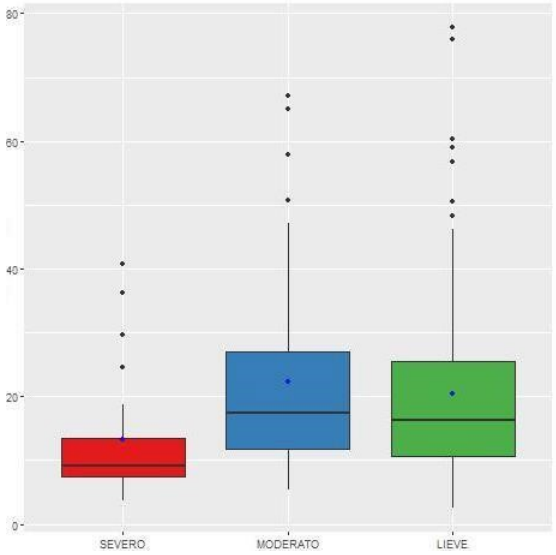


Figura 1: Numero di tokens per turno

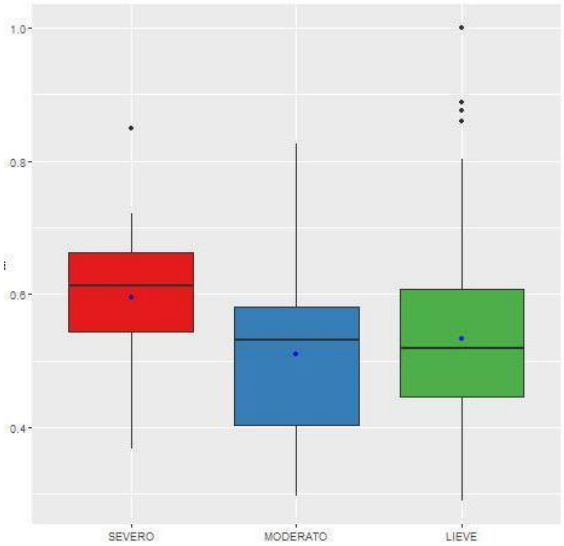


Figura 2: Types-to-Token Ratio (TTR)

Riferimenti bibliografici

[1] Alzheimer's-Association. "2018 Alzheimer's disease facts and figures." *Alzheimer's and Dementia*, 2018: 367-429

[2] Cuetos, F., Arango-Lasprilla, J. C., Uribe, C., Valencia, C., and Lopera, F. (2007). Linguistic changes in verbal expression: a preclinical marker of Alzheimer's disease. J. Int. Neuropsychol. Soc. 13, 433–439. doi: 10.1017/S1355617707070609

[3] Ahmed S, Haigh AM, de Jager CA, Garrard P. Discorso connesso come indicatore della progressione della malattia nel morbo di Alzheimer testato dall'autopsia. Cervello. 2013 dicembre;136 (Pt 12):3727-37.

[4] Toth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z, et al. Una soluzione basata sul riconoscimento vocale per il rilevamento automatico di lievi disturbi cognitivi dal discorso spontaneo. Curr Alzheimer Ris. 2018;15(2):130-8.

[5] Reilly J, Peelle JE, Antonucci SM, Grossman M. Anomia come marker di distinti disturbi della memoria semantica nella malattia di Alzheimer e nella demenza semantica. Neuropsicologia. 25 luglio 2011 (4): 413–26.

[6] Boschi V, Catricalà E, Consonni M, Chesi C, Moro A, Cappa SF. Discorso connesso nei disturbi neurodegenerativi del linguaggio: una revisione. Psicologo anteriore. 2017 Mar;8:269.

[7] König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Analisi vocale automatica per la valutazione dei pazienti con predemenza e malattia di Alzheimer. demenza di Alzheimer. 2015 marzo; 1 (1): 112-24.

[8] de la Fuente Garcia, S., Haider, F., & Luz, S. (2020). Cross-corpus Feature Learning between Spontaneous Monologue and Dialogue for Automatic Classification of Alzheimer's Dementia Speech. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 5851-5855). IEEE.

[9] Nicola Benvenuti, Andrea Bolioli, Alessio Bosca, Alessandro Mazzei, Pietro Vigorelli, The "Corpus Anchise 320" and the analysis of conversations between healthcare workers and people with dementia, Proc. of Seventh Italian Conference on Computational Linguistics, Bologna, 1-3 March 2021. http://hdl.handle.net/2318/1761583

[10] Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, *12*(3), 189-198

[11] Vigorelli P., Bonalume M., Cocco A., Lacchini C., Maramonti A., Negri Chinaglia C., Peduzzi A., Pezzano D., Riedo E., Sertorio S. L'Approccio Capacitante nella cura degli anziani fragili e delle persone con deficit cognitivi. 10 anni di esperienza. Psicogeriatria 2011; 2: 58-70.

[12] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In Association for Computational Linguistics (ACL) System Demonstrations. 2020.

[13] Harvey R.G. Alzheimer's disease: clinical diagnosis and management strategies. Clinician, Vol.15, n.1, G-C C. Ltd

IL PARLATO IN AMBITO MEDICO:
Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD:
Linguistic analysis, technological applications and clinical tools

WEDNESDAY 15 FEBRUARY
POSTER SESSION I

Markers of dialogue shifts in doctor-patient interactions

Sarah Bigi[1], Vittorio Ganfi[2], Fabrizio Macagno[3], Maria Grazia Rossi[3]
[1]Università Cattolica del Sacro Cuore, [2]Università degli Studi Roma 3, [3]Universidade NOVA de Lisboa

Background: The quality of dialogical interactions between doctors and patients has attracted growing attention since the description of the patient-centered paradigm of care (Moja, Vegni 2000). In particular, the process of shared decision making (SDM) has been identified as a crucial precondition for full patient empowerment and involvement in the process of care (Charles, Gafni, Whelan 1997, 1999). However, also in this case, it has been repeatedly reported that in real-life practice SDM is often poorly realized or simply absent. A recent debate is questioning the definition of SDM and some researchers have suggested alternative conceptualizations of the process, such as "contextualizing decisions" (Gerwing, Gulbrandsen 2019).

The most problematic aspect seems to be the fact that SDM is seldom considered from a linguistic-pragmatic perspective. In other words, very few studies focus on the linguistic structures that are called into play in the dialogical realization of SDM. This has been done in the fields of Conversation Analysis (CA), argumentation theory (Labrie, Schulz 2014), and discourse analysis (Sarangi 2010). In this paper, we share the approach of 'dialogism', which stresses the "dynamic use of linguistic resources in the construal of meaning in interaction and contexts" (Linell, 2009: 97). This approach assumes that grammatical constructions have meaning potentials tied to them, and that these potentials become actualized in combination with specific contextual factors (Linell 1998, 2009; Rocci 2003).

Aim: Within this theoretical starting point, our aim is to contribute to the description of SDM in clinical interactions by resorting to conceptual and methodological resources drawn from research on argumentation practices and linguistics. Our goal is to describe the dialogical realization of deliberations within a corpus of medical interactions, by analyzing the dialogue moves performed by participants. In particular, we are interested in understanding how, interactionally, a participant shifts to a different type of dialogue by letting the interlocutor understand that a shift is taking place and, on the other side, the other party responds by acknowledging the shift and 'playing along'. We are particularly interested in understanding if there are recurrent linguistic structures that signal shifts to deliberation moves and, if there are, which are they.

Method: For our analysis, we rely on a corpus of 53 transcriptions of doctor-patient interactions (264,185 tokens) that have been video-recorded in a diabetes outpatient clinic in Northern Italy during the years 2012-2014 (Bigi 2014). The recordings have been performed after obtaining signed consensus from all the participants involved and approval from the clinic's Ethical Committee. The recordings have been transcribed following the Jeffersonian transcription conventions. For the analysis, our main conceptual tool is the 'dialogue move', which is the minimal unit in a coding procedure described in Macagno & Bigi (2017, 2018) and Rossi, Macagno & Bigi (2022). Within this analytical framework, we defined seven coding categories for the analysis of dialogue moves in a diabetes care context: 1 – information sharing personal, 2 – information sharing procedural, 3 – information sharing clinical, 4 – clinical proposal, 5 – procedural proposal, and 6 – persuasion.

Results: Preliminary results from the coding performed on a subsection of the corpus show that shifts from moves 3 (information sharing clinical) to 4 (clinical proposal) happen very frequently through connectives expressing consequence (*allora, quindi, per cui*). The expression of the clinical proposal happens also very frequently through the use of hedges and bushes (Caffi 1999) that downgrade the directive illocution and increase the degree of intimacy between participants, also realizing politeness strategies. Given the description in the literature of a tight connection between intensity, politeness, prosody and context (Gili Fivela & Bazzanella 2014), we are particularly keen to obtain from the research

community present at the conference insights on how to develop the analysis of prosody in relation to the dialogic, syntactic and contextual factors already described through our analysis.

References

Bigi S. (2014), Healthy Reasoning: The Role of Effective Argumentation for Enhancing Elderly Patients' Self-management Abilities in Chronic Care, in G. Riva - P. Ajmone Marsan - C. Grassi (eds.), *Active Ageing and Healthy Living A Human Centered Approach in Research and Innovation as Source of Quality of Life*, Amsterdam, IOS Press, pp. 193–203.

Bigi, S. (2022). *Le strutture interrogative nelle interazioni in contesto clinico*. Milano, Vita & Pensiero. Caffi, C. (1999), *On mitigation*, Journal of Pragmatics, 31, 7, pp. 881-909.

Charles C. - Gafni A. - Whelan T. (1997), *Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango)*, Social Science & Medicine, 44, 5, pp. 681-692.

Charles C. - Gafni A. - Whelan T. (1999), *Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model*, Social Science and Medicine, 49, pp. 651-661.

Gerwing J. - Gulbrandsen P. (2019), *Contextualizing decisions: Stepping out of the SDM track*, Patient Education and Counseling, 102, pp. 815-816.

Gili Fivela, B. – Bazzanella, C. (2014), *The relevance of prosody and context to the interplay between intensity and politeness. An exploratory study on Italian*, Journal of Politeness Research, 10, 1, pp. 97-126.

Labrie N. - Schulz P.J. (2014), *Does argumentation matter? A systematic literature review on the role of argumentation in doctor-patient communication*, Health Communication, 29, 10, pp. 996-1008.

Linell P. (1998), *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*, Amsterdam/Philadelphia, John Benjamins Publishing.

Linell, P. (2009). Grammatical constructions in dialogue, in Bergs, A. – G. Diewald, G. (eds.), *Contexts and constructions* (Vol. 9), Amsterdam/Philadelphia, John Benjamins Publishing, pp. 97-110.

Macagno, F. - Bigi, S. (2017). *Analyzing the pragmatic structure of dialogues*, Discourse Studies, 19, 2, pp. 148–168.

Macagno, F. - Bigi, S. (2018). Types of dialogue and pragmatic ambiguity. In S. Oswald – J. Jacquin – T. Herman (eds.), *Argumentation and language*, Cham, Switzerland, Springer, pp. 191–218.

Macagno, F. - Bigi, S. (2020). *Analyzing dialogue moves in chronic care communication. Dialogical intentions and customization of recommendations for the assessment of medical deliberation*, Journal of Argumentation in Context, 9, 2, pp. 167–198.

Rocci, A. (2003), La testualità, in G. Bettetini - S. Cigada - S. Raynaud (eds.), *Semiotica II*, Brescia, La Scuola, Rossi, M.G. - Macagno, F. - Bigi, S. (2022). *Dialogical functions of metaphors in medical interactions*, Text and Talk, 42, 1, pp. 77–103.

Sarangi S. (2010), Practising discourse analysis in healthcare settings, in I. Bourgeault - R. DeVries - R. Dingwall (eds.), *The SAGE Handbook of Qualitative Methods in Health Research*, London, Sage, pp. 397-416.

Vowel Formant Variability in the Speech of Patients with Dementia: a Qualitative and Quantitative analysis

Gloria Gagliardi[1], Chiara Meluzzi[2]
[1]*Università di Bologna,* [2]*Università degli Studi di Milano*

Adult neurodegenerative disorders have been described in relation to communication problems involving word retrieval, but also motor control and other speech disturbances that may affect language production or comprehension to various degrees. As for dementia, many studies have demonstrated a reduction of durational pattern in comparison with healthy speech, from a phonetic (e.g., vowel duration, word duration) and a phonological perspective (e.g., syllable duration, but also phonological selection of more frequent structures, see Carozza et al. 2011, Behforuzi et al. 2013 among many others). However, only a few studies have been devoted to the analysis of vowel formant variability in people with mild-stage dementia as opposed to a control group of the same (old) age. Nishikawa et al. (2022) represent a significant exception since they tried to develop a protocol for the early identification of dementia through different acoustic measures, including formant analysis of the five Japanese vowels /a/, /e/, /i/, /o/, /u/. Their results demonstrated a general reduction of the vowel space in speakers with dementia, thus possibly resulting in lower speech intelligibility. Furthermore, they have shown a significant variation in F1 more than in F2, thus indicating a general lowering of the tongue in people with dementia compared to the control group.

By directly referring to Nishikawa et al. (2022) pioneering study, we investigate formant variability in the spontaneous speech of 2 Italian-speaking patients (1M, 73 years-old; 1F, 63 years-old) diagnosed with dementia coming from the corpus collected by Martinelli (2022; Bioethics Committee of the University of Bologna n. 0072032/2022). The enrolled patients lived in Basilicata. Their productions were compared with 2 healthy control speakers balanced by sex and age. After the neuropsychological evaluation (i.e., MMSE, MoCA, semantic and phonemic fluency), a sample of semi-spontaneous speech was collected by administering three elicitation tasks: two autobiographical narratives (i.e., journey and Christmas tradition) and the description of a complex figure ("Esame del Linguaggio II", Ciurli et al., 1996).

We isolated the words and annotated the stressed vowel in three different PRAAT layers (*sentence*, *word*, *phone*, cf. Fig. 1). The beginning and ending of vowels have been identified by the starting and ending of F2. Vowel transcription used X-SAMPA and the following tags: "_ST" for stressed vowels, "_UN" for unstressed vowels, "_NAS" for presence nasalization, "_ROT" for presence of rhoticization. Nasalization and rhoticity have been tagged whenever their presence occurred for about a quarter of the overall vowel length.

Through a PRAAT script, we collected formants' values at the middle point and on 7 target points through the vowel duration, to qualitatively analyze the dynamic variability of vowels in different consonantal contours. We plotted vowel dynamics and vowel space through the online tool Visible Vowel (Heeringa & Van de Velde 2018). A statistical ANOVA analysis was run on the different formants with the midpoint values to compare speakers with dementia and controls by dividing the analysis according to the target vowel.

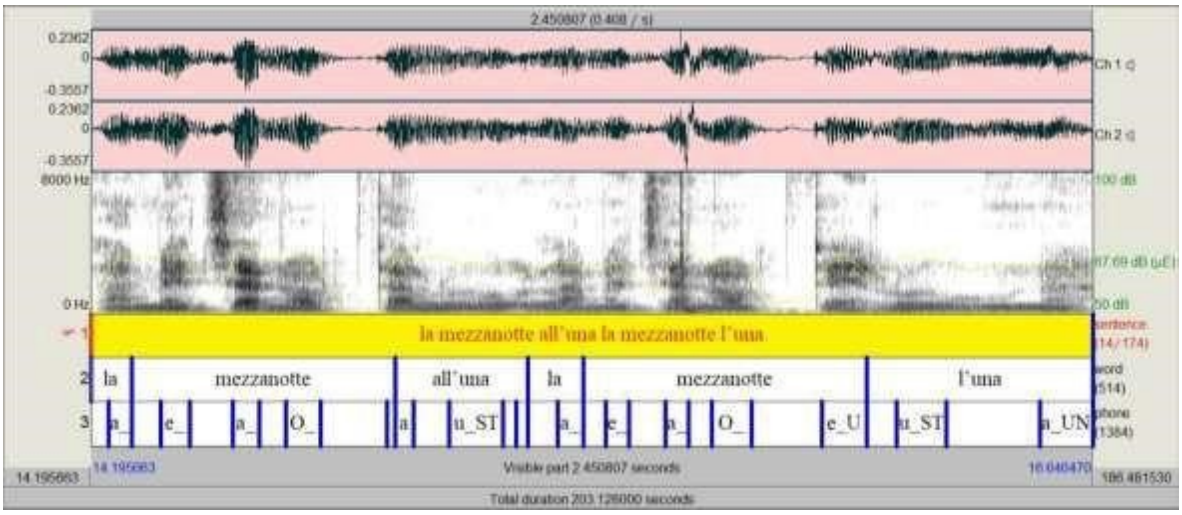


Fig. 1 Tagging example of the vowels in our corpus

So far, we hypothesize that, in line with Nishikawa *et al*. (2022), F1 will be significantly lower in speakers with dementia compared to the control group. In line with other works that have examined the lexical-phonetic interface in pathological speech (e.g., Behforuzi et al. 2013), we also assume a possible role played by word and syllable frequency in determining different behavioral patterns in the productions of speakers with dementia.

References

Behforuzi, H., Brandon Burtis, D., Williamson, J.B., Stamps, J.J. & Heilman, K.M. (2013) Impaired initial vowel versus consonant letter-word fluency in dementia of the Alzheimer type, *Cognitive Neuroscience* 4(4): 163-170.

Carozza, L., Quinn, M., Nack, J. & Bell-Berti, F. (2011) Temporal structure in the speech of a person with dementia: a longitudinal case study, *Acta Neuropsychologica* 9(4): 383-392.

Ciurli, P., Marangolo, P., and Basso, A. (1996) *Esame del Linguaggio II*. Firenze: Giunti OS. Martinelli, E. (2022) *I correlati linguistici del deterioramento cognitivo: raccolta e analisi di un corpus di eloquio patologico di individui anziani italofoni affetti da demenza*. Master Thesis, University of Bologna.

Heeringa, W. & Van de Velde, H. (2018) Visible Vowels: a Tool for the Visualization of Vowel Variation, *Proceedings CLARIN Annual Conference 2018*, 8 - 10 October, Pisa, Italy. CLARIN ERIC. Nishikawa, K., Hirakawa, R., Kawano, H. & Nakatoh, Y. (2022) Analysis of prosodic features and formant of dementia speech for Machine Learning, *Proceedings of 5th International Conference on Information and Computer Technologies (ICICT)*, pp. 173-176.

Vogel, A.P., Poole, M.L., Pemberton, H., Caverlé, M.W.J., Boonstra, F., Low, E., Darby, D. & Brodtmann, A. (2017) Motor speech signature of behavioral variant frontotemporal dementia. Refining the phenotype, *American Academy of Neurology* 89: 837-844.

Exploring Alternative Subwords Approaches For E2e ASR

Sara Picciau, Domenico De Cristofaro, Marco Matassoni, Roberto Gretter, Alessio Brutti
Free University of Bozen, Fondazione Bruno Kessler

There are different ways to design the discrete units used in end-to-end Automatic Speech Recognition (E2E-ASR) to represent the basic tokens that model sequences of words. Subwords represent an convenient solution: the proper combination of groups of characters of different length can efficiently cover the vast majority of actual words without the need of a large vocabulary. Recent works show that subwords are a convenient model for ASR, comparing favorably with respect to characters or words [1]. These units are usually obtained through algorithms that rely on frequency, probability and likelihood, the most popular being WordPiece, Byte Pair Encoding and Sentencepiece, and they perform very well when the amount of data is relevant. However, these tokenization approaches can lead to suboptimal results when little training data is available or in a multilingual setting; in such cases, injecting additional (phonological) information in the tokenizer can be an interesting strategy to address this issue. The aim of this work is to investigate two alternative approaches to build effective subwords sets to be implemented in a E2E ASR system by exploiting acoustic and phonological information. In both cases, the textual data is represented in phonemes rather than graphemes.

The first approach starts from a set of subwords obtained through the WordPiece [2] on the training data and aims to obtain a more refined and optimal subwords set by exploiting the acoustic confusability among the units. he metric of Phonetic Confusability (PC) to define the ideal vocabulary size of subwords is based on the phonetic similarity between phonetic subword strings. To calculate the similarity score we adapt the Levenshtein algorithm. Usually the phoneme substitution operation, in classical Levenshtein algorithm, results in an unambiguous value that does not consider the articulatory and acoustic nature of the substituted phoneme. In this approach we consider as acoustic quantification of the phonemes, the acoustic inner representation of the model. For weights assignment, we used the corpus APASCI [3] which is meticulously phonetically transcribed and contains timestamps for each segment present in an audio transcription. After passing the APASCI audio data to an Italian ASR WavLM model trained with a phoneme tokenizer, we extract the last hidden layer vectors and force align them to their labels considering the original timestamps of the corpus as ground truth. Once collected all the vectors representing each phoneme of the APASCI corpus, we calculate the centroids and use them as weights for the substitution operation of the PED. The distance matrix for each subwords vocabulary is calculated in order to decide how to prune it. Identifying subwords vocabulary from their phonetic complexity aims to select the best suitable units for the tokenizer, avoiding confusability among tokens that share a similar or distant acoustic structure.

In the second approach the set of subwords contains linguistically significant elements, namely phonological syllables. These units consist of at least a nucleus, a segment characterized by high sonority (typically avowel) that can be surrounded by consonantal segments that constitute the onset and the coda of the syllable and have lower sonority. The distribution of the elements inside the syllabic units is defined by two main rules: firstly, the Sonority Sequencing Principle, according to which syllable sonority increases before the nucleus until and then decreases; secondly, the Maximum Onset Principle, stating that consonantal elements are preferred to be in the onset position rather than in the coda [4, 5]. Language-specific rules added to these two universal principles make the syllable inventory of a language easily predictable. Moreover, since syllables can be formed by combining the segments in a restricted number of ways according to the SSP and the MOP and in a limited number of patterns, being the CV the least marked one, they constitute a convenient inventory in terms of size. Recently, experiments with syllables as subwords in E2E ASR have been made on data of Sino-Tibetan languages [6, 7]. Within this works the target language will be Italian.

The experiments will consist in fine-tuning the pre-trained model WavLM-large using supervised Italian data (i.e. waveforms plus transcriptions). For both experimental set-ups, a grid of experiments has been

designed with sets of different amounts of audio data including 5, 10 and 20 hours and three subword vocabularies composed of ~100, ~250 and ~500 units. The data used for the experiments is taken from the corpus Common Voice [8].

Within the first approach we present a new subwords regularization method, which selects the most acoustically similar or distant phonetic subwords to implement in the Wav2Vec2PhonemeCTCTokenizer. The distance thresholds are set to 0.5 for the most acoustically similar and 0.7 for the distant one. The experiments are conducted for both sets of subwords, i.e., those that are most similar to each other and those that are furthest apart. In the second approach a custom rule-based tokenizer has been designed and implemented to tokenize the data according to the syllabification rules using the Wav2Vec2PhonemeCTCTokenizer as source.

In the final paper we will report the comparison between the two approaches in terms of two evaluation metrics: the Token Error Rate (TER) that calculates the percentage of subwords that were not predicted correctly, and Phoneme Error Rate (PER) that computes the amount of inconsistencies at a phoneme level.

Some preliminary results with small amounts of data are shown in the table below:

h **units**		**5h**		**10h**	
		TER	PER	TER	PER
~100	sub_sim	18.58%	8.21%	16.01%	6.89%
	sub_dist	17.11%	7.52%	16.41%	6.94%
	sub_syl	10.3%	8.7%	9.7%	8.1%
~250	sub sim	39.29%	9.76%	36.98%	8.53%
	sub dist	28.79%	9.9%	26.71%	8.24%
	sub_syl	12.0%	9.6%	10.3%	8.4%

Table 1. Comparison of TER and PER implementing different subword sets

References

[1] Papadourakis, Vasileios & Müller, Markus & Liu, Jing & Mouchtaris, Athanasios & Omologo, Maurizio. (2021). Phonetically Induced Subwords for End-to-End Speech Recognition. 1992-1996. 10.21437/Interspeech.2021-1787.

[2] Kudo, Taku. (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. 66-75. 10.18653/v1/P18-1007

[3] Picciau, S., De Cristofaro, D., Matassoni, M., Gretter, R., (2022). "Efficient Phonetic Representation for Multilingual ASR system". Submitted to Conference AISV, May 2022.

[4] B. Angelini, F. Brugnara, D. Falavigna, D. Giuliani, R. Gretter, M. Omologo, "Speaker Independent Continuous Speech Recognition Using an Acoustic-Phonetic Italian Corpus". In Proceedings of ICSLP94, Vol. III, pp. 1391-1394, Yokohama, Japan, 1994.

[5] Clements, G. N. (1990). The role of the sonority cycle in core syllabification. Papers in laboratory phonology, 1(283-333).

[6] Kahn, D. (1976). Syllable-based generalizations in English phonology. PhD diss. Massachusetts Institute of Technology.

[7] Zhou, S., Dong, L., Xu, S., & Xu, B. (2018). A comparison of modeling units in sequence -to-sequence speech recognition with the transformer on mandarin chinese. In Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25 (pp. 210-220). Springer International Publishing.

[8] Vu, N. T. (2014). Automatic speech recognition for low-resource languages and accents using multilingual and crosslingual information (Doctoral dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), Diss., 2014).

[9] Ardila, R., Branson, M., Davis, K, Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., and Weber, G., (2020). Common voice: A massively-multilingual speech corpus. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, May 2020, 4218-422.

Infrastructures for Research and Services. CLARIN for Speech Sciences and Education in Linguistics, Phonetics and Phonology

Francesca Frontini[1], Monica Monachini[1], Silvia Calamai[2]
[1] *Cnr-Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa & CLARIN IT*
[2] *DFCLAM, Università di Siena*

Questo contributo mira a presentare l'infrastruttura CLARIN ERIC per le risorse e tecnologie del linguaggio, mettendo in luce in particolare le applicazioni degli strumenti CLARIN per la didattica in linguistica, fonetica, fonologia.

La didattica universitaria negli ultimi decenni si è orientata verso percorsi di insegnamento più interattivi, in grado di rendere gli studenti e le studentesse più partecipanti attivi nel processo di apprendimento (vd. Brew 2006). Per la didattica delle lingue e nell'insegnamento della fonetica e della fonologia nei curricula dei corsi di laurea in lingue, mediazione linguistica, interpretariato, tali percorsi possono essere facilitati dalla presenza di dati linguistici messi a disposizione seguendo i principi della scienza aperta e resi FAIR (Findable, Accessible, Interoperable and Reusable ovvero trovabili, accessibili, interoperabili e riutilizzabili - per maggiori informazioni sui principi FAIR si veda Wilkerson et al 2016).

In questo contributo ci concentreremo sui dati e gli strumenti messi a disposizione da CLARIN (https://www.clarin.eu/), la più importante infrastruttura europea nel settore dei dati linguistici. CLARIN (Common Language Resources and Technology Infrastructure) è un ERIC (European Research Infrastructure Consortium) articolato in centri che ospitano e offrono dati, servizi e competenze nel settore delle risorse e tecnologie del linguaggio. Se i centri sono diffusi in diversi paesi (sia europei che extraeuropei), i servizi centrali dell'infrastruttura rendono il loro reperimento ed uso più facile attraverso l'applicazione di standard e protocolli comuni. Grazie a tali servizi e cataloghi, è possibile accedere a grandi collezioni testuali, corpora linguistici scritti, orali e multimodali, piattaforme di interrogazione ed analisi linguistica in diverse lingue.

Se CLARIN nasce come infrastruttura di ricerca, per facilitare l'uso e il riuso delle risorse linguistiche da parte di ricercatori nel settore delle tecnologie del linguaggio, gli strumenti di CLARIN sono già utilizzati in aula in numerose istituzioni accademiche europee e permettono di sviluppare competenze digitali oltre che disciplinari.

Per facilitare tale utilizzo, CLARIN ha lanciato di recente una serie di iniziative volte allo sviluppo e alla condivisione di materiali didattici (https://www.clarin.eu/content/teaching- clarin). Negli ultimi due anni, grazie anche all'istituzione del *Teaching with CLARIN Award*, sono stati già condivisi diversi corsi sviluppati da ricercatori attivi nell'infrastruttura e che utilizzano risorse CLARIN. Per quanto riguarda il settore del trattamento dei dati orali si citerà il corso in sei lezioni messo a disposizione da Lennes (2021) dedicato allo studio, trascrizione, annotazione ed analisi di corpora orali, e contenente moduli dedicati a strumenti quali Praat ed ELAN, oltre che una guida all'uso delle risorse della *Language Bank of Finland*.

Se tali corsi possono fornire interessanti spunti, si tratta comunque di materiali concepiti per un particolare contesto d'uso, che non sono quindi immediatamente riutilizzabili da parte di altri docenti senza un adattamento preliminare. Per questo motivo, CLARIN sta anche mettendo a punto una serie di moduli didattici sviluppati ex novo e orientati all'uso dei suoi principali servizi, che possono quindi facilmente essere integrati in diverse tipologie di corsi nelle discipline del testo e del linguaggio. Qui di seguito descriveremo alcuni di questi servizi e il loro uso nei moduli didattici.

Il **VLO (Virtual Language Observatory** https://vlo.clarin.eu/**)** è il meta catalogo di CLARIN. Esso permette il reperimento di risorse e collezioni di interesse indipendentemente dal centro in cui queste sono ospitate. Attraverso l'uso del VLO, ma anche delle **Resource Families** (https://www.clarin.eu/resource-families) - agili raccolte di risorse compilate per tipo - gli studenti possono scoprire numerose tipologie di dati, imparare ad identificare gli standard e i livelli di annotazione ed acquisire importanti nozioni di base su preservazione, metadatazione e citazione delle risorse linguistiche.

Una volta reperite le risorse di interesse, il **CLARIN Language Resources Switchboard** (https://switchboard.clarin.eu/) permette di analizzarle con catene di analisi del linguaggio multilingui e allo stato dell'arte. Ciò facilita l'introduzione di alcuni dei concetti fondamentali del trattamento automatico del linguaggio, quali i moduli e livelli di analisi del testo e del linguaggio, oltre che dei formati di input ed output, attraverso interfacce e visualizzazioni semplici e intuitive che non richiedono installazione o competenze tecniche avanzate.

La **Federated Content Search** (https://www.clarin.eu/content/content-search), e le molte piattaforme di esplorazione di corpora messe a disposizione dai vari centri sono un buon punto di partenza per introdurre nozioni di base di linguistica dei corpora, e per apprendere i primi rudimenti di linguaggi di esplorazione quali CQL, la Corpus Query Language.

Infine, la **Transcription Chain** mira a mettere a disposizione strumenti di facile utilizzo per la trascrizione e l'allineamento con la forma d'onda di dati parlati, in diverse lingue europee (si veda il sito https://speechandtech.eu/transcription-portal, che conduce direttamente al portale dedicato).

L'utilizzo degli strumenti dell'infrastruttura CLARIN in ambito didattico è attualmente in fase di pilotaggio, nel contesto della partecipazione di CLARIN al progetto ERASMUS+ UPSKILLS (https://upskillsproject.eu). I primi moduli saranno presto disponibili su una piattaforma Moodle dedicata oltre che sul portale dell'infrastruttura. Al fine di facilitare il loro utilizzo, saranno inoltre messe a punto delle linee guida dettagliate su come integrare questi ed altri strumenti nei corsi di linguistica e lingue.

L'Italia aderisce a CLARIN attraverso il consorzio CLARIN-IT (https://www.clarin-it.it/) e contribuisce all'infrastruttura con diversi centri, tra cui un centro B (di dati e servizi), ILC4CLARIN (https://ilc4clarin.ilc.cnr.it/), che ospita anche la piattaforma Archivio ViVo (https://archiviovivo.clarin-it.it/), dedicata alla conservazione e diffusione degli archivi orali. Grazie all'adesione a CLARIN, studenti italiani possono già ora utilizzare gli strumenti dell'infrastruttura, inclusi quelli ad accesso riservato, con le loro credenziali accademiche. Nei prossimi anni, anche grazie al sostegno dei finanziamenti del PNRR Infrastrutture di Ricerca, CLARIN-IT si attiverà per rendere FAIR il patrimonio di risorse linguistiche italiane, inclusi i numerosi corpora e archivi orali. Contestualmente, si avvieranno anche attività volte alla traduzione, adattamento e divulgazione dei moduli qui descritti, oltre che allo sviluppo di altri moduli dedicati per le risorse italiane, con particolare attenzione alle risorse orali.

Questo intervento mira quindi da una parte presentare alla comunità dell'AISV i principali moduli didattici di interesse, con esempi di utilizzo che possano permettere già ora di introdurre questi strumenti nei percorsi di studio delle scienze del linguaggio, lingue e filologia, e dall'altra a rafforzare il dialogo con la comunità al fine di raccoglierne le esigenze e le competenze già in essere, in vista di auspicate collaborazioni future.

Bibliografia

Brew, Angela (2006) Research and teaching: beyond the divide. London: Palgrave Macmillan.

Fung, Dilly (2017). A Connected Curriculum for Higher Education. UCL Press.

Jenkins, Alan, and Mick Heale (2009). "Developing the student as a researcher through the curriculum." Innovations in practice 2.1. 3-15.

Jong, Franciska de, Dieter Van Uytvanck, Francesca Frontini, Antal van den Bosch, Darja Fišer, and Andreas Witt. (2022). "Language Matters. The European Research Infrastructure CLARIN, Today and Tomorrow." In CLARIN. The Infrastructure for Language Resources, by Darja Fišer and Andreas Witt, 1:31–58. Digital Linguistics. Berlin, Boston: De Gruyter.

Lennes, Mietta (2021). "Introduction to Speech Analysis" (Puheen analyysin perusteet) [Online course material].

Monachini, Monica, and Francesca Frontini (2016). "CLARIN, l'infrastruttura Europea Delle Risorse Linguistiche per Le Scienze Umane e Sociali e Il Suo Network Italiano CLARIN-IT." IJCoL - Italian Journal of Computational Linguistics, Special Issue on NLP and Digital Humanities, 2 (2): 11–30.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. (2016). "The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3 (March): 160018.

Temporal profile of linguistic information in speech and language hierarchy and its neural correlates

Cosimo Iaia[1,2,3], Alessandro Tavano[3], Edoardo Pinzuti[4], Adriana Paladini[5], and Mirko Grimaldi[1,2]
[1]Università del Salento, Lecce, Italy; [2]Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL), Lecce, Italy; [3]Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany; [4]Leibniz Institute for Resilience Research, Mainz, Germany; [5]UO Neuroradiologia, Ospedale Vito Fazzi ASL/LE, Italy

Introduction

Rhythmicity is the main expression of neuronal activity. Thus, it is thought that neuronal oscillations represent a temporal mechanism for perception and action. Traditionally, psychophysiological, and neurophysiological studies classify neural oscillations into five bands: delta, 0.5-3.5 Hz, theta, 4-7.5 Hz, alpha, 8–13 Hz, beta, 14–30 Hz, and gamma, above 30 Hz (low gamma = 30-80 Hz, high gamma = 90-150 Hz, see Buzsáki 2006).

According to influential oscillatory theories of within the neurobiology of language research, the key mechanism for tracking linguistic information at multiple timescales is the so-called *neural entrainment* (Giroud and Poeppel 2012; in this study *entrainment in the broad sense* according to Obleser and Kayser 2019): that is, the property of the neural activity to temporally align to specific frequencies in the acoustics (i.e., speech envelopes) in order to track meaningful information decoding, at least, the acoustic signal into linguistic representations (Ding et al. 2016). In particular, it has been shown that the syllabic timescale is entrained to theta oscillations (4-8 Hz) (Morillon et al. 2012), specifically with a peak around 4 to 5 Hz (Poeppel and Assaneo 2020) and the phonemic timescale falls within the low gamma range (30 Hz) (Di Liberto et al. 2015). Subsequent studies have tried to apply the same mechanism of information tracking to higher and more abstract units, such as syntactic phrases: in the seminal study by Ding and colleagues (2016), syntactic phrases are entrained by delta oscillations (0.1-4 Hz), reflecting their specific timescale in a frequency tagging paradigm where stimuli are designed to have same duration per word (one syllable each word) and are presented at a constant rate. Following these results, other researchers have linked oscillations to physical timescales of linguistic units in the acoustics (Keitel et al. 2018, Coopmans et al. 2021).

The implicit assumptions for such theories to account for a temporal alignment of the human brain to the acoustic input is that linguistic unit durations, ranging from phonemes to sentences, are unimodally distributed at each level and non-overlapping among them. In fact, only unimodality of the distribution of durations (that is, not more than one clear peak in the distribution) can allow for one oscillator per speech/language level.

How does temporal variability in linguistic information characterize the brain's neural response? In the present study, we test whether the temporal structure of linguistic information at each timescale is, in fact, unimodally distributed and distinguishable allowing for a neural tracking of each timescale. Instead of the speech envelopes in the acoustics, we investigated an alternative approach for tracking linguistic units based on linguistic annotations.

Methods

To do so, we designed an EEG experiment with quasi-naturalistic stimuli: 23 participants (5 male, mean age = 23.3, std ± 3.5) were asked to listen to the first chapter of two different audiobooks in the Italian language (roughly 18 minutes) with no experimental manipulation. To check for attention and comprehension, every 50 seconds participants are asked to answer a two-choice written question. As a control condition, at the end of one chapter and before moving to the second one, reversed and non-intelligible versions of the same stimuli are presented (audio files were reversed in Audacity).

Temporal references (i.e., onset and offset) for phonemes, syllables, words, and sentence (speech units) were manually annotated in PRAAT (Boersma et al. 2022). As for syntactic phrases (language units), constituency analysis was performed with Stanza (Qi et al. 2020), a library in Python that provides a bottom-up parser for the Italian language. Syntactic Phrases were then manually annotated and checked for

mistakes: Noun Phrase (NP), Verb Phrases (VP), Prepositional Phrases (PP) and Clauses were considered for this analysis. All other syntactic units were excluded for insufficient number of occurrences in our material.

Duration of each unit was calculated as the difference of offset minus onset (performed in Python). As a first step, we used the Coefficient of Variation (CV), calculated as standard deviation divided by the mean, to compare across different timescales (from milliseconds to seconds). All speech units showed a CV below 1 with phonemes and syllables having a significant difference in variance from words and sentences. On the other hand, syntactic phrases showed a higher CV on average, with Noun Phrases above 1, yielding the highest variance. To assess their distribution, we then constructed six models to test, based on unimodality (either gaussian or lognormal) and bimodality (a combination of the previous models). Best fit to our distributions was assessed using both R-squared and Akaike Information Criterion (AIC), both common measures for model selection. As results, we found that larger variance for words and sentences was due to a bimodal distribution, whereas phonemes and syllables are unimodally distributed. Interestingly, syntactic phrases showed a unimodal distribution except for NPs. Using results from the previous model fits, we estimated the highest posterior density interval (HPD) for the corresponding frequencies to be chosen for filtering acoustic and EEG signals.

At this point, two sets of subsequent analyses were performed: replicating previous literature (e.g., Keitel et al. 2018), we extracted acoustic envelopes by filtering the acoustic signal at different frequencies based on the estimated HPD intervals. We then filtered the EEG signal at the same frequencies and used a Mutual Information approach to determine how dependent the two signals are. Conversely, we constructed new vectors of linguistic information by binarizing the time series of the acoustics, with each offset corresponding to 1 in the new series. This offset marking linguistic vectors were convolved with kernels estimated for each sound segment and band-passed filtered at the intervals assessed with HPD. Finally, we used a Mutual Information approach again between the new vectors and the EEG signal.

Results and discussion

Our findings show that with an approach based on the acoustics, as in previous literature, linguistic information at multiple timescales is not distinguishable in a naturalistic setting: while phonemes, syllables, and words overlap largely, sentences evoked a different neural response. All phrases analyzed were also undistinguishable. The main difference with previous studies (Keitel et. al 2018; Coopmans et al. 2022) and the reason why neural tracking to low frequencies in the acoustics does not work in our study is that we first model the exact temporal profile of the information without making any assumptions on the mean duration of linguistic units.

Using linguistic vectors instead of the acoustic envelopes led to distinguishable neural responses across levels. Such analysis also allowed for separating the information type between phrases.

Overall, our results show that an approach based solely on the acoustic analysis does not hold considering the exact temporal profile of linguistic units of interest. In particular, the findings of the present research suggest that the common assumption in the oscillatory framework that there exists one oscillator per time scale, as a consequence of unimodality in the distribution of their duration, does not hold in the present study. On the one hand, some linguistic units exhibit a bimodal distribution that does not allow for such tracking; on the other hand, a temporal alignment to linguistic information superimposed onto the acoustics leads to a distinguishable neural response at different levels of granularity.

These results address the question whether and how the temporal profile of linguistic information plays a role in the neural entrainment mechanism. At a more general level, our study shows that the brain does not extract linguistic information directly from the acoustics in a pure bottom-up manner and neural entrainment as a mechanism does not hold in a non-controlled experimental setting. On the other hand, top-down linguistic information is essential for language comprehension: in fact, we show that the brain temporally aligns to different types of information, from phonemes to sentences, in a different way, possibly based on internal top- down knowledge.

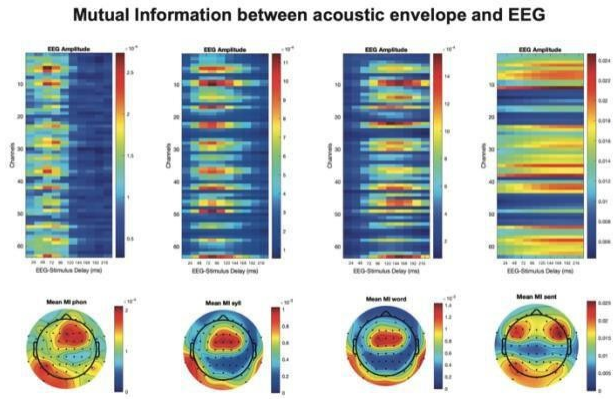


Fig.1 Mutual information between acoustic envelope and EEG. Neural response for phonemes, syllables, words, and sentences (from left to right). No distinguishable response for 3 levels

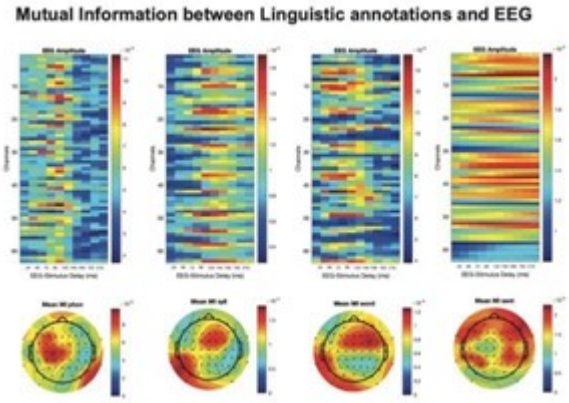


Fig.2 Mutual information between Linguistic vectors and EEG. Neural response for phonemes, syllables, words, and sentences (from left to right). All units elicit a distinguishable response

References

Audacity, T. (2014). Audacity.

Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.23, retrieved 8 October 2022 from http://www.praat.org/

Buzsáki, Gyorgy 2006. *Rhythms of the Brain*. Oxford: Oxford University Press.

Coopmans, C. W., De Hoop, H., Hagoort, P., & Martin, A. E. (2022). Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. *Neurobiology of Language*, 1-56.

Di Liberto, G. M., O'sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457-2465.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*, *19*(1), 158-164.

Giraud, AL., Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**, 511– 517 (2012). https://doi.org/10.1038/nn.3063

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS biology*, *16*(3), e2004473.

Morillon, B., Liegeois-Chauvel, C., Arnal, L., Bénar, C., & Giraud, A.-L. (2012). Asymmetric Function of Theta and Gamma Activity in Syllable Processing: An Intra-Cortical Study. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00248

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in cognitive sciences*, *23*(11), 913– 926. https://doi.org/10.1016/j.tics.2019.08.004

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature reviews neuroscience*, *21*(6), 322-334.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Arabic foreign language: acquisition of new phonological categories

Nicholas Nese
Università degli Studi di Pavia

Il presente contributo si inserisce all'interno di un progetto di dottorato volto a indagare i processi acquisizionali relativi a nuove categorie fonologiche in apprendenti italofoni di arabo LS, sia a livello di produzione sia di percezione. Due delle domande di ricerca che hanno guidato la progettazione di questo studio sono volte a verificare se 1) esiste una correlazione tra il livello degli apprendenti italofoni di arabo LS e le loro produzioni/percezioni di nuove categorie fonologiche; 2) se esistono dei suoni o dei contrasti più difficili da acquisire nella produzione e/o nella percezione dell'arabo LS. Ad oggi non esiste un modello teorico completo, in grado di rendere conto dei processi acquisizionali sia a livello di produzione sia di percezione (Colantoni et al. 2015: 31); per questo motivo si è deciso di testare tre diversi modelli teorici e di valutarne l'adeguatezza. I modelli presi in esame sono la *Markedness Differential Hypothesis* di Eckman (1977), lo *Speech Learning Model* (SLM) di Flege (1995, SLM-r 2021) e il *Perceptual Assimilation Model* (PAM) di Best (1995; Best & Tyler, PAM-L2 2007). La varietà linguistica oggetto di studio di questa ricerca è l'arabo standard (*fuṣḥa*), varietà che viene insegnata nei corsi di lingua araba all'estero. Le variabili fonetiche-fonologiche che sono state indagate riguardano un gruppo di suoni estranei all'inventario fonologico dell'italiano, ovvero i suoni faringalizzati [sˤ] [dˤ] [tˤ] [ðˤ], faringali [ʕ][ħ] e l'occlusiva uvulare [q].

La ricerca è stata condotta in piena pandemia covid-19, pertanto la raccolta dati è stata svolta a distanza con l'ausilio di Gorilla, *tool* per la realizzazione di esperimenti online. L'esperimento prevedeva un breve questionario preliminare, un task di produzione (lettura di parole in isolamento), due task percettivi (di identificazione e discriminazione AX) e un questionario finale. Per ogni task è stata stilata una lista di 42 parole costituita da monosillabi reali, con il suono target in posizione iniziale, e seguito da vocale /a i u/ quasi sempre lunga. Gli stimoli percettivi sono stati registrati da 5 madrelingua, tre donne e due uomini di 4 nazionalità diverse (Marocco, Tunisia, Egitto e Giordania), tutti con incarichi di docenza di lingua araba a livello universitario. Hanno preso parte all'esperimento 125 soggetti di cui 64 sono stati inclusi nel campione ai fini dell'analisi. I soggetti sono stati quindi suddivisi in livello di 'esperienza': base (25%), intermedio (41%) e avanzato (34%). Il livello di esperienza dei partecipanti è stato definito sulla base dei CFU conseguiti e del numero di corsi frequentati. È stato inoltre reclutato un gruppo di controllo, per il solo task di discriminazione AX, costituito da 15 studenti universitari che non hanno mai studiato la lingua araba.

L'analisi del task di produzione (lettura di parole in isolamento) è stata eseguita utilizzando come variabile dipendente il giudizio relativo al grado di accettabilità attribuito agli stimoli: ogni parola registrata dai partecipanti è stata valutata da chi scrive come accettabile (*target like*) o non accettabile (*non target like*), su base percettiva, rispetto alla realizzazione fonetica dell'arabo standard. È stata eseguita un'ulteriore valutazione di conferma sul 10% del corpus da parte di due docenti di lingua araba, di cui uno arabo madrelingua, con un *agreement* dell'85%. Complessivamente le risposte *target like* (TL) sono risultate essere il 70%. È stata successivamente integrata la variabile 'esperienza' per verificare la variazione rispetto al livello degli apprendenti. I partecipanti di livello base hanno ottenuto una media del 64% di risposte TL, il gruppo intermedio ha ottenuto una media del 67% e quello avanzato del 77%. È stata condotta un'ANOVA ($F(2, 61) = 4.987$, $p < .010$, $\eta 2 = .141$) che ha rivelato una differenza statisticamente significativa tra i diversi gruppi. Un test post hoc di Tukey ha rivelato una differenza statisticamente significativa tra il gruppo base e il gruppo avanzato ($p < .016$) e tra il gruppo intermedio e il gruppo avanzato ($p < .034$). Non è emersa significatività tra il gruppo base e quello intermedio ($p < .812$).

Per ciascun fono è stata calcolata la media degli esiti valutati come realizzazioni TL. Un'ANOVA di Welch ($F(13, 10.418) = 66{,}487$ $p < .001$) ha confermato la significatività statistica delle differenze fra i vari esiti. È sato quindi eseguito un test post hoc di Games- Howell grazie al quale è stato possibile individuare

tre gruppi di foni etichettabili come difficili [ðˤ] [ʕ] [sˤ] [tˤ] [dˤ], medio-difficili [q][ħ][h] [ð], facili [ʔ][s][t] [k] [d]. L'analisi di entrambi i task percettivi è stata eseguita utilizzando come variabile dipendente il D prime (d')[1]. Per il task di identificazione (82% di risposte TL) è stato eseguito un test di Kruskal-Wallis (H (2) = 8.386, p< .015, a due code) che ha rivelato una differenza statisticamente significativa tra i diversi livelli degli apprendenti, nello specifico tra il gruppo base e il gruppo avanzato (p<.010) nonché tra il gruppo intermedio e quello avanzato (p< .018). La differenza tra gruppo base e intermedio non è risultata statisticamente significativa (p<.608). Sono stati successivamente calcolati i valori relativi ai singoli foni: un'ANOVA di Welch (F (13, 10.595) = 7.778 p = .001) ha rivelato una differenza statisticamente significativa tra [ðˤ]-[ʔ] (p < .045) nonché una differenza, seppur di poco non significativa, tra [ðˤ]-[ʕ] (p= .057). Per l'analisi del task di discriminazione AX (65% di risposte TL) è stata eseguita un'ANOVA (F (2, 61) = .023, p < .978) che ha rivelato l'assenza di una differenza statisticamente significativa tra i tre gruppi di partecipanti. Anche l'analisi (ANOVA, F (6, 35) =1.376, p < .252) dei diversi contrasti non ha rivelato una differenza statisticamente significativa.

Sulla base dei dati raccolti è possibile riconoscere una correlazione tra il livello degli apprendenti e le loro produzioni e identificazioni. Questo risultato conferma quanto emerso nello studio di Shehata (2018), dove si evidenzia una differenza statisticamente significativa solo tra gli studenti di livello base-intermedio e gli studenti di livello avanzato. Per poter osservare un miglioramento significativo risulta quindi necessario considerare un lunga esposizione alla LS\L2, in linea con Flege & Bohn (2021) secondo i quali l'apprendimento di una L2 sarebbe un processo lento. Non risulta invece correlato il livello degli apprendenti con la capacità di discriminare i suoni. È stato possibile identificare un diverso grado di difficoltà dei suoni e dei contrasti, che si è rivelato statisticamente significativo nei task di lettura e identificazione. È inoltre emerso che la difficoltà riscontrata dai partecipanti, rispetto al singolo suono, può variare a seconda che si consideri il fenomeno sul piano della produzione o della percezione. Infine, sebbene nessuno dei tre modelli teorici adottati si sia rivelato pienamente adeguato per rispondere a tutte le domande di ricerca, il PAM-L2 è risultato il più accurato rispetto al singolo quesito di ricerca relativo alla percezione di contrasti non nativi. Riteniamo inoltre che il MDH sia più adeguato rispetto allo SLM-r in studi in cui l'interesse è rivolto a un campione di apprendenti ampio ed eterogeneo, laddove il modello di Flege & Bohn sarebbe più indicato qualora la ricerca fosse orientata a indagare l'acquisizione linguistica da parte di singoli apprendenti o di un campione ridotto.

BIBLIOGRAFIA

Best, C. (1995) *A direct realist perspective on cross-language speech perception* in W. Strange, ed. *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, 171–204.

Best, C. T., & Tyler, M. D. (2007). *Nonnative and second-language speech perception: Commonalities and complementarities* in M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production*, Amsterdam: John Benjamins, 13-34.

Colantoni, L., Steele J., & Escudero, P. (2015) *Second Language Speech. Theory and Practice*, Cambridge, Cambridge University Press.

Eckman, F. R. (1977) *Markedness and the contrastive analysis hypothesis*, Language learning, 27, 315-330.

Flege, J. E., & Bohn, O. S. (2021) *The revised speech learning model (SLM-r).* Second language speech learning: Theoretical and empirical progress, 3-83.

Macmillan, N.A. & Douglas Creelman, C. (2005) *Detection Theory: A User's Guide*, second edition, Mahwah, New Jersey, Lawrence Erlbaum Associates.

Shehata, Asmaa (2018), Native English speakers perception and production of Arabic consonants, in Alhawary, Mohammad T. (a cura di), The Routledge Handbook of Arabic Second Language Acquisition, Routledge, New York (NY), pp. 56-69.

[1] Indice di sensibilità della Detection Theory (MacMillan & Creelman 2005) che tiene conto dei Hit (veri positivi) e dei False alarm (falsi positivi): d'=z(H)-z(F).

. The coarticulation of /n/ before the glide /w/ in Italian: acoustic data and phonological implications

Piero Cossu
Università di Pisa

Il lavoro si propone di fornire dati acustici preliminari sul comportamento della nasale alveolare /n/ davanti al legamento labiovelare /w/ in italiano al confine di parola, contesto scarsamente studiato dal punto di vista fonetico.

Tra i fenomeni che interessano la nasale /n/, uno dei più noti è la coarticolazione con la consonante seguente, sia all'interno ("incastro" /inkastro/ > [iŋkastro]) che al confine di parola ("con Carlo" /kon karlo/ > [koŋ karlo]). Il processo è categorico in italiano (Bertinetto e Loporcaro, 2005: 134), ma può talvolta essere influenzato dalla velocità del parlato, in particolare al confine di parola (Celata et al., 2013). Un altro aspetto riguardante /n/ in posizione di coda al confine di parola è la sua risillabificazione in posizione di attacco quando la parola seguente inizia con vocale: "con Alberto" > [ko.nal.bɛr.to] (Bertinetto e Loporcaro, 2005: 142).

Davanti a un legamento /w j/, la nasale non acquisisce il luogo di articolazione del segmento seguente; questo indica che le approssimanti hanno un comportamento vocalico in questo contesto (Bertinetto, 2010: 12). A partire da questa osservazione, si può supporre che anche davanti ai dittonghi ascendenti sia in atto un processo di risillabificazione: "un uomo" > [u.nwɔmo], come accade nell'incontro post-lessicale tra una nasale finale e una vocale iniziale. C'è tuttavia motivo di credere che lo scenario appena descritto possa essere soggetto a variazione. Infatti, Bertinetto e Loporcaro (2005: 142) e Baroni (2020) osservano che davanti ai dittonghi /wV/ provenienti da prestiti si seleziona l'articolo determinativo maschile preconsonantico il ("il walkie-alkie), mentre, in linea con il comportamento vocalico dei legamenti di cui sopra, il lessico nativo seleziona l'allomorfo prevocalico l' in questo contesto ("l'uomo").

Baroni (2020) attribuisce la selezione di il davanti a /wV/ nei prestiti allo statuto consonantico del grafema <w> che rappresenta il legamento labiovelare. La lettera <w> implicherebbe un'interpretazione consonantica di /w/, altrimenti vocalica. Se fosse questo il caso, dovremmo osservare un effetto coarticolatorio sulla nasale che precede /w/ scritto <w> e contestualmente l'assenza di risillabificazione. Ci si aspetterebbe pertanto che davanti a parole come western, word e whiskey la nasale /n/ sia articolata nella regione velare (ad es. "un whiskey" > [uŋ wiski])

Per testare l'ipotesi del duplice statuto di /w/, consonantico o vocalico, a causa di un'influenza ortografica, è stata effettuata una raccolta dati che ha coinvolto otto parlanti (di cui quattro donne), studenti universitari di età compresa tra i 20 e i 30 anni. I partecipanti sono stati sottoposti a un test di lettura in cui gli stimoli, inseriti in una frase cornice, erano parole inizianti con /w/ precedute dall'articolo indeterminativo maschile "un" (ad. es. "devo dire un uomo tante volte", "devo dire un whiskey tante volte"). Oltre alle parole reali, erano presenti anche forme in cui <u> era sostituito da <w> e viceversa: "un womo", "un uischi". Come controllo, sono state utilizzate parole inizianti con consonanti occlusive sorde, velare e alveolare /k t/, come in "calo" e "toner", e le vocali /i u/, come in "inno" e "urlo". Questi filler sono stati utilizzati per effettuare il confronto tra le produzioni delle nasali prevelari /n#k/, prealveolari /n#t/ e prevocaliche /n#V/ con quelle relative alle nasali pre-legamento /n#w/.

È stata quindi svolta un'analisi acustica con il software Praat. I parametri maggiormente utilizzati per determinare il luogo di articolazione della nasale sono la prima larghezza di banda (BW1), la prima antiformante (FZ1) e F2 (Tabain et al., 2016). Inoltre, informazioni sul luogo di articolazione della nasale possono provenire dalle transizioni formantiche nel passaggio da vocale a nasale (Reetz, Jongman, 2020: 222-3). I valori di BW1 e FZ1 si sono rivelati scarsamente affidabili al fine di individuare il punto di articolazione di /n/ nei nostri dati, probabilmente a causa della posizione della nasale negli stimoli (posizione finale di determinanti, atoni) e della sua conseguente scarsa prominenza. Non abbiamo pertanto potuto svolgere un'analisi approfondita dei correlati acustici di cui sopra (cfr. Ohala & Ohala, 1993; Lai, 2009; Kerdpol, 2012). Al contrario, i valori di F2 e, soprattutto, della distanza tra F1 e F2 hanno mostrato un più alto grado di informatività. Anche la misurazione della variazione formantica, in particolare di F2, dal punto centrale all'offset della vocale prenasale, è risultato un metodo valido per differenziare il luogo di articolazione di /n/ nei dati raccolti, principalmente grazie alla presenza sistematica di /u/ nel nucleo prenasale.

Dalle nostre misurazioni è emerso che in contesto velare (/n#k/), il valore di åF della nasale si aggira intorno ai 400 Hz, mai superiore a 600 Hz, mentre in contesto alveolare, ossia /n#t, n#V/, il valore corrisponde mediamente a 900-1000 Hz, mai inferiore a 800 Hz (la differenza è statisticamente significativa, $p < 0,006$). Le misurazioni formantiche della nasale che precede il legamento /w/ mostrano un allineamento verso le produzioni relative al contesto velare, se il legamento corrisponde a <w> nello scritto. Viceversa, i valori ottenuti per la nasale pre-legamento, quando questo è scritto con il grafema vocalico <u>, sono in linea con quelli relativi al contesto alveolare. La diffrenza tra i valori di åF della nasale davanti a /w/ scritto <w> e davanti a /w/ scritto <u> raggiunge la significatività statistica ($p < 0,006$). Per quanto riguarda le transizioni formantiche, notiamo un innalzamento di F2 di circa 300 Hz verso la fine della vocale quando questa ricorre davanti a /n#t/ e /n#V/, mentre si riscontra un abbassamento di F2 davanti al nesso /n#k/. Anche in questo caso emerge un parallelismo tra il contesto /n/#<w> e il contesto velare (/n#k/) e tra /n/#w/ dove il legamento corrisponde a <u> e il contesto alveolare (/n#t/, /n#V/), corroborando ulteriormente l'interpretazione velare e alveolare, rispettivamente, della nasale.

È possibile fornire una spiegazione di tipo fonologico per tenere conto dei risultati ottenuti. Proponiamo infatti che, quando è scritto <w>, /w/ riceve interpretazione consonantica e occupa di conseguenza la posizione di attacco. Per quanto riguarda la nasale, l'italiano non ammette nessi /nC/ tautosillabici. Conseguentemente, nel contesto /n/#<w>, la nasale occuperebbe la posizione di coda sillabica, contesto in cui è soggetta all'assimilazione del luogo di articolazione della consonante seguente. La sua realizzazione sarà pertanto velare [ŋ], se seguita dall'approssimante labiovelare /w/. Quando il legamento /w/ è scritto invece <u>, la sua interpretazione è vocalica e occupa quindi la posizione di nucleo sillabico. Il contesto /n#V/ favorisce la risillabificazione post-lessicale della nasale, che la conduce a occupare la posizione di attacco e a emergere di conseguenza come alveolare [n], in linea con quanto osservato nei casi di "un inno" e "un urlo" nei nostri dati. L'ipotesi della diversa sillabificazione di /w/ dipendente dalla sua grafia, consonantica o vocalica, è in grado di tener conto anche del diverso allomorfo selezionato davanti a /w/, scenario discusso sopra.

Bibliografia

Baroni, A. (2020), Le semivocali in francese e italiano. Alternanza con le vocali alte e selezione dell'articolo, in M. Bacquin, P. Bernardini, V. Egerland, & J. Granfeldt (a cura di), *Écrits sur les langues romanes à la mémoire d'Alf Lombard*, Media-Tryck, Lund University, Lund, 17-42.

Bertinetto, P. M. (2010). Fonetica italiana, in *Quaderni del laboratorio di linguistica*, 9(1), 1-30.

Bertinetto, P. M., e Loporcaro, M. (2005), The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome, in *Journal of the International Phonetic Association*, *35*(2), 131–151.

Celata, C., Calamai, S., Ricci, I., e Bertini, C. (2013), Nasal place assimilation between phonetics and phonology: An EPG study of Italian nasal-to-velar clusters, in *Journal of Phonetics*, *41*(2), 88–100.

Kerdpol, K. (2012), Formant transitions as effective cues to differentiate the places of articulation of Ban Pa La-u Sgaw Karen nasals, in *Manusya: Journal of Humanities*, *15*(2), 21–38.

Lai, Y. (2009), Acoustic correlates of Mandarin nasal codas and their contribution to perceptual saliency, in *Concentric: Studies in Linguistics*, *35*(2), 143–166.

Ohala, J. J., e Ohala, M. (1993), The phonetics of nasal phonology: Theorems and data, in M. K. Huffman &

R. A. Krakov (a cura di), *Nasals, nasalization, and the velum* (Vol. 5), Elsevier, Cambridge, 225– 249.

Reetz, H., e Jongman, A. (2020), *Phonetics: transcription, production, acoustics and perception*, John Wiley & Sons Inc.

Tabain, M., Butcher, A., Breen, G., Beare, R. (2016), An acoustic study of nasal consonants in three Central Australian languages, in *Journal of Acoustical Society of America*, 139(2), pp. 890-903

Effects of lexical status and neighborhood in the realisation of phonemic contrasts in Italian

Eleonora Ridolfi*, Chiara Celata*, Olga Dmitrieva°
** Università degli studi di Urbino Carlo Bo; ° Purdue University*

Le proprietà lessicali e statistiche delle parole influenzano il modo in cui esse vengono prodotte, e gli ascoltatori sono in grado di discriminare differenze acustiche sottili che sono sistematicamente legate a tali proprietà, usandole per interpretare correttamente il significato. Il dettaglio fonetico varia sistematicamente in rapporto a fattori come la frequenza lessicale, la lunghezza della parola, l'ortografia e il significato pragmatico (Sereno & Jongman 1995, Gahl 2008, Drager 2011, Cohen Goldberg 2015). Ad esempio, Martinuzzi & Schertz (2021) hanno trovato che molteplici differenze acustiche subfonemiche distinguono sorry come richiesta di scuse da sorry come segnale di richiesta di attenzione, e queste differenze sono utilizzate dagli ascoltatori per identificare correttamente la parola. Similmente, la densità del vicinato lessicale e la presenza di vicini minimi (minimal competitors) influenzano il modo in cui i suoni vengono prodotti (Baker & Bradlow 2009, Gahl et al. 2012, Goldrick et al. 2013, Clopper & Tamati 2014). Ad esempio, in inglese le occlusive sorde hanno un VOT più lungo quando sono incluse in parole che hanno un vicino minimo con occlusiva sonora (es. teen, che contrasta con dean) rispetto a quando sono incluse in parole che non ce l'hanno (es. table; *dable) (Baese-Berk & Goldrick 2009). Questo insieme di risultati sembra perciò indicare che il dettaglio fonetico graduale è funzionale alla realizzazione dei contrasti paradigmatici tra le parole.

In questo studio ci chiediamo se lo status lessicale e la presenza di vicini minimi influenza la realizzazione acustica del contrasto fonologico tra consonanti scempie e geminate in italiano. Lo studio riprende l'impostazione di Celata et al. (2018) sul contrasto di sonorità nelle occlusive italiane, ed è ancora in corso; ci limitiamo qui a presentare la struttura dell'esperimento e il modo in cui saranno sottoposte a verifica le ipotesi sperimentali.

Ottanta parole reali e non-parole con scempia e geminata intervocalica ([n] vs. [n:], [t] vs. [t:]) sono state prodotte in un test di lettura da 30 locutori di una varietà centrale di italiano. Gli stimoli sono stati selezionati per far parte di quadruplette come quella in Tabella 1. Ogni quadrupletta comprende una coppia minima di parole realmente esistenti (es. cane - canne), una coppia di non-parole, foneticamente simili alle parole reali (es. *bane - *banne), e due ulteriori coppie di parole, in ciascuna delle quali una è una parola realmente esistente e l'altro no (es. tane - *tanne, dove la parola reale contiene la scempia, e *zane - zanne, dove la parola reale contiene la geminata). Questo disegno fattoriale permette di valutare sia gli effetti dello status lessicale (parola vs. non-parola), sia quelli della presenza o assenza di un vicino minimo sul dettaglio acustico nella produzione del contrasto di geminazione. Gli stimoli sono bilanciati per frequenza lessicale e probabilità fonotattica. I membri di ogni coppia minima sono stati divisi in due liste sperimentali in modo tale che ogni partecipante fosse esposto solo a un membro di ogni coppia. Ogni lista sperimentale conteneva 40 stimoli con scempia o geminata e 82 filler bisillabici con struttura fonotattica simile. Ogni lista è stata randomizzata in tre randomizzazioni differenti; ogni partecipante era esposto a tutte e tre le randomizzazioni, dunque leggeva ogni parola della lista tre volte, in un ordine diverso. Ogni stimolo della lista era presentato in forma isolata, al centro dello schermo del pc e a distanza di 2 secondi dal precedente; ai partecipanti era richiesto di leggere ad alta voce ogni stimolo al momento della sua comparsa sullo schermo. Il corpus sperimentale finale si compone di 3600 stimoli totali (40 stimoli x 3 ripetizioni x 30 soggetti).

Per ogni stimolo, è stata misurata in Praat (Boersma & Weenink 2008) la durata della consonante scempia o geminata e quella della vocale tonica precedente. L'analisi, tuttora in corso, comprende due modelli lineari misti, per due diverse variabili dipendenti: durata di C, durata di V. I fattori fissi sono status (parola vs. non- parola), lunghezza (scempia vs. geminata) e vicinato (presenza vs. assenza di un vicino minimo), tipo di consonante (occlusiva vs. nasale), anche in interazione tra di loro.

Le attese sono le seguenti. Se le proprietà lessicali dello stimolo influiscono sulla realizzazione del contrasto di lunghezza, ci aspettiamo che in presenza di un vicino minimo nel lessico le differenze

acustiche tra stimolo con scempia e stimolo con geminata siano maggiori (ipotesi della dissimilazione): geminate più lunghe, scempie più brevi (e l'opposto per le rispettive V precedenti). Si può inoltre ipotizzare che l'effetto del vicinato sia esclusivo (o più forte) nel caso delle parole, rispetto alle non-parole. In alternativa, o in aggiunta, si può ipotizzare che la presenza di un vicino minimo rallenti in generale l'accesso lessicale (per effetti di competizione/inibizione nella selezione lessicale) e ciò si rifletta in durate complessivamente maggiori, sia per le scempie che per le geminate (ipotesi dell'interferenza). I risultati così ottenuti saranno commentati rispetto all'ipotesi generale che la realizzazione acustica dei contrasti fonologici vari in funzione delle conoscenze implicite che i parlanti hanno del lessico e delle proprietà statistiche delle parole.

	scempia		geminata	
	parola	non-parola	parola	non-parola
Coppia 1	cane		canne	
Coppia 2		*bane		*banne
Coppia 3	tane			*tanne
Coppia 4		*zane	zanne	

Tabella 1. Esempio di quadrupletta di coppie minime usate nell'esperimento. Gli stimoli (sia parole esistenti che non-parole) accoppiati a una parola esistente sono considerati stimoli con vicino minimo per lunghezza consonantica.

Bibliografia

Celata, C., Dmitrieva, O., Meluzzi, C. V. C., & Concu, V. (2018). The effects of lexical status and lexical competitors on production of Italian stops. Poster presented at 16th Conference on Laboratory Phonology, Lisbon, 19- 22 June 2018.

Baese-Berk, M., & Goldrick, M., 2009. Mechanisms of interaction in speech production. Language and Cognitive Processes, 24(4), 527-554.

Baker, R. E., & Bradlow, A. R., 2009. Variability in word duration as a function of probability, speech style, and prosody. Language and Speech, 52(4), 391-413.

Boersma, P., & Weenink, D., 2008. Praat: doing phonetics by computer. Computer program.

Clopper, C.G., & Tamati, T.N., 2014. Effects of local lexical competition and regional dialect on vowel production. The Journal of the Acoustical Society of America, 136(1), 1-4.

Cohen-Goldberg, A.M., 2015. Abstract and Lexically Specific Information in Sound Patterns: Evidence from /r/-sandhi in Rhotic and Non-rhotic Varieties of English. Language and Speech, 58(4), 522–548.

Drager, K., 2011. Sociophonetic variation and the lemma. Journal of Phonetics, 39, 694-707.

Gahl, S., 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. Language, 84(3), 474-496.

Gahl, S., Yao, Y., & Johnson, K., 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. Journal of Memory and Language, 66(4), 789-806.

Goldrick, M., Vaughn, C., & Murphy, A., 2013. The effects of lexical neighbors on stop consonant articulation. The Journal of the Acoustical Society of America, 134(2), EL172-EL177.

Martinuzzi, C., Schertz, J., 2022. Sorry, Not Sorry: The independent role of multiple phonetic cues in signaling the difference between two word meanings. Language and Speech, 65(1), 143–172.

Sereno, J. A., Jongman, A., 1995. Acoustic Correlates of Grammatical Class. Language and Speech, 38(1), 57–76.

Intonational variation in right dislocated interrogatives: A study on Veneto Italian

Claudia Crocco[1], Barbara Gili Fivela[2], Mariapaola D'Imperio[3], Giuseppe Magistro[1]
[1] *Ghent University,* [2] *University of Salento,* [3] *Université & Laboratoire Parole et Langage (LPL)*

In this paper we present a phonological analysis of yes-no questions (YNQ), comparing the intonational organization of broad focus YNQs with unmarked word order (Subject-Verb-Object: SVO), with the prosodic pattern found in YNQs containing a Right Dislocation (RD). Previous research on RD [1, 2] indicate the intonational distinction between broad and narrow informational focus may not be clear-cut in several Italian varieties. In this study we explore this hypothesis on a corpus of dialogical, semi-spontaneous and read speech. Since the lack of a clear-cut distinction between broad and narrow focus intonation seems to be a feature of Italian intonation as a whole [1], we consider more than one variety in this study, taking a group of Venetan varieties (Padua, Venice, Vicenza) as a case study. We examine the phonological organization of the tonal prominences in the relevant contexts, leaving out variety-specific features such as pitch accents structure and inventory.

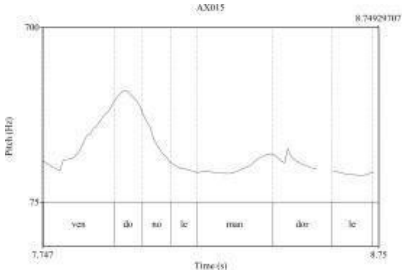
It is generally assumed [3, 4] that Italian SVO (e.g. *Maria legge il libro*, 'Mary reads the book') and RD utterances (e.g. *Maria lo legge, il libro* 'Mary reads [it] the book') differ at the prosodic level for their internal phrasing and the position of the main prominence. In particular, RDed sentences, a boundary is expected to precede the RDed constituent(s), while in SVO sentences, O is not phrased apart from the preceding V [5]. Moreover, while in broad focus SVO sentences the main prominence is expected on the rightmost lexical stress (O), in RDed structures the main accent is located within the core of the clause, before the RDed constituent, which in contrast is lowered in pitch [1, 3, 4, 6, cf. also 7]. From the prosodic point of view, therefore, RD are generally considered as narrow focus structures. However, when a RD occurs in a question (e.g. *lo vedi il libro?*, 'do you see [it] the book?'), there is evidence that it can be realized with an unmarked prominence pattern [1], suggesting that clause type (interrogative vs. declarative) may affect the prosody of a RD sentence. This leads to the hypothesis that the prosodic difference between RD and SVO could be blurred in questions, independently of the speech style.

To explore this hypothesis, we examine the tonal organization of such cases, by comparing verb+ object (VO) and right-dislocated (i.e., obj. clitic + verb + co-referent obj.: clVO) structures in read, scripted and dialogical speech in Veneto Italian, including the varieties spoken in Venice, Padua and Vicenza. The analysis is based on a corpus of semi-spontaneous speech elicited using a *Discourse Completion Task* combined with dialogical and read speech [8, 9]. Approx. 250 YNQs produced by 13 speakers (6F; 7M), age ranging between 30 and 40, with university- level education. We examined 4 syntactic-phonological structures: (i) broad focus YNQs with SVO (unmarked word order for Italian); (ii) YNQs with RDed O; (iii) YNQs with RDed S, and (iv) YNQs with RD of both O and S. The contexts used for DCT and reading task were designed in a comparable way to elicit the desired pragmatic interpretation. The dialogues are extracted from the CLIPS corpus [10], and are collected using the *map task* and the *spot the difference* paradigms. The phonological analysis proposed in this study builds on the available literature on the intonation of Veneto Italian [11, 12].

Preliminary results indicate that the realization of RDed and SVO YNQs may vary. While [1] highlighted the possibility for RDed YNQs to be realized with the unmarked intonational pattern of SVO questions, the results of the present study indicate that in SVO YNQs the largest pitch movement can be located on V rather than on the rightmost lexical stress (O; cf.

fig.1, 3 and 4, read and dialogical speech), resulting in a pattern similar to the one used for RDs (cf. fig.2). This SVO pattern, which has also been previously observed in [13], is not meant to induce a pragmatic narrow focus on V.

The observed variability in the intonation pattern of SVO and RDed YNQs calls for a phonological explanation to address the issue of the metrical relationship between the tonal prominences on the verbal phrase and the RDed constituent. Along the lines of [9] we propose the accent on the RDed functions as an intonational nucleus in association with the preceding accent.



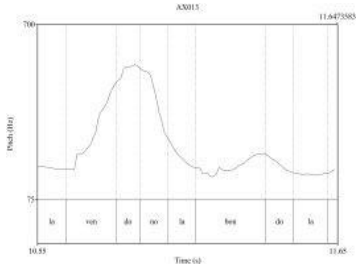
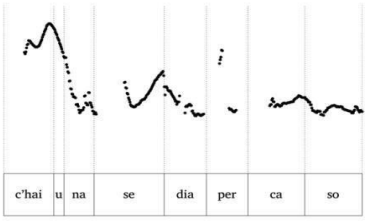


Figure 1. YNQ with (S)VO word order: vendono le mandorle? *Figure 2. YNQ with RD of O: (I Tebaldi) la vendono la bondola?*



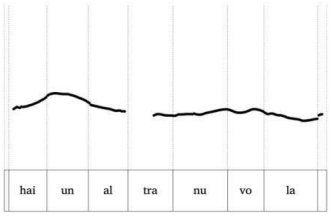


Figure 3. YNQ with SVO order -dialogical speech (Venice) *Figure 4. YNQ with SVO order -dialogical speech (Venice)*

References

[1] Crocco, C. 2013. Is Italian Clitic Right Dislocation grammaticalized? A prosodic analysis of yes/no questions and statements. *Lingua* 133: 30-52.

[2] Crocco, C. Badan L. 2020. Dislocazioni a destra interrogative tra grammatica e discorso. Revue Romane. Published online: 12 June 2020. https://doi.org/10.1075/rro.18017.cro

[3] Cardinaletti A. 2001. A second thought on Emarginazione: Destressing vs. "Right Dislocation". In Cinque G., Salvi G. (Eds.) *Current Studies in Italian Syntax. Essays Offered to Lorenzo Renzi*. Amsterdam, Elsevier: 117-135.

[4] Cardinaletti A. 2002. Against optional and zero clitics. Right Dislocation vs. Marginalization. *Studia Linguistica* 56.1: 29-57.

[5] D'Imperio, M., G. Elordieta, S. Frota, P. Prieto & M. Vigário. 2005. Intonational Phrasing in Romance: The role of prosodic and syntactic structure. In S. Frota, M. Vigário & M. J. Freitas (Eds.), *Prosodies. Phonetics & Phonology Series*. Berlin: Mouton de Gruyter: 59- 97.

[6] Bocci, Giuliano. 2013. *The syntax-prosody interface from a cartographic perspective: evidence from Italian.* Amsterdam/Philadelphia: John Benjamins.

[7] D'Imperio, M. 2002. Italian intonation: An overview and some questions. *Probus* 14(1): 37-69.

[8] Blum-Kulka, S., House, J., & Kasper, G. 1989. Cross-cultural Pragmatics: Requests and Apologies.
Norwood, NJ: Alblex Publishing Corporation.

[9] Gili-Fivela B. et alii. 2015. Varieties of Italian and their intonational phonology. In S. Frota & P. Prieto (Eds.), Intonational Variation in Romance. Oxford: OUP: 140-197.

[10] CLIPS*: Corpora e Lessici di Italiano Parlato e Scritto*. http://www.clips.unina.it [accessed on 7/11/22].

[11] Savino, M. 2012. The intonation of polar questions in Italian: Where is the rise? Journal of the International Phonetic Association 42(1):23-48.

[12] Badan, Linda, and Claudia Crocco. 2019. Focus in Italian Echo Wh-Questions: An Analysis at Syntax-Prosody Interface. *Probus* 31 (1): 29–73.

[13] Herraiz Devis, E. 2007. La prosodia nell'interferenza tra L1 e L2: il caso delle interrogative polari tra veneti e catalani. *Estudios de fonética experimental*, XVI:119- 146.

IL PARLATO IN AMBITO MEDICO:
Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD:
Linguistic analysis, technological applications and clinical tools

THURSDAY 16 FEBRUARY
KEYNOTE SPEECHS AND ORAL PRESENTATIONS

On Deep Neural Networks for Parkinson Detection in Spanish and Italian

Raffaello Caserta, Marco Siniscalchi
Università degli Studi di Enna Kore

Extensive research has been conducted to develop accessible, affordable, and non-invasive techniques for automatic detection of Parkinson's disease. It has been shown that speech can be used to distinguish between healthy patients (control) and afflicted patients (patients), and by exploiting deep neural networks several substantial steps forward have been made in both automatic speech recognition and synthesis. Most of the research effort has however focused on well-crafted speech corpora, where clean audio and corresponding transcripts are made available to machine learning practitioners. In real-world scenarios, it is nevertheless highly unlikely to acquire patients' spoken utterances in a quite and favourable environment. Therefore, there is very little material available to train robust automatic system for Parkinson disease detection. In addition to that, Italian accents and dialects introduce sociolinguistic features that make even the simple binary detection (control vs patient) task even more challenging. We leverage knowledge-transfer and convolutional neural networks (CNNs) to build the first automatic Parkinson detection system based on speech for the Italian dialect spoken in Salento. We will show that despite the poor quality of the audio and the strong sociolinguistic characteristics, our preliminary results are encouraging and worth of further investigation.

The grammar of (dys)prosody: a link between perception and production?

Sonia Frota
University of Lisbon

Prosody is an essential component of language. The prosodic grammar of a given language establishes the patterns of intonation, prominence, and speech chunking that characterize the language, as well as the prosodic form-meaning relations that convey linguistic and communicative meanings. Prosodic impairments may thus seriously disrupt the ability to successfully interact with others. Dysprosody has been described as a common feature in motor speech disorders, including Parkinson's disease (PD). However, the ability of PD patients to use the prosodic categories and structures of the prosodic grammar of their native language, beyond the acoustics of prosody, remains largely unknown. Following methods of the prosodic and intonational phonology frameworks, I will present findings showing that PD patients, compared to healthy controls, have a decreased ability to use intonation and speech chunking. Medication improved intonation but did not help with dysprosodic chunking. In turn, disease duration and motor fluctuations affected chunking patterns but had no impact on intonation. Moreover, and exploring potential links between perception and production, the impaired production data was related to language/speech measures of self-reports, clinical assessment, and speech perception by other (healthy) speakers. For the latter, perception studies were conducted. The results suggest that impairments in intonation are closely associated to self-reports of speech and communication difficulties (especially before medication intake). Findings from the perception studies indicate that PD patients are clearly distinguished from healthy controls, both on the basis of intonation and chunking patterns. Overall, the grammar of prosody is impaired in PD, and different prosodic categories and structures (intonation, chunking) may be differently affected in production, as well as show distinct impact on self-perception and the perception of others of PD patients' speech and communication difficulties. Implications to the understanding of the mechanisms underlying prosodic impairments, together with language-driven approaches to their treatment in PD, will be discussed.

Cross-Lingual Transferability of Voice Analysis Models: a Parkinson's Disease Case Study

Claudio Ferrante and Vincenzo Scotti
DEIB, Politecnico di Milano

Traditionally, speech analysis has always relied on a set of very informative features like (Mel) spectrogram or Mel Fre- quency Cepstral Coefficients (MFCC) to solve many different problems and build incredible speech powered applications (Jurafsky and Martin, 2009). In this sense, speech analysis distinguished itself from other areas, like computer vision, where deep learning has become a pervasive framework. However, recently, deep learning models for the extraction of acoustic features have allowed improving significantly the state of the art in many speech-related applications, like Automatic Speech Recognition (ASR) (Baevski et al., 2020; Radford et al., 2022), speaker identification (Wan et al., 2018), or speech emotion recognition (Scotti et al., 2020). Basically, these models are deep neural networks trained in an unsuper- vised or self-supervised (Aytar et al., 2016; Hershey et al., 2017; Baevski et al., 2020) way on large collections of acoustic data (not necessarily speech). Starting from the features computed by these deep neural network models, we can build classification and regression models to solve speech analysis-related problems with improved results.

With our work, we focus the analysis on the cross-lingual transferability of these deep learning features for speech analysis. The idea is to understand whether and how well a classification model trained on deep acoustic features in a source language works in a new target language. The problem we are dealing with is thus a domain adaptation problem (Redko et al., 2019). We selected Parkinson's disease recognition as the use case of our experiments: the goal is to train a binary classifier that discriminates from speech between healthy patients and patients affected by Parkinson's disease.

We used two data sets for our experiments: one in English and one in Telugu, the latter served as an under-resourced language example. We experimented using both languages first as the source domain data and as the target domain data in two sets of experiments. The English data set is the Mobile Device Voice Recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls (hag). It contains both spontaneous and read audio clips from both Parkinson's disease patients and healthy patients, respectively it contains 266 clips of healthy persons and 199 clips of persons affected by Parkison's disease. The Telugu data set, instead, is a private collection of audio clips from Parkinson's disease patients (81 clips); we combined it with samples taken from the Telugu split of Open SLR (He et al., 2020), a data set for speech recognition (we subsampled 200 clips from it). The latter language motivated our work; in fact, the scarcity of data for some problems or languages usually forces us to adapt solutions from different domains.

Building a classifier to recognise Parkinson's disease from the speech is not a new problem. This task has already been addressed using machine learning-based techniques (Williamson et al., 2015; Karan et al., 2020; Rahman et al., 2021; Kurada and Kurada, 2020; Toye and Kompalli, 2021). Usually, these approaches involve the extraction of acoustic and prosodic features (mainly MFCC, Jitter, Shimmer, and Pitch), which allows the building of very discriminative models. In some cases, these features are further processed through dimensionality reduction transformations to keep only the most relevant components of the vectors encoding the audio clips to classify.

Very recent results already explored the role of deep learning features to learn a classifier for this Parkinson's disease detection problem (Kurada and Kurada, 2020), reaching more than 80% recognition accuracy on the same English data set we considered and outperforming the other considered classifiers based on acoustic features. Other works started focusing on building classifiers for other languages (Toye and Kompalli, 2021): their results showed that acoustic and prosodic features could be used to build high-performing classifiers (reaching more than 90% recognition accuracy) on English and Italian. However, none of these previous works analysed the effects of directly transferring trained classification models for Parkison's disease recognition across languages (usually models are trained from scratch in the new language, which isn't always possible due to data availability problems) or the

effects of applying domain adaptation on these classifiers to see how it affects the performances in the target language, without explicit retraining. With this work, we are interested in experimenting with classification models to tackle these two analyses

The classification pipeline we propose is completely based on data-driven techniques. The first step we apply is denoising through RNNoise (Valin, 2018), a deep neural network for voice enhancement and noise suppression. We process the denoised audio directly with one of the selected deep features models, which yields a sequence of feature vectors. Each sequence vector covers a specific time window of the input signal. Since the final classification step of machine learning models expect information to be encoded into a single vector, we tested different approaches to convert the sequence of feature vectors into a single vector: flattening (or unrolling), Global Average Pooling (GAP) (Lin et al.,2014), and Global Max Pooling (GMP) (Lin et al., 2014). Finally, we used a Feed Forward Neural Network to perform the classification task from the extracted feature vectors.

As premised, we considered also a technique for unsupervised domain adaptation, Deep CORAL (Sun and Saenko, 2016), as an alternative path of our pipeline. We re-trained the classification network adding the Deep CORAL objective on the covariance of the final projections before classification for adaptation. We considered this adaptation step to understand how robust the classifier is with and without explicit adaptation.

In our work, we considered three different models for computing deep acoustic features: Wav2Vec (Baevski et al., 2020), VGGish (Hershey et al., 2017), and SoundNet (Aytar et al., 2016). The former was specifically designed for speech analysis problems and is used as input for state-of-the-art ASR models. The other two instead are actually mode generic models though for acoustic analysis not necessarily aimed at speech. VGGish was trained on a large audio classification data set containing many labels. SoundNet, on the other hand, was trained to predict from the audio track of video clips the pseudo labels generated from object recognition and scene recognition using a neural network that processed the images of the video clips. Additionally, we trained some classification models using traditional acoustic and prosodic features. We selected the set of features from recent works on the same problem and data set (Toye and Kompalli, 2021).

The results on individual languages we obtained are in line with previous works. Even with few samples, with Telugu, we got > 90% accuracy with some feature and pooling configurations. In English, we reached > 90% accuracy as well. We achieved these results using either traditional or deep features, showing that for this kind of speech classification problems, transferring deep features is not necessary to achieve the best results. Concerning cross-lingual results, we noticed that in some cases, without explicit adaptation, many classifiers still achieve reasonable results (≥ 70% accuracy) when applied to the other (target) language. From the results of domain adaptation, we noticed that when the zeroshot behaviour of the unadapted classifier produced low scores on the target language, adaptation helped improve the performances, lowering the scores on the source language though. The results of adaptation are mixed, and a more extensive analysis of these techniques is necessary to understand how to achieve the best from these approaches. As for now re-training and zero-shot model transfer seem to be the most effective solutions. For completeness and reproducibility, we share the reference to the repository with the source code (https://github.com/vincenzo-scotti/voice_analysis_parkinson).

References

Yusuf Aytar, Carl Vondrick, and Antonio Torralba. Soundnet: Learning sound representations from unlabeled video. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 892–900, 2016.

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Fei He, Shan-Hui Cathy Chu, Oddur Kjartansson, Clara Rivera, Anna Katanova, Alexander Gutkin, Isin Demirsahin, Cibu Johny, Martin Jansche, Supheakmungkol Sarin, and Knot Pipatsrisawat. Open-source multi-

speaker speech corpora for building gujarati, kannada, malayalam, marathi, tamil and telugu speech synthesis systems. In Nicoletta Calzolari, Fre´de´ric Be´chet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, He´le`ne Mazo, Asuncio´n Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 6494–6503. European Language Resources Association, 2020.

Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin W. Wilson. CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2017, New Orleans, LA, USA, March 5-9, 2017*, pages 131–135. IEEE, 2017.

Dan Jurafsky and James H. Martin. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition*. Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education International, 2009.

Biswajit Karan, Sitanshu Sekhar Sahu, and Kartik Mahto. Parkinson disease prediction using intrinsic mode function based features from speech signal. *Biocybernetics and Biomedical Engineering*, 40(1):249–264, 2020.

Sruthi Kurada and Abhinav Kurada. Poster: Vggish embeddings based audio classifiers to improve parkinson's disease diagnosis. In *5th IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2020, Crystal City, VA, USA, December 16-18, 2020*, pages 9–11. IEEE, 2020.

Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. *OpenAI blog*, page 28, 2022.

Wasifur Rahman, Sangwu Lee, Md Saiful Islam, Victor Nikhil Antony, Harshil Ratnu, Mohammad Rafayet Ali, Abdullah Al Mamun, Ellen Wagner, Stella Jensen-Roberts, Emma Waddell, et al. Detecting parkinson disease using a web-based speech task: Observational study. *Journal of medical Internet research*, 23(10):e26305, 2021.

Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younes Bennani. *Advances in domain adaptation theory*. Elsevier, 2019.

Vincenzo Scotti, Federico Galati, Licia Sbattella, and Roberto Tedesco. Combining deep and unsupervised features for multilingual speech emotion recognition. In Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani, editors, *Pattern Recognition. ICPR International Workshops and Challenges - Virtual Event, January 10-15, 2021, Proceedings, Part II*, volume 12662 of *Lecture Notes in Computer Science*, pages 114–128. Springer, 2020.

Baochen Sun and Kate Saenko. Deep CORAL: correlation alignment for deep domain adaptation. In Gang Hua and Herve´ Je´gou, editors, *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*, volume 9915 of *Lecture Notes in Computer Science*, pages 443–450, 2016.

Adedolapo Aishat Toye and Suryaprakash Kompalli. Comparative study of speech analysis methods to predict parkinson's disease. *CoRR*, abs/2111.10207, 2021.

Jean-Marc Valin. A hybrid dsp/deep learning approach to real-time full-band speech enhancement. In *20th IEEE International Workshop on Multimedia Signal Processing, MMSP 2018, Vancouver, BC, Canada, August 29-31, 2018*, pages 1–5. IEEE, 2018.

Li Wan, Quan Wang, Alan Papir, and Ignacio Lopez-Moreno. Generalized end-to-end loss for speaker verification. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*, pages 4879–4883. IEEE, 2018.

James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Joseph Perricone, Satrajit S. Ghosh, Gregory A. Ciccarelli, and Daryush D. Mehta. Segment-dependent dynamics in predicting parkinson's disease. In *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015*, pages 518–522. ISCA, 2015.

Phonetic analysis of dysarthric speech in Parkinson's Disease: framing individual strategies

Barbara Gili Fivela[1], Raffaello Caserta[3], Salvatore Monteleone[2], Sonia d'Apolito[1], Anna Chiara Pagliaro[1], Marco Siniscalchi[3]
[1]University of Salento – CRIL-DReAM, Lecce, Italy, [2]Niccolò Cusano University, Rome, Italy
[3]Kore University of Enna, Italy

Speech accuracy is considered a good metric for assessing the severity level of motor speech disorders, such as dysarthria. Among the objective methods available to assess accuracy, phonetic analysis provides very detailed information, shedding light on individual strategies, even with reference to specific speech sounds or gestures. For instance, acoustic studies have shown that vowels produced by parkinsonian dysarthric speakers (PDs) occupy a reduced acoustic space compared to those of healthy controls (HCs), show specific acoustic characteristics, such as increases in jitter and noise-to-harmonics ratio (a.o., [1]), or larger horizontal tongue gestures than control's [2]. Phonetic investigation results may be correlated to severity levels that are actually guaranteed by overall, usually subjective evaluations of dysarthria (e.g., Robertson's Dysarthria Profile [3, 4]) or that may offer an overall and not speech-specific or dysarthria-specific evaluation (e.g., UPDR and Hoehn and Yahr [5,6]). Nevertheless, especially in the case of articulatory studies, investigated speech sounds or even populations may often not be wide enough to allow for robust generalization [7, 8, 2]. On the other hand, automatic learning algorithms are also used for objectively evaluating speech accuracy by considering global speech characteristics, even though no detailed observation on specific sound properties is provided [13, but see also 9]. A question then arises as to whether the speech events under investigation in phonetic studies are representative of the overall speech accuracy or if they rather describe local, individual strategies that do not necessarily fit the severity level assigned to the experimental subject.

In this paper, a comparison is offered of a phonetic (acoustic and articulatory) analysis and an automatic evaluation of speech aiming at telling apart productions by dysarthric and healthy speakers and possibly rating the severity level of dysarthria. The goal is to observe if and how the two classifications perform differently, given their obvious specificities, and if they may be integrated. The hypothesis is that the two procedures do not always classify speech samples consistently, given the abovementioned specificities. Still, they may be integrated, offering equally valuable, though different, evaluations of the same speech.

To reach our goal we compared the phonetic and automatic analysis of a set of three-word nominal phrases in Italian, such as *Lu/a X blu* 'The blue X', where Xs are ˈCVCV disyllables, Vs are /a/ - /u/ or /u/ - /a/, and initial or intervocalic Cs may be voiced or unvoiced bilabial stops (/p/, /b/). Target disyllables are [ˈpu.pa], [ˈpa.pu], [ˈbu.ba], [ˈba.bu]. The corpus under investigation was acoustically and articulatorily (AG501) recorded by four mild-to-severe dysarthric PD speakers from Lecce – Southern Italy (age 65-80; two PDs were diagnosed a mild-to-severe and two a mild to moderate hypokinetic dysarthria) and four peer HC from the same geo-linguistic area (matching age), who read aloud (7 times) the target sentences. The phonetic investigation focused on the use of the phonetic vowel space and the possible influence of unvoiced and voiced bilabials, more specifically if PDs show unexpectedly greater amplitude in PDs than in Controls tongue gestures along the horizontal, anterior-posterior dimension (as shown in [2, 10] on different vowel cycles). The automatic evaluation was performed on samples by one of the PD speakers. The automatic system was built around a convolutional neural model trained in a cross-language setting. More specifically, the convolutional neural model was first bootstrapped on the PC-GITA database [11], which is a Spanish speech corpus database for the analysis of people suffering from Parkinson's disease covering data from 50 PD speakers, and 50 HC speakers. Next, speakers which were not used in the testing phase, were rather used to fine-tune the neural parameters. Since the Italian data were extremely noisy, a preprocessing step was carried out using a deep neural network (DNN) based speech enhancement system [12] with the goal of improving the sound quality before PD detection.

Despite the scarcity of data used to train the automatic classifier, results confirm the lack of systematic correspondence between phonetic and automatic results. The experimental results show that the proposed cross-language approach is viable and high detection accuracy can be attained. The in-depth analysis offered by phonetics shows speakers' specific strategies that do not always match the dysarthria severity level clinically assigned to speakers. All in all, the automatic processing allows to frame the phonetic analysis within a wider objective evaluation of speaker's speech productions and offers support to the proposal of individual, different strategies by speakers belonging to the same severity group.

References

[1] Young-Im Bang, Kyunghoon Min, Young H Sohn, Sung-Rae Cho (2013). Acoustic characteristics of vowel sounds in patients with Parkinson disease. NeuroRehabilitation, 32(3):649-54.

[2] Fivela, B. G., d'Apolito, S. I., & Di Prizio, G. (2020). Labialization and Prosodic Modulation in Italian Dysarthric Speech by Parkinsonian Speakers: A Preliminary Investigation. In Proc. 10th International Conference on Speech Prosody 2020 (pp. 828-832).

[3] Robertson S.J. (1982) Dysarthria profile. Background and development. College Speech Therapist Bullettin, 359.3.82.

[4] Fussi F, Cantagallo A. (1999): *Profilo di valutazione della disartria. A*dattamento italiano del test di Robertson, raccolta di dati normativi e linee di trattamento (1982) Ed. Omega

[5] S. Skodda, W. Gronheit, and U. Schlegel, "Impairment of Vowel Articulation as a Possible Marker of Disease Progression in Parkinson's Disease," *PLoS ONE*, vol.7, no.2, pp.1–8, 2012Liss JM, Utianski R, Lansford K. Crosslinguistic application of English-centric rhythm descriptors in motor speech disorders. Folia Phoniatrica et Logopaedica. 2013; 65:3–19.

[6] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease (2003). State of the Art Review The Unified Parkinson's Disease Rating Scale (UPDRS): Status and Recommendations. *Movement Disorders,* Vol. 18, No. 7, 2003, pp. 738–750

[7] M.N. Wong, B.E. Murdoch, and B.M. Whelan, "Kinematic analysis of lingual function in dysarthric speakers with Parkinson's disease: An electromagnetic articulograph study,"*Int. JS-LPathology*, vol.12, pp.414–425, 2010.

[8] M.N. Wong, B.E. Murdoch, and B.M. Whelan, "Lingual Kinematics in Dysarthric and Nondysarthric Speakers with Parkinson's Disease," *Parkinson's Disease*, pp.1-8, 2011.

[9] N. Audibert, C. Fougeron, C. Fredouille, C. Meunier and O. Panseri, "Evaluation d'un alignement automatique sur la parole dysarthrique," Mons: Actes des 28e Journées d'Etudes sur la Parole (JEP'10), pp. 353-356, 2010.

[10] Fivela, B. Gili, et al. "Italian Vowel and Consonant (co) articulation in Parkinson's Disease: extreme or reduced articulatory variability?." *10th ISSP, May 5-8, Cologne, Germany, Proceedings* (2014): 146-149.

[11] J. R. Orozco-Arroyave et al., "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in Language Resources and Evaluation Conference, (LREC), 2014, pp. 342–347.

[12] S. M. Siniscalchi, "Vector-to-Vector Regression via Distributional Loss for Speech Enhancement," in *IEEE Signal Processing Letters*, vol. 28, pp. 254-258, 2021.

[13] Qiyue Wang, Yan Fu1, Baiyu Shao, Le Chang, Kang Ren, Zhonglue Chen and Yun Ling (2022). Early detection of Parkinson's disease from multiple signal speech: Based on Mandarin language dataset, Front Aging Neurosci. 14: 1036588.

Analisi di alcuni indici temporali per una valutazione dei disturbi della fluenza

Valentina De Iacovo, Dario Strangis, Antonio Romano
Università di Torino

Questo contributo riguarda un progetto che si propone l'obiettivo di dare una caratterizzazione sperimentale delle modalità con cui si manifestano i disturbi della fluenza in soggetti di madrelingua italiana in riferimento alle conoscenze elaborate e adottate nella pratica clinica in altri spazi linguistici.

I primi passi di uno studio-pilota sono stati indirizzati alla discussione dei principali modelli teorici e alla definizione di categorie di rappresentazione di alcuni frequenti fenomeni relativi alla fluenza già valutati da altri autori (Zmarich 1999, Manfrellotti *et al.* 2003, Dovetto 2013, Sorianello 2017) nell'ottica di una descrizione oggettiva delle qualità vocali e linguistiche di soggetti che si discostano da quelle descritte per un parlato "ortofonico" nei grandi progetti nazionali (Cresti & Moneglia 2007, Savy & Cutugno 2009).

In questo contributo ci proponiamo in particolare di dettagliare alcuni dei primi risultati che riguardano disfluenze riconducibili a fenomeni legati all'organizzazione temporale (pausazione, velocità d'eloquio) rilevabile negli atti individuali di un campione di soggetti interessati da balbuzie e/o cluttering.

Disturbi della fluenza quali il cluttering e la balbuzie possono essere distinti su base percettiva, in quanto il loro parlato ha caratteristiche differenti.

Le valutazioni degli aspetti udibili e visibili di entrambe le condizioni non sono però da considerarsi esaustive e devono essere sempre complementari a un'indagine delle caratteristiche non visibili e udibili, come le condotte di evitamento (es. sostituzione di una parola, circonlocuzioni, evitamento di una situazione comunicativa), gli aspetti cognitivi, affettivi e comportamentali, e più in generale l'impatto sulla qualità di vita e sulla partecipazione sociale della persona (Yaruss & Quesal 2004).

Se al livello internazionale diventa sempre più decisivo il riferimento a banche dati che permettano una comparazione più solida in termini quantitativi (Ratner & MacWhinney 2018), sembra mancare un riferimento omogeneoper la lingua italiana. Per questo motivo, in seguito a prime collaborazioni (Romano *et al.* 2011), abbiamo messo a punto un metodo oggettivo per la classificazione delle voci sulla base di riferimenti robusti in merito alla variazione di f_0 (Romano & De Iacovo 2021), una proposta di classificazione di vari fenomeni fonetici presenti nel parlato patologico (Vernero & Romano 2017), e incominciato a lavorare su velocità d'eloquio e pausazione (De Iacovo *et al.* 2020).

Se infatti da un lato possiamo definire qualsiasi tipo di parlato sulla base di parametri ricavabili acusticamente, risulta ancora difficile stabilire quantitativamente criteri o soglie che permettano di capire univocamente se la loro presenza/assenza o correlazione con altri parametri siaindicativo di undisturbo della fluenza specifico e quindi permettere una diagnosi differenziale chiara.

Con l'obiettivo di incrementare la banca dati, arricchendola di modalità di elicitazione diverse,mostriamo per ora i risultati ottenuti da un primo confronto basato sulla lettura di un testo di circa 500 parole svolta da 4 persone il cui eloquio rientra nel campo dei disturbi della fluenza, con caratteristiche percettive differenti.

Ci siamo proposti di analizzare alcuni casi in cui si manifestano interruzioni, esitazioni, rallentamenti e fenomeni che caratterizzano la costruzione/frammentazione degli enunciati di un testo.

Per accelerare la prima fase di rilevamento/etichettatura (anche in vista di applicazioni cliniche che non possono sottrarre tempo prezioso alla classificazione, investendosi in operazioni di segmentazione manuale) abbiamo fatto ricorso a segmentatori automatici.

In un primo momento si è voluto capire il grado di affidabilità della segmentazione automatica di WebMaus a partire dalla lettura di un testo di riferimento (Martino *et al.*, 2011), l'allineamento è stato quindi rivisto manualmente e sono stati etichettati i casi specifici di interruzione, riformulazione, ripetizione (sul modello di CLIPS).

Tralasciando altre interessanti caratteristiche del parlato (come gli indici di costruzione di Cresti & Moneglia 2007) che potrebbero essere considerate successivamente, ci siamo concentrati sulle misurazioni dei valori acustici della velocità d'eloquio e delle pause a partire dall'annotazione. Dalle prime valutazioni sommarie sono emersi indicatori promettenti, ma si sono anche presentate varie difficoltà che vorremmo esporre a una comunità scientifica plurale anche in vista di una convergenza su misure solidamente definite. Se, infatti, esistono dei riferimenti precisi (Zmarich *et al.* 1997) per la classificazione di pause brevi o pause lunghe, di *speech rate* o indice di scioltezza, *fluency rate*,o *mean articulatory rate* (van Zaalen & Reichel 2015, Cosyns *et al.* 2018), non sono ancora stati messi a punto

dei metodi di segmentazione ed estrazione automatica che permetterebbero invece l'analisi e il confronto di tali indici temporali nelle popolazioni oggetto di studio.

Riferimenti bibliografici

Ambrose N.G. & Yairi E. (1999). Normative disfluency data for early childhood stuttering. Journal of Speech, Language, and Hearing Research, 42(4), 895-909.

Cosyns M., Meulemans M., Vermeulen E., Busschots L., Corthals P. & Van Borsel J. (2018). Measuring Articulation Rate: A Comparison of Two Methods. *Journal of speech, language, and hearing research*, 61(11), 2772-2778.

Cresti, E. & Moneglia M. (2005). *C-ORAL-ROM: integrated reference corpora for spoken romance languages*. John Benjamins Publishing.

De Iacovo V., Colonna V. & Romano A. (2020). La pausazione. *Bollettino del LFSAG*, 5, 41-48.

Dovetto F. (2013). *Il parlar matto. Schizofrenia tra fenomenologia e linguistica: il corpus CIPPS*. Roma: Aracne.

Eggers K., Van Eerdenbrugh S., Byrd CT. (2020). Speech disfluencies in bilingual Yiddish-Dutch speaking children. *Clin. Linguist. Phon.*, 34(6), 576-592.

Manfrellotti O.M., Savy R., Gomez-Paloma F. (2003). Un procedimento di codifica fonologica basato sull'analisi segmentale del parlato di bambini ipoacusici, In: A. Regnicoli (a cura di), La fonetica acustica come strumento di analisi della variazione linguistica in Italia, Roma: Il Calamo, 211-218.

Martino M.G., Pappalardo F., Re A.M., TressoldiP.E., Lucangeli D. & Cornoldi C. (2011). La valutazione della dislessia nell'adulto. Un contributo alla standardizzazione della Batteria dell'Università di Padova. *Dislessia*, 8(2), 119-134.

RatnerN.B. & MacWhinney B. (2018). Fluency Bank: A new resource for fluency research and practice. *Journal of fluency disorders*, 56, 69-80.

Romano A., De Iacovo V. (2021). Statistiche di f0 per 200 parlanti di italiano. *Bollettino LFSAG*, 8, 21-33.

Romano A., Vernero I. & Strangis D. (in c. di p.). Trascrizione fonetica e parlato patologico: tra modelli teorici, indicazioni operative e prassi. Com. pres. al Convegno "(Tra) medici e linguisti (4)" (Napoli, Accademia Pontaniana, 13-14 dicembre 2021). In c. di pubbl. a cura di F. Dovetto, Medici e Linguisti.

Savy R., Cutugno F. (2009). CLIPS. Diatopic, diamesic and diaphasic variations in spoken Italian, in *Proceedings of 5th Corpus Linguistic Conference* (Liverpool 2009), 1-24.

Sorianello P. (2017). *Il linguaggio disturbato. Modelli, strumenti, dati empirici*. Roma: Aracne.

Vernero I. & Romano A. (2017). La trascrizione del parlato patologico. In: L. Romito & M. Frontera (a cura di), *La scrittura all'ombra della parola* (no. monografico dei Quaderni di Linguistica dell'Università della Calabria, VII (5)), 11-31.

Yaruss J.S. & Quesal R.W. (2004). Stuttering and the International Classification of Functioning, Disability, and Health: an update. *J. of Commun. Disorders*, 37(1), 35-52.

van Zaalen Y. & Reichel I. (2015). *Cluttering: Current views on its nature, assessment and treatment*. wC: iUniverse.

van Zaalen Y. Strangis D. (2022). An Adolescent Confronted With Cluttering: The Story of Johan. *Perspectives of the ASHA Special Interest Groups*, 1-13.

Zmarich C. (1991). Una revisione critica degli studi 'linguistici' sulla balbuzie.*Acta Phoniatrica Latina*, XIII, 4, 495-514.

Zmarich C., Magno Caldognetto E.& Ferrero F. (1997). Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza. In F. Cutugno (a c. di), *Fonetica e Fonologia degli stili dell'italiano parlato* (Atti delle XXIV Giornate di Studio del Gruppo di Fonetica Sperimentale dell'A.I.A., Napoli, 14-15 novembre 1996), Roma: Tipografia Esagrafica, 111-139.

Leveling mechanisms in a university environment: the Gorgia Toscana in Florentine subjects studying in Bologna

Giuditta Avano[a], Cinzia Avesani[b], Mario Vayra[b,c]
[a]Università di Pisa; [b] ISTC-CNR; [c]Università di Bologna

Introduzione. Nel mondo anglosassone sono stati fatti vari studi sull'accomodamento [9] tra varietà della stessa lingua (Second dialect acquisition studies, raccolti da [17]) e in particolare sulla convergenza degli accenti in contesti universitari ([16]; [7]; [1]), contesti in cui, più che entrare in contatto con la varietà della città di arrivo o con una lingua standard, ad entrare in contatto sono più varietà della stessa lingua. Nel contesto italiano, in cui non c'è una varietà standard, una pronuncia senza tratti regionali è estremamente rara anche in situazioni molto formali e tra i parlanti con più alto livello di educazione [4], in quanto "Pronunciation is less subject to the pressure of the standard norm than other levels of the language system" (ivi :20). Anche nelle università, dunque, i parlanti si esprimono con una pronuncia in cui sono presenti tratti fonetici della propria varietà di origine. Il contatto con altre varietà ha però delle conseguenze sull'italiano regionale degli studenti? Si osservano fenomeni di accomodamento che portano ad un livellamento [10], ovvero a una riduzione dei tratti regionalmente più marcati o avvertiti come non prestigiosi? Il diverso prestigio di cui le varietà d'italiano godono ha un ruolo nel determinare se vi sarà un livellamento verso una pronuncia regionalmente meno marcata?

Ad oggi, l'unico studio italiano pubblicato sull'accomodamento in contesto universitario è quello di Nese e Meluzzi [14] in cui si osserva e confronta l'accomodamento delle affricate dentali verso una realizzazione settentrionale da parte di studenti fuorisede che soggiornano da sei mesi a Pavia, e nella comunità italofona di Bolzano. Scopo dello studio era verificare se processi migratori diversi per tipo, durata e prospettiva portino o meno a "un identico processo di accomodamento linguistico" (ivi, 68). Poiché a Pavia gli studenti di provenienza meridionale avevano parzialmente aumentato le realizzazioni di tipo settentrionale delle affricate ed erano i soggetti femminili a guidare il mutamento, così come a Bolzano, gli autori concludono che si sia di fronte a un processo di accomodamento simile che tende "verso una varietà sentita come maggiormente prestigiosa" (ivi:80). Occorre però notare che l'accomodamento nei due contesti migratori è stato analizzato con due modalità differenti (diacronia reale a Pavia vs diacronia apparente a Bolzano), riguarda contesti diversi sia di partenza (vari) che di arrivo (Pavia e Bolzano), e usa corpora diversi.

Scopo. Il nostro studio vuole, in primo luogo, investigare se le dinamiche di contatto in ambito universitario portino a un livellamento dei tratti regionali dopo l'intero ciclo di studi (ovvero dopo un intervallo più lungo rispetto alla ricerca pavese in [14]) tramite l'analisi delle realizzazioni delle occlusive sorde intervocaliche in studenti fiorentini residenti a Bologna da 5 anni. Lo studio si concentra sulle occlusive sorde intervocaliche perché, nella varietà di italiano regionale toscana, sono il principale target del fenomeno di spirantizzazione noto come Gorgia Toscana (GT) [8], tratto di cui i fiorentini sembrano essere consapevoli, ma solo per il luogo di articolazione velare (stereotipo di "toscanità", [5]). La GT è stata ritenuta interessante per suo il prestigio controverso: alto nella comunità di origine [3], più basso fuori regione [12]. Per comprendere se negli studenti universitari fuorisede vi sia stato un accomodamento che ha portato a un livellamento della GT, abbiamo valutato se (a) dopo cinque anni di residenza a Bologna, i soggetti presentino percentuali minori di realizzazioni spirantizzate delle occlusive rispetto ai soggetti del gruppo di controllo. Ci siamo poi chiesti (b) se il genere possa avere un effetto sull'accomodamento, come in [14]; (c) se influisca sulla GT in modo diverso a Firenze e a Bologna; (d) se l'eventuale livellamento è lo stesso per tutte e tre le occlusive coinvolte nel fenomeno della GT e infine (e) se le reti sociali degli studenti e l'"identità di luogo", (la parte di identità che si definisce in relazione ad un luogo a cui il soggetto attribuisce un significato sociale) abbiano un ruolo nel determinare il grado di livellamento del tratto ([13],[15],[19]). In secondo luogo, abbiamo confrontato i risultati con quelli di un nostro precedente studio sull'accomodamento della GT in lavoratori residenti a Bologna da 20 anni (in revisione). Ciò che si vuole indagare è se,

come in [14], tipi di migrazione diversi - lavorativa, a lungo termine, che porta al contatto con una varietà predominante di italiano vs a breve termine, in un contesto in cui non prevale alcuna varietà - portino a un identico processo di accomodamento, ovvero se emerga una stessa influenza dei fattori sociolinguistici e linguistici sul fenomeno. Rispetto a [14], il presente studio presenta il vantaggio di mettere a confronto l'accomodamento nei due contesti migratori, prendendo in considerazione lo stesso contesto di provenienza (Firenze) e di arrivo (Bologna), lo stesso numero di parlanti, utilizzando lo stesso metodo e corpus di analisi.

Metodo. Per il presente studio, sono stati coinvolti quattro soggetti fiorentini "fuorisede" (bilanciati per genere, età 24-28, di estrazione sociale simile) che hanno svolto l'intero percorso universitario (5 anni) a Bologna (città con più della metà degli studenti fuorisede, contesto ideale per osservare la realtà universitaria) e quattro studenti fiorentini rimasti nella città di origine come "controllo" (bilanciati per genere, età: 20-24, anch'essi di simile estrazione sociale). Gli 8 soggetti sono stati registrati durante un compito di denominazione di immagini e lettura di frasi. Per indagare l'influenza dell'identità di luogo, a ogni soggetto è stato chiesto di indicare quanto (da 1 a 5) si identificasse nella città di origine (Firenze) e nella città di arrivo (Bologna). Per indagare le reti sociali [13], ad ogni soggetto è stato chiesto inoltre di indicare la provenienza dei dieci contatti più stretti. Il corpus per l'esperimento è stato costruito scegliendo, dal VdB [6], 60 sostantivi target (tutte parole piane) contenenti occlusive sorde /p/, /t/, /k/, in contesto intervocalico, equamente distribuite in sillaba atona finale e in sillaba tonica prefinale. Ognuno di essi è stato pronunciato quattro volte, una volta in isolamento, elicitato tramite immagine, e tre volte letto in contesto frasale. In questa condizione, la parola è posta in una diversa posizione all'interno dei costituenti prosodici: posizione finale di sintagma fonologico, posizione finale di sintagma intonativo e posizione finale di enunciato. Il corpus ottenuto consta di 1920 occorrenze. Adattando i criteri di classificazione di [11] e [18], sono state individuate 6 realizzazioni allofoniche, da considerarsi in una scala crescente di indebolimento: O*cclusive>lenite>semifricative>fricative>fricative sonore>approssimanti*. Ogni occorrenza è stata classificata come uno degli allofoni sopra indicati, tramite analisi acustica e spettrografica con PRAAT [2]. Per misurare l'influenza dei fattori sociolinguistici e linguistici sulla GT, i dati sono stati analizzati utilizzando un test di regressione logistica ordinale, in cui la variabile dipendente ordinale è la realizzazione allofonica delle occlusive sorde intervocaliche e gli effetti fissi sono il "gruppo" (*fuorisede* vs *controllo*), il genere (*Maschi* vs. *Femmine*), il luogo di articolazione (*velare* vs *bilabiale* vs *dentale*) e le loro interazioni. Osservando i risultati per singolo parlante e confrontandoli con le indicazioni fornite dai soggetti sulla propria appartenenza identitaria alle città di origine e di arrivo e sui contatti più stretti abbiamo verificato inoltre l'effetto dell'identità di luogo e delle reti sociali sul grado di indebolimento delle occlusive dovuto a GT.

Risultati. Gli studenti fuorisede presentano complessivamente un grado di spirantizzazione delle occlusive sorde intervocaliche minore rispetto ai soggetti del gruppo di controllo, per tutte e tre le occlusive considerate, risultato che suggerisce che il contesto universitario bolognese abbia portato a un livellamento del tratto. Il genere sembra solo in prima analisi influire sul fenomeno, con una maggior differenza nel grado di GT tra fuorisede e controllo tra i soggetti femminili rispetto ai soggetti maschili. Osservando però l'effetto del genere nei due contesti (c) si osserva che a Bologna studenti e studentesse presentano una distribuzione allofonica simile, mentre a Firenze sono i soggetti femminili a presentare maggior spirantizzazione. Tra i soggetti fuorisede, più che il genere, sembrano essere le reti sociali a spiegare meglio il maggiore o minore livellamento della GT: chi ha mantenuto più saldi i rapporti con soggetti della città di origine presenta un maggior grado di spirantizzazione delle occlusive. L'identità di luogo non sembra invece avere un effetto sul livellamento del tratto. Anche nello studio sui soggetti lavoratori residenti a Bologna da venti anni era emerso un minore grado di indebolimento per tutte e tre le occlusive considerate in chi viveva a Bologna e il genere era emerso come un fattore che influiva significativamente sull'accomodamento (maggior livellamento della GT tra i soggetti femminili) . Il genere però (contrariamente al presente studio) influiva sulla realizzazione della GT a Bologna (molta più GT tra gli uomini rispetto alle donne) e era ininfluente a Firenze. L'identità di luogo sembrava avere

un certo ruolo nello spiegare la variabilità intersoggettiva in quanto i soggetti che si identificavano più con Firenze che con la città di arrivo erano quelli che avevano maggiormente mantenuto la GT. I due tipi di migrazione portano dunque entrambi ad un accomodamento verso una pronuncia regionalmente meno marcata. Se i fattori linguistici sembrano influire sul fenomeno in maniera simile, i fattori sociolinguistici mostrano invece una influenza diversa.

Riferimenti bibliografici.

[1] BIGHAM, D. S. (2010). Mechanisms of Accommodation among Emerging Adults in University Setting. In *Journal of English Linguistics*, 38(3), 193-210

[2] BOERSMA, P., WEENINK, D. (2021). PRAAT: doing phonetics by computer. [software] Versione 6.1.40. https://www.praat.org/.

[3] CALAMAI, S. (2011). Per una storia della pronuncia degli italiani: opinioni e atteggiamenti intorno alla pronuncia fiorentina. In NESI, A., MORGANA, S., & MARASCHIO, N. (a cura di), *Storia della lingua italiana e storia dell'Italia unita*. Firenze: Franco Cesati, 175-184.

[4] CERRUTI, M. (2011). Regional varieties of Italian in the linguistic repertoire. In *Int'l. J. Soc. Lang,* 210, 9-28.

[5] CRAVENS, T. (2000). Sociolinguistic subversion of a phonological hierarchy. In *Word,* 51, 1-19.

[6] DE MAURO, T. (2016). Nuovo Vocabolario di Base della lingua italiana. In *Internazionale*.

[7] EVANS, B.G., IVERSON, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. In *Journal of the Acoustical Society of America*, 121, 3814-3826.

[8] GIANNELLI, L., SAVOIA L. M. (1978). L'indebolimento consonantico in Toscana, I. In *Rivista Italiana di Dialettologia,* 2, 25-58.

[9] GILES, H. (1973). Accent mobility: a model and some data. In *Anthropological Linguistics,* 15, 87-105.

10] KERSWILL, P. (2013). Koineization. In CHAMBERS, J.K., SHILLING, N. (a cura di), *The Handbook of Language Variation and Change*. Wiley, 519-536

[11] MAROTTA, G. (2001). Non solo spiranti: La Gorgia Toscana nel parlato di Pisa. In *L'Italia Dialettale,* 62, 27-60.

[12] MAROTTA, G. (2014). New parameters for the sociophonetic indexes. Evidence from the Tuscan varieties of Italian. In CELATA, C., CALAMAI, S. (a cura di), *Advances in Sociophonetics*. Amsterdam-Philadelphia, John Benjamins, 137-168

[13] MILROY, L (1987). *Language and Social Networks.* Oxford: Basil Blackwell.

[14] NESE, N., MELUZZI C., (2017). Accomodamento ed emergenza di varianti fonetiche: le affricate dentali intermedie a Pavia e Bolzano. In BERTINI, C., CELATA, C., LENOCI, G., MELUZZI, C. & RICCI, I (a cura di), *Fattori sociali e biologici nella variazione fonetica. Studi AISV 3*, 67-82.

[15] NYCZ, J. (2018). Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity. In *Language variation and change,* 30, 175-202.

16] PARDO, J. S. (2006). On phonetic convergence during conversational interaction. In *Journal of the Acoustical Society of America,* 85, 2088-2392

[17] SIEGEL, J. (2010). *Second Dialect Acquisition*. Cambridge, Cambridge University Press.

[18] SORIANELLO, P. (2001). Un'analisi acustica della Gorgia fiorentina. In *L'Italia Dialettale,* 62, 61-94

[19] VIETTI, A. (2017). Italian in Bozen/Bolzano: The formation of a "new dialect". In CERRUTI, M., CROCCO, C, MARZO S. (a cura di), *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian.* Berlin: De Gruyter, 176-212.

Understanding the role of prosody and timbre in voicediscrimination using synthetic stimuli

Elisa Pellegrino[1], Debora Beuret[1], Thayabaran Kathiresan[1], Claudia Roswandowits[1], Sascha Fruholz[2], Volker Dellwo[1]
[1]*Department of Computational Linguistics, University of Zurich, Switzerland*
[2]*Department of Psychology, University of Oslo,*
Norway

While decades of research revealed the importance of spectral segmental information in voice recognition (cf. Schweinberger et al. 2014 for a review), the role of prosody is yet under debate. In most common automatic voice recognition systems, prosodic information plays at best a subordinate role (Adami et al. 2003). The role of prosody in voice recognition needs, instead, further investigations since speaker-specific prosodic information can be the key to distinguish natural voices from their synthesized counterparts. The purpose of this study is thus to examine the role of prosody in relation to timbre in human voice discrimination.

To produce the stimuli for the discrimination task, voice conversion technique was applied by which the timbre (short-term feature representation, such as instantaneous F0 and spectrum, Wu and Lee, 2014) was transferred from one speaker (source) to four different speakers (target), while the prosodic characteristics (rate, rhythm, intonation) of the source speaker were maintained (see a.o. Kobayashi and Toda, 2018).This technique allowed to test the separate contribution of timbre and prosody. Source and target speakers were male, aged between 22 and 45 years old, and native speakers of Standard German. In the experiment, the voices in a trial were combined under three conditions in which:

a) the timbre was maintained between speakers, but prosody varied (henceforth condition‘same timbre’),
b) prosody varied but timbre was maintained (henceforth condition ‘same prosody’),
c) both timbre and prosody varied between speakers (henceforth condition “none”).

In view of the known effects of sentence length on voice recognition (see a.o. Cook and Wilding, 1997), long (5-word) and short (2-word) declarative utterances were used that contained different degree of prosodic information. The voice discrimination experiment was administered to 40 Swiss German listeners, aged between 18 and 35 years old. With such an experimental design, the following hypotheses regarding the role of timbre and prosody in relation to sentence length on voice discrimination (measured in % same) were tested. We expected that the % same for stimuli in condition «same prosody» would be (a) higher in longer than in short utterances, (b) lower than stimuli in condition «same timbre» for short utterances. In the absence of an interaction between features combination and sentence length, we also expected that (a) the %same in condition ‘same prosody’ would be higher than in condition “none” but lower than in condition “same timbre”; (b) the % same in longer utterances would be higher than in short utterances.

The significance of the interaction and main effect of feature combination (same timbre/same prosody/none) and sentence length (short/long utterance) on %same (dependent variable) were tested with a two-way repeated measures ANOVA. The overall results of the study showed that (a) longer utterances allowed listeners to make more informed decision about the speaker identity; (b) in both long and short utterances listeners used prosody and timbre equally to decide whether two voices stemmed from the same or different speakers.

Beyond informing about the centrality of prosodic feature for voice discrimination, these results also suggest that when voices are manipulated such that the identity of one speaker is

converted into that of another speaker, prosodic differences between source and targetspeakers can be used to distinguish natural voices from their synthetic copies.

References

Adami, G., Mihaescu, R. Reynolds, D. A. and Godfrey, J. J. (2003). Modeling prosodicdynamics for speaker recognition, *Proceedings ICASSP '03,* IV-788.

Schweinberger S R, Kawahara H, Simpson AP, Skuk V G, Zäske R. (2014). Speakerperception. Wiley Interdiscip Rev Cogn Sci. Jan;**5(1):**15-25.

Cook, S., and Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear thelength," Appl. Cogn. Psychol. 11(2), 95–111.

Kobayashi, K. and T. Toda (2018). Sprocket: Open-Source Voice Conversion Software. *Proc. Odyssey*, 203-210, June 2018.

Wu and Lee, 2014, Voice conversion versus speaker verification: an overview, APSIPATransactions on Signal and Information Processing, Volume 3.

Prestigious features and over-distancing from the local variety in the Italian of Neapolitan radio speakers

Giovanni Abete
Università degli Studi di Napoli 'Federico II'

Il presente contributo prende in esame la variabilità fonetica in un corpus di parlato radiofonico, con particolare attenzione alle scelte stilistiche effettuate dagli speaker in rapporto a una serie di variabili situazionali (cfr. Coupland 2001). L'analisi si concentra su due storiche radio napoletane, Radio Marte e Radio Kiss Kiss Napoli, le quali, pur essendo fortemente radicate nella realtà locale, hanno un bacino d'utenza piuttosto ampio che abbraccia l'intera regione e che, grazie alle potenzialità di internet e della tv digitale, si spinge anche oltre i confini regionali. Gli speaker analizzati si esprimono prevalentemente in italiano e tendono a rimuovere le caratteristiche fonetiche più marcate in senso locale e di minor prestigio; tuttavia, in alcune circostanze e per alcuni scopi comunicativi selezionano anche registri più bassi (cfr. Atzori 2016; Maraschio 2011). In questo contributo ci si concentrerà sulle pronunce di maggior prestigio e sulla tendenza, particolarmente avanzata in alcuni speaker, all'iperdistanziamento dalla varietà locale (Berruto 1993: 59). La ricerca consentirà di riflettere sulle dinamiche in atto nei processi di livellamento degli italiani regionali e nella formazione di nuovi modelli di prestigio (cfr. Cerruti, Crocco, Marzo 2017).

Il corpus consta di circa venti ore di parlato provenienti da sei programmi radiofonici delle due emittenti sopra citate. Gli speaker analizzati sono anch'essi sei e costituiscono i conduttori dei rispettivi programmi. Sono tutti nati e cresciuti a Napoli. Le trasmissioni trattano di calcio e sono incentrate sulla squadra di calcio del Napoli. In riferimento agli interventi dei conduttori radiofonici, emergono diverse tipologie di parlato: monologico (nelle parti più informative, spesso a inizio trasmissione); dialogico con ascoltatori da casa (il cosiddetto "filo-diretto"); dialogico con esperti in presenza o a distanza; parlato letto (il conduttore legge messaggi WhatsApp inviati dagli ascoltatori).

L'analisi si concentra su 10 varianti fonetiche che consentono di valutare il grado di distanziamento degli speaker dalle pronunce locali e l'adesione a modelli sovraregionali:

1. Dittonghi [jɛ] e [wɔ] (es. [ˈpjɛde], [ˈwɔmo] vs [ˈpjede], [ˈwomo])
2. Avverbi in -m[e]nte (es. [indubˈbjamente] vs [indubˈbjamɛnte])
3. Pronuncia scempia di /b/ e /dʒ/ intervocaliche (es. [staˈdʒone] vs [stadˈdʒone])
4. – Raddoppiamento Fonosintattico (es. [a ˈnapoli] vs [a nˈnapoli])
5. Mantenimento di /tʃ/ intervocalica (es. [pjaˈtʃere] vs [pjaˈʃere])
6. Sonorizzazione di /s/ intervocalica (es. [ˈrɔza] vs [ˈrɔsa])
7. Resa sorda delle occlusive sorde postnasali (es. [ˈaŋke] 'anche' vs [ˈaŋge])
8. Resa alveolare di /s/ davanti a C[– coronale] (es. [skuˈdɛtto] vs [ʃkuˈdɛtto])
9. Mancata affricazione di /s/ postnasale (es. [ˈpɛnso] vs [ˈpɛntso])
10. Mantenimento di /r/ davanti a consonante (es. [ˈfɔrtsa] vs [ˈfɔttsa])

Come si vede dagli esempi, ogni variante si contrappone a una variante locale caratteristica dell'italiano regionale campano (cfr. Maturi 2006: 88-105, De Blasi 2014: 103-104). Ogni coppia di varianti individua quindi una variabile sociolinguistica (ulteriori indicazioni sul trattamento binario delle variabili saranno fornite nella versione estesa del lavoro). Per la presente ricerca, sono stati selezionati circa 50 minuti di parlato per ciascuno speaker, all'interno dei quali tutte le occorrenze delle variabili in esame sono state segmentate, etichettate, e trascritte foneticamente. L'analisi è stata condotta con un metodo essenzialmente uditivo e impressionistico, integrato dall'ispezione visiva delle caratteristiche acustiche del segnale, in particolare della forma dell'onda e dello spettrogramma (cfr. Abete 2021ab). Tale scelta è funzionale a un'analisi qualitativa e globale del parlato e consente di prendere in esame un'ampia gamma di fenomeni fonetici in relazione alle dinamiche interazionali che si sviluppano durante la trasmissione radiofonica.

Questa metodologia consente di delineare i profili sociofonetici degli speaker e di valutare il grado di distanziamento dalla varietà locale. A tale proposito, l'analisi dimostra come gli speaker napoletani si orientino verso pronunce di prestigio in gran parte deregionalizzate (cfr. Vietti 2019: 4). I tratti più

marcatamente locali sono evitati, in particolare la palatalizzazione di /s/ prima di consonante non coronale. La sonorizzazione di /s/ intervocalica si presenta invece molto avanzata e in diversi speaker è addirittura sistematica. Come è noto, si tratta di un tratto di prestigio in espansione da nord (Vietti 2019: 13; Giordano, Savy 2012; De Blasi, Fanciullo 2002: 644 segnalavo già il fenomeno nell'italiano di alcune zone residenziali di Napoli). Altre caratteristiche meridionali, come la resa doppia di /b/ e /dʒ/ intervocaliche, sono evitate nei contesti più formali ed emergono solo nei momenti di minore controllo. Gli speaker che vi prestano più attenzione presentano anche casi di mancato raddoppiamento fonosintattico e qualche scempiamento di consonanti etimologicamente lunghe (cfr. Valentini, Calamai 2021: 397), interpretabili come forme di iperdistanziamento dalla varietà locale.

Dopo aver presentato le caratteristiche del campione e i risultati generali della ricerca, si proporrà un'analisi qualitativa di alcuni scambi conversazionali tra conduttore radiofonico e ascoltatori. L'analisi qualitativa delle dinamiche interazionali che si realizzano alla radio è infatti cruciale per la comprensione dei meccanismi che regolano la selezione delle varianti. La variazione fonetica che ogni speaker ha a disposizione viene impiegata come una risorsa per raggiungere determinati scopi comunicativi: in particolare, nel corpus analizzato, la selezione di varianti di prestigio e l'iperdistanziamento dalla varietà locale vengono frequentemente utilizzati come strategie per aumentare la distanza con gli ascoltatori – che presentano invece caratteristiche fonetiche piuttosto marcate in senso locale – e per evidenziare l'obiettività e la professionalità del conduttore (cfr. gli studi condotti nell'ottica dello *speaker design* discussi in Schilling-Estes 2002: 388 ss.). I risultati della presente ricerca offrono quindi diversi spunti di riflessione sulle tensioni tra modelli locali e sovralocali che condizionano le scelte fonetiche dei parlanti, e sulle dinamiche in atto nella formazione di nuovi modelli di prestigio.

Riferimenti bibliografici

Abete, G. (2021a). Trascrizione impressionistica e analisi della variazione fonetica: aspetti teorici e metodologici. In: Bertin, A. et al. (eds.), *Réflexions théoriques et méthodologiques autour de données variationnelles*. Actes du colloque international DIA V, Nanterre, 5-7 septembre 2018. Chambéry: PUSMB. 85-97.

Abete, G. (2021b). Profili sociofonetici di alcuni parlanti di origine campana. In: Baggio, S. (ed.), *Chi siamo, come parliamo. Inchiesta sociolinguistica nel Dipartimento di Lettere e Filosofia dell'Università di Trento*. Trento: Università degli Studi di Trento. 87-117.

Atzori, E. (2016). La lingua della radio. In: Bonomi, I., Morgana, S. (eds.), *La lingua italiana e i mass media*. Roma: Carocci. 41-79.

Berruto, G. (1993). Varietà diamesiche, diastratiche, diafasiche. In: Sobrero, A.A. (ed.), *Introduzione all'italiano contemporaneo*. Vol. II, *La variazione e gli usi*. Roma-Bari: Laterza. 37-92.

Cerruti, M., Crocco, C., Marzo, S. (eds.) (2017). *Towards a new standard. Theoretical and empirical studies on the restandardization of Italian*. Berlin: De Gruyter.

Coupland, N. (2001). Language, situation, and the relational self: theorizing dialect-style in sociolinguistics. In: Eckert, P., Rickford, J. (eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. 185-210.

De Blasi, N. (2014). *Geografia e storia dell'italiano regionale*. Bologna: il Mulino.

De Blasi, N., Fanciullo, F. (2002). La Campania. In: Cortelazzo, M. et al. (eds.), *I dialetti italiani. Storia, struttura, uso*. Torino: UTET. 628-678.

Giordano, R., Savy, R. (2012). Sulla standardizzazione del consonantismo dell'italiano: consonanti geminate, rafforzate e fricative alveolari in contesto intervocalico. In: Bianchi, P. et al. (eds.), *La variazione nell'italiano e nella sua storia*. Atti dell'XI Congresso SILFI, Napoli, 5-7 ottobre 2010. Firenze: Cesati. 431-445.

Maraschio, N. (2011). Radio e lingua. In: Simone, R., Berruto, G., D'Achille, P. (eds.), *Enciclopedia dell'Italiano*. Roma: Treccani.

Maturi, P. (2006). *I suoni delle lingue, i suoni dell'italiano. Introduzione alla fonetica*. Bologna: il Mulino.

Schilling-Estes, N. (2002). Investigating stylistic variation. In: Chambers, J.K., Trudgill, P., Schilling-Estes, N. (eds.), *The handbook of language variation and change*. Malden (MA): Blackwell. 375-401.

Valentini, C., Calamai, S. (2021). Sull'insegnamento della pronuncia italiana negli anni Sessanta a bambini e a stranieri. In: Bernardasci et al. (eds.), *l'individualità del parlante nelle scienze fonetiche: applicazioni tecnologiche e forensi*. Studi AISV 8. 391-402.

Vietti, A. (2019). Phonological variation and change in Italian. In: *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. 1-28.

On the Role of Dialogue Context in Predicting Speaking Style

Vincenzo Scotti and Roberto Tedesco
DEIB, Politecnico di Milano – Via Golgi 42, 20133, Milano (MI), Italy

Text to Speech (TTS) synthesis is a problem almost as old as Natural Language Processing (NLP). The focus of this problem is on creating tools capable of generating a voice uttering a given text. These tools have many applications, from assistive technologies to chatbots and virtual assistants. Older solutions to build TTS tools relied on concatenative synthesis (either diphone or unit selection) (Jurafsky and Martin, 2009). Nowadays, deep learning powered solutions have shown impressive generative capabilities are becoming the standard technology to build TTS tools (Tan et al., 2021): models like Tacotron (Skerry-Ryan et al., 2018), DeepVoice (Ping et al., 2018), or FastSpeech (Ren et al., 2019) can generate incredibly natural voices. The naturalness of the generated speech is also helped by the use of neural vocoders that convert spectrograms into fluent and natural speech waveforms, replacing to the Griffin-Limm algorithm (Zhu et al., 2007).

Most recent deep learning models try to go beyond mere speech generation from the text. As a result, newer models try to factorise the probability predicted by these generative models to the condition it on various aspects: speaker's voice, speaking style, or prosody (Tan et al., 2021). While in some cases, as speaker conditioning, it is possible to build separate and re-usable models to extract the latent representation that encoders the desired information (Jia et al., 2018), in others, like speaking style prediction, the encoder of the latent representation is an integral part of the TTS model (Wang et al., 2018), yielding to a strong coupling between style encoder and TTS that makes the sub-modules hardly re-usable.

In this work, we focused on speaking style conditioning and detaching the style prediction from the actual speech synthesis, to promote the re-usability and modularity of these models. State-of-the-art TTS models use an unsupervised approach called Global Style Tokens (GSTs) that extracts the style representation from a given reference audio (Wang et al., 2018). This representation is computed internally by the TTS as a weighted sum of some learnt latent vectors. Changing the weights in the combination, the resulting style vector used to condition the spectrogram generation changes and as a result the speaking style changes (usually, these vector affects aspects like speech rate or intensity). These weights are computed from a reference audio clip (the one we want to emulate the speaking style); consequently, the resulting model is bound by the need for a reference database of speaking styles to choose from and a way to retrieve the reference style audio from the database, introducing the need of a human in the loop selecting manually the reference audio clip for the style. Alternatively, it would be necessary to have a separate module predicting the speaking style latent vector from the text the TTS needs to utter. This prediction problem is the object of this study.

Our work focuses on predicting the speaking style from the given text inside a conversation, for the application to conversational agents and chatbots. To this end, we developed and trained a neural network module working as a connection module between the textual component and the speech synthesis component of a chatbot. While some works have already addressed this problem more generically (Stanton et al., 2018), we are interested in understanding the role of the context of the conversation (i.e., the preceding turns in the dialogue) in the choice of the appropriate response speaking style.

We developed this connection module to help improve the human-likeness of chatbots and conversational agents. In fact, the choice of the appropriate tone and style while talking during a conversation is a consequence of the empathetic capabilities that characterise humans; thus, being able to control this aspect would improve the perception of the underlying agent. While this aspect does not seem to be crucial in assistive technologies applications, in other use cases like developing mental healthcare chatbots it is. The use case we propose for this technology is the development of counselling/psychotherapy chatbots, which require the simulation of empathetic traits to be perceived by the users as human thus favouring sympathy and openness from the user towards the agent.

To investigate the role of conversational context in the choice of the appropriate speaking style, we started from deep neural networks trained for dialogue language modelling –DialoGPT (Zhang et al., 2020) and Therapy-DLDLM (the latter is a custom dialogue model trained on open domain and therapy conversations)– and conditioned TTS model –namely Mellotron TTS (Valle et al., 2020), a variation of the Tacotron TTS–. The dialogue language models work as embedding models: we leverage their latent contextual representation of the dialogue as input to predict the speaking style.

To explore the possible solutions, in the experiments we compared textual models having different complexities. Moreover, to understand the role of context, we compared, for each model, different encoding approaches: (i) response embeddings only, (ii) contextualised response embeddings, (iii) combined context-response embeddings. The latter pre-trained model provides a reference encoder for the GSTs. In fact, the employed TTS model encapsulates a reference encoder to extract the latent prosody representation used to compute the style latent vector.

Besides the two pre-trained models for text analysis and spectrogram generation, we leveraged also a pre-trained vocoder model –namely WaveGlow (Prenger et al., 2019)– to convert the Mel spectrogram synthesised by the TTS into a raw waveform. This allowed us to assess qualitatively that the audio synthesised by the TTS was still intelligible even if generated from the predicted GSTs, rather than the GSTs of a reference audio clip.

Following similar works on GSTs prediction (Stanton et al., 2018), we addressed the problems in two ways: (i) predicting the combination weights to compose the style vector, (ii) predicting directly the raw embedding vector The difference in the two targets results in slightly different approaches to the problem. In fact, the former approach requires a model working as a classifier, yielding a probability distribution that corresponds to the combination weights; we used the Kullback–Leibler divergence (KL divergence) of the predicted distribution from the target one as loss function. The latter approach, instead, requires a model doing regression, in this case, we trained the model minimising the Euclidean norm of the difference between the predicted and the target style vector. Note that to train the model, independently from the approach, we need to have a spoken dialogue data set providing the transcriptions. From the audio clips in the training data set, we can extract the target weights distribution or the target style vector we want to learn to predict on new data.

Table 1: Results of the two approaches to reconstruct the GST (lower is better).

Model	No. of parameters	MSE			KL-Divergence		
		Response	Response from context	Context and response	Response	Response from context	Context and response
DialoGPT	117M	0.0473	0.0515	0.0559	0.1238	0.1377	0.1526
	345M	0.0451	0.0449	0.0478	0.1148	0.1055	0.1249
	762M	0.0492	0.0557	0.0599	0.1236	0.1450	0.1540
Therapy-DLDLM	762M	0.0427	**0.0399**	0.0441	0.1047	**0.0958**	0.1109

We conducted the experiment training and evaluating the style prediction module on the IEMOCAP corpus (Busso et al., 2008), which offers scripted and spontaneous dialogues in English displaying different emotions that result in a wide variety of speaking styles1. Given this highly variated corpus, the neural network module could learn how to choose the appropriate speaking style given the output text and the conversational context, simulating the desired empathetic behaviour we were looking for. We evaluate the considered models by computing the prediction error of the speaking style (using both weight prediction and embedding prediction approaches) on the test split of IEMOCAP. The resulting figures, reported in Table 1, indicate an important role of context and language model complexity in achieving good predictive capabilities. In fact, the model with the highest number of parameters using contextualised response embeddings achieves the best results with both approaches.

References

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. IEMOCAP: interactive emotional dyadic motion capture database.

Lang. Resour. Evaluation, 42(4):335–359, 2008.

Ye Jia, Yu Zhang, Ron J. Weiss, Quan Wang, Jonathan Shen, Fei Ren, Zhifeng Chen, Patrick Nguyen, Ruoming Pang, Ignacio Lopez-Moreno, and Yonghui Wu. Transfer learning from speaker verification to multispeaker text-to-speech synthesis. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolo` Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montre´al, Canada*, pages 4485–4495, 2018.

Dan Jurafsky and James H. Martin. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition*. Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education International, 2009.

Wei Ping, Kainan Peng, Andrew Gibiansky, Sercan O¨ mer Arik, Ajay Kannan, Sharan Narang, Jonathan Raiman, and John Miller. Deep voice 3: Scaling text-to-speech with convolutional sequence learning. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

Ryan Prenger, Rafael Valle, and Bryan Catanzaro. Waveglow: A flow-based generative network for speech synthesis. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, United Kingdom, May 12-17, 2019*, pages 3617–3621. IEEE, 2019.

Yi Ren, Yangjun Ruan, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, and Tie-Yan Liu. Fastspeech: Fast, robust and controllable text to speech. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alche´-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 3165–3174, 2019.

R. J. Skerry-Ryan, Eric Battenberg, Ying Xiao, Yuxuan Wang, Daisy Stanton, Joel Shor, Ron J. Weiss, Rob Clark, and Rif A. Saurous. Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsma¨ ssan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 4700–4709. PMLR, 2018.

Daisy Stanton, Yuxuan Wang, and R. J. Skerry-Ryan. Predicting expressive speaking style from text in end-to-end speech synthesis. In *2018 IEEE Spoken Language Technology Workshop, SLT 2018, Athens, Greece, December 18-21, 2018*, pages 595–602. IEEE, 2018.

Xu Tan, Tao Qin, Frank K. Soong, and Tie-Yan Liu. A survey on neural speech synthesis. *CoRR*, abs/2106.15561, 2021.

Rafael Valle, Jason Li, Ryan Prenger, and Bryan Catanzaro. Mellotron: Multispeaker expressive voice synthesis by conditioning on rhythm, pitch and global style tokens. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2020, Barcelona, Spain, May 4-8, 2020*, pages 6189–6193. IEEE, 2020.

Yuxuan Wang, Daisy Stanton, Yu Zhang, R. J. Skerry-Ryan, Eric Battenberg, Joel Shor, Ying Xiao, Ye Jia, Fei Ren, and Rif A. Saurous. Style tokens: Unsupervised style modeling, control and transfer in end-to-end speech synthesis. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsma¨ssan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 5167–5176. PMLR, 2018.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. DIALOGPT : Large-scale generative pre-training for conversational response generation. In Asli Celikyilmaz and Tsung-Hsien Wen, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, ACL 2020, Online, July 5-10, 2020*, pages 270–278. Association for Computational Linguistics, 2020.

Xinglei Zhu, Gerald Beauregard, and Lonce L. Wyse. Real-time signal estimation from modified short-time fourier transform magnitude spectra. *IEEE Trans. Speech Audio Process.*, 15(5):1645–1653, 2007.

[1] Links to the source code: https://github.com/vincenzo-scotti/dialoguegst , Mellotron TTS API: https//github.com/vincenzo-scotti/tts_mellotron_api; , Thearapy-DLDLM: https://github.com/vincenzo-scotti/dldlm

IL PARLATO IN AMBITO MEDICO:
Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD:
Linguistic analysis, technological applications and clinical tools

THURSDAY 16 FEBRUARY
POSTER SESSION II

A bimodal methodology for the analysis of rhythmic priming in speech performance

Leonardo Contreras-Roa[1], Paolo Mairano[2], Caroline Moreau[2], Anahita Basirat[2]
[1]Université de Picardie Jules Verne, [2]Université de Lille

Introduction. In this contribution, we propose a data treatment methodology for the study of the influence of non-linguistic rhythmic priming on the speech produced by French-speaking patients with Parkinson's disease (PwPD). PwPD and control participants were recorded performing a sentence read-aloud task under three conditions: one after listening to a regular rhythmic prime coinciding with the accentual structure of the sentences, one after listening to an irregular rhythmic prime incongruous with the accentual structure of the sentences, and one after listening to two seconds of silence, with the aim of testing the hypothesis that regular rhythmic priming favors a more natural speech rhythm performance for patients. The methodology is devised by combining analyses of speech rhythm metrics traditionally used in phonetic cross-linguistic analysis (Arvaniti, 2012; Mairano & Romano, 2011; Low, Grabe & Nolan, 2000; Ramus et al., 1999) and increasingly used in the study of dysarthric patients (Lowit et al., 2018; Liss et al. 2009), along with automatic gradual prominence detection methods –as opposed to strictly categorical detection methods– based on acoustic prosodic parameters (Goldman & Simon, 2020). In this sense, our approach is bimodal, as it considers two different sets of parameters that are thought to contribute to the perception of speech rhythm (Lowit et al., 2018). The objective of this contribution is to document the data treatment protocol and to reflect upon its advantages and potential uses in other areas of speech analysis.

Methodology. The data treatment workflow illustrated in Figure 1 encompasses 4 main stages: semi-automatic speech annotation and segmentation, prominence detection, rhythm metrics calculation, and overall data extraction. Semi-automatic speech annotation and subsequent syllabification are carried out with two services provided by the WebMAUS platform: WebMAUS Basic (Kisler et al., 2017) and Pho2Syll (Reichel & Kisler, 2014). The resulting annotated Praat TextGrid is manually corrected to address potential alignment errors. Audio and annotation data are then fed to the ProsoProm tool, part of the ProsoBox3 package (Goldman & Simon, 2020), a Praat plugin which allows to compute prominence as a gradual feature by analyzing silent pauses along with the relative duration and pitch averages within an adjustable window of analysis: 2 syllables before and 1 syllable after, in this case. We adjusted the duration threshold for silent pauses to 250 ms to better account for pathological speech. ProsoProm outputs four extra textgrid tiers: one signaling syllables that are considered categorically prominent following the given threshold, and three tiers specifying the cumulative contribution of relative duration, relative pitch and intra-syllable pitch movement to the gradual prominence of every syllable, whether they are classified as prominent or not. Subsequently, rhythm metrics reported in Lowit et al. (2018) are calculated individually per speaker via a Praat script using textgrids from the previous step: %V, ΔC, ΔV, varcoC, varcoV, rVPV-C, nPVI-V, varcoVC, rPVI-VC, nPVI- VC. Finally, data is extracted onto two different datasets, one per type of analysis. ProsoProm prominence results detail syllable-per-syllable data, by speaker, by priming condition. Rhythm metrics results contain global values per speaker, by priming condition.

Discussion and perspectives. We propose that combining both methods provides a multimodal view of rhythmic and prosodic changes in the speech of PwPD. On the one hand, gradual prominence evaluation allows for a detailed quantitative study of the contribution of different acoustic correlates of stress. On the other hand, rhythm metrics have already shown to be effective for the study of speech pathologies like dysarthria (e.g. Lowit et al., 2018; Maffia et al., 2021). Combined results can be used to create more robust statistical models to explore the interaction, and potential correlation between rhythmic and prominence data. In addition, this methodology allows for examining the impact of the metrical structure of a prime on speech production. A pilot case study (Contreras-Roa et al., 2022) showed substantial differences in F0 and duration values between one control subject and one PwPD that point towards there being intonational differences in the underlying prosodic structure of sentences produced by both groups, as shown in figure 1. Rhythm metrics displayed slightly different behaviors by priming condition and by group as well. The

results are promising and will hopefully be confirmed and expanded on data that we are currently collecting and analyzing. If proven effective, our workflow may be expanded to other fields, such as cross-linguistic typological studies, infant L1 acquisition, or L2/L3 acquisition in adults.

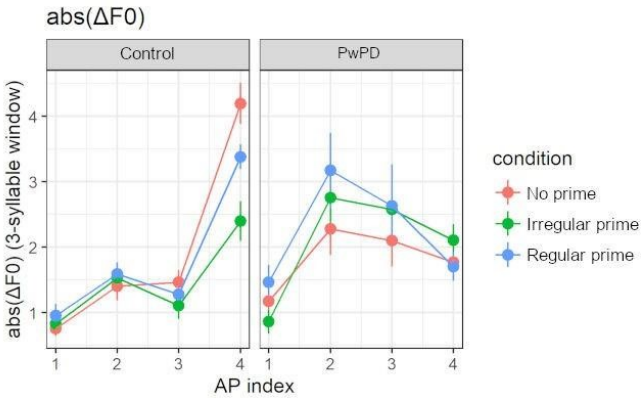


Figure 1. Absolute values of F0 movement within target prominent syllables, by speaker and priming condition (adapted from Contreras Roa et al., 2022). The F0 movements in the Control speaker's data reflect canonic pitch behaviors in prosodic boundaries in French.

References

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. Journal of Phonetics, 40(3), 351-373.

Contreras Roa, L., Mairano, P., Moreau, C. & Basirat, A. (2022). Impact de l'amorçage rythmique sur la production de la parole chez des personnes atteintes de la maladie de Parkinson : étude de cas. Actes des 34e Journées d'Études sur la Parole JEP 2022.

Goldman J.-P. & Simon-Hustinx A. C. (2020). Prosobox, a Praat plugin for analysing prosody. *Proc. of Speech Prosody*, 1009–1013.

Kisler, T. & Reichel U. D. and Schiel, F. (2017): Multilingual processing of speech via web services.

Computer Speech & Language, Volume 45, September 2017, pages 326–347.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., et al. (2009). Quantifyingspeech rhythm abnormalities in the dysarthrias. Journal of Speech, Language, and Hearing Research, 52, 1334–1352.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable- timing in Singapore English. Language and Speech, 43, 377–401.

Lowit, A., Marchetti, A., Corson S. & Kuschmann A. (2018). Rhythmic performance in hypokinetic dysarthria : Relationship between reading, spontaneous speech and diadochokinetic tasks. *Journal of Communication Disorders*, 72, 26–39.

Maffia, M., De Micco, R., Pettorino, M., Siciliano, M., Tessitore, A., & De Meo, A. (2021). Speech Rhythm Variation in Early-Stage Parkinson's Disease: A Study on Different Speaking Tasks. Frontiers in Psychology, 12, 2216.

Mairano P. & Romano, A. (2011). Rhythm metrics for 21 languages. Proceedings of the 17th International Congress of Phonetic Sciences, p. 1318–1321, Hong Kong.

Ramus F., Nespor M. & Mehler J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition, 73(3), 265–292.

Reichel, U.D., Kisler, T. (2014). Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In: Hoffmann, R. (Ed.): Elektronische Sprachverarbeitung. Studientexte zur Sprachkommunikation, 71, pp 42-49. TUDpress, Dresden.

Teaching speech sciences

Silvia Calamai[1], Chiara Celata[2]
[1]Università degli Studi di Siena, [2]Università di Urbino 'Carlo Bo',

Nella didattica universitaria italiana, l'insegnamento delle scienze del parlato e della voce compare spesso come un modulo all'interno dell'insegnamento di Linguistica generale; soltanto nei corsi di laurea magistrale può talvolta avere una maggiore autonomia. Si tratta, nella maggior parte di casi, di insegnamenti erogati all'interno di corsi di laurea umanistici, in cui risulta talvolta difficile introdurre concetti di fisica acustica e di anatomia: nella percezione degli studenti la materia è spesso giudicata ostica poiché richiede un armamentario proprio delle scienze dure. Viceversa, laddove lo studio della voce ha un risvolto più tecnologico – in particolare nei corsi di laurea di ingegneria o di scienze informatiche – diventa arduo approfondire concetti più squisitamente linguistici, quali ad esempio il mutamento fonetico. Di questo carattere intrinsecamente pluridisciplinare cercano in vario modo di rendere conto i vari manuali universitari, soprattutto di ambito anglosassone, che offrono una notevole ricchezza di prospettive e di approfondimenti (si vedano i bilanci 'europei' di Hazan e van Dommelen 1999 e di Mompeán et al 2011). Nel panorama italiano è d'uopo citare il fortunatissimo Albano Leoni, Maturi (2018[3]), che faceva seguito al il pionieristico Giannini, Pettorino (1992): su questi testi, cui possiamo aggiungere – da altre angolature disciplinari – il trattato curato da Croatto (1983) e il volume di Ferrero et al (1979) si sono formati generazioni di studiosi che hanno praticato la fonetica e la fonologia dentro i laboratori di ricerca italiani.

Tenendo conto di questo sfondo, riportiamo la nostra esperienza con un manuale introduttivo di scienze del parlato per studenti universitari italiani di lingue e linguistica, attualmente in fase di avanzata elaborazione [anonimo per revisione]. Il libro è concepito per offrire una guida allo studio linguistico dei suoni del parlato, dalla produzione all'acustica, all'udito e alla percezione, e con illustrazioni pratiche di procedure comuni per la registrazione, archiviazione (secondo i principi FAIR), annotazione e analisi acustica del parlato. Avendo una struttura modulare, con materiali integrativi per gli studenti più avanzati, il libro non trascura di presentare alcune nozioni dell'analisi fonologica (contrasto fonologico, neutralizzazione, restrizione fonotattica, peso sillabico ecc.). Un'ultima sezione è dedicata alle applicazioni della fonetica in vari ambiti professionali (disfunzioni del linguaggio, linguistica forense, insegnamento della pronuncia nelle lingue seconde, trattamento automatico del linguaggio).

Il libro assume il Laryngeal Articulator Model (LAM) come modello della produzione del parlato (Esling et al 2019). L'approccio LAM è emerso negli ultimi decenni come alternativa a un modello glottidale-linguale del tratto vocale. Il LAM offre un migliore potere esplicativo in quanto riconosce il fatto che la laringe è un articolatore complesso che consiste in una serie di strutture, implicate tanto nella fonazione quanto nell'articolazione, e rende conto delle interazioni tra le articolazioni del tratto vocale inferiore e quelle del tratto vocale orale.

L'assunzione del LAM come punto di partenza per una corretta descrizione delle dinamiche di produzione del parlato introduce alcune novità nell'insegnamento della fonetica e della fonologia, e sulle quali ci concentreremo in questo conributo. Ad esempio, una di questi ha a che fare con la caratterizzazione dei timbri vocalici. Lo 'spazio' vocalico quadrangolare basato sulla tradizionale intersezione tra due dinamiche linguali, una di antero-posteriorità e una di altezza-abbassamento (con il contributo eventuale dell'arrotondamento delle labbra), viene riconcettualizzata sulla base delle acquisizioni pià recenti della fonetica articolatoria come risultante da tre diverse dinamiche articolatorie (Esling 2005). Due di esse, l'avanzamento e l'innalzamento, sono dimensioni orali, controllate dai movimenti della lingua e della mandibola; la terza, che corrisponde alla ritrazione, è faringea nella misura in cui è controllata dallo sfintere ariepiglottale. Questo cambiamento di prospettiva si riflette anche sulla terminologia descrittiva da adottare. La nuova concettualizzazione è più realistica in termini articolatori e più intuitiva dal punto di vista didattico perché è direttamente ricollegabile alla contrazione dei tre muscoli principali che modificano la forma della lingua nella realizzazione dei tibri vocalici (Honda 1996).

Il LAM ha un impatto diretto sul modo in cui il tratto vocale stesso è rappresentato concettualmente.

Una prima deviazione dalle prospettive tradizionali è che la parte posteriore della lingua fa parte del tratto vocale inferiore (mentre il resto della lingua appartiene al tratto orale o vocale superiore). Vari fenomeni che implicano una costrizione della laringe lo dimostrano bene, come ad esempio la cosiddetta distinzione tra vocali [+ATR] e [-ATR]. Invece di concentrarsi solo sulla dinamica linguale, la spiegazione basata sul LAM si concentra sulla co-occorrenza dei movimenti della lingua con i movimenti verticali della laringe, dinamica che complessivamente porta all'espansione o alla restrizione della cavità faringea. La distinzione tra vocali tese e rilassate (tipicamente implementata in lingue come l'inglese) e quella tra vocali medio-alte e medio-basse (come tipicamente si indica per lingue come l'italiano) può, di conseguenza, essere spiegata all'interno di un quadro più ampio e molto più informativo. In tale quadro, un ruolo fondamentale è giocato dall'osservazione delle sinergie esistenti tra gli articolatori. Concentrarsi sulle sinergie piuttosto che sulla parcellizzazione del tratto vocale potrebbe essere erroneamente interpretato come una "complicazione" dal punto di vista pedagogico. Al contrario, è il modo più realistico per spiegare come funziona la produzione del parlato; e il realismo ha il vantaggio (tra vari altri) di elevare il livello di intuitività nella spiegazione di molti fenomeni linguistici.

Questi e altri aspetti innovativi del manuale verranno discussi per le loro implicazioni teoriche, didattiche e applicative e per il possibile impatto sulla pratica della fonetica e della fonologia in Italia.

Riferimenti bibliografici

Albano Leoni F., Maturi P. 2018. *Manuale di fonetica*, 3a edizione, Roma, Carocci. Croatto L. (a cura di) 1983. *Trattato di foniatria e logopedia*, Padova, La Garangola.

Esling J. 2005. There are no back vowels. The Laryngeal Articulator Model. Canadian Journal fo Lingusitics 50, 13-44.

Esling J. et al. (2019) *Voice quality. The Laryngeal Articulator Model*, Cambridge, CUP. Ferrero F. et al. 1979. *Nozioni di fonetica acustica*, Torino, Omega.

Giannini A., Pettorino M. 1992. *La fonetica sperimentale*, ESI, Napoli.

Hazan, V., van Dommelen, W. 1997. A survey of phonetics education in Europe. *EuroSpeech97*, 1939-42.

Honda, K. 1996. Organization of tongue articulation for vowels. *Journal of Phonetics* 24: 39–52. Mompeán, J.A., Ashby, M., Fraser H. 2011. Phonetics Teaching and Learning: An Overview of Recent Trends and Directions. *17th ICPhS*.

Reference, sound symbolism, and variation. In search of an elusive link

Duccio Piccardi
Università degli Studi di Siena/Università degli Studi di Pavia

Introduction and scope

Almost a century ago, Köhler [1] and Sapir [2] sparked the interest of psychologists and linguists on (apparently) non- arbitrary associations between sounds and referential features. Since then, research on sound symbolism has substantially grown [3]; in particular, we are currently witnessing a rekindled attention on this topic [4], and linguistics subfields are striving to find a place for sound symbolism in their frameworks (e.g., [5]). Third-wave variationist sociolinguistics considers sound symbolism as one of the potential source of indexical values attributed to variants in individual styles [6: 96-97]. By referring to Ohala's frequency code [7], studies suggested that frequency modulations may display, e.g., affect [8] and refinedness [9]. Indeed, Ohala's code pertains to both socio-indexical and referential meaning [7]. A biologically-determined indexical feature [10], i.e., body size signaled through frequencies inversely proportional to the dimensions of the vibrating membranes, gave birth to both socio-semiotic strategies (e.g., low F_0 evoking dominance of the utterer) and denotational sound symbolism (e.g., low vowel F_2 evoking big referents). Ohala himself deemed the latter «more problematic[ally]» explained through the tenets of his code, ultimately justifying this osmosis between meaning domains by referring to general mechanisms of hearer reaction to a communicated concept [7: 5-6]. Thus, the frequency code is deeply rooted in indexicality [11], and it could be assumed that, being the product of multiple socio-semiotic reinterpretations, most of the stylistic phenomena mentioned above do not necessarily imply a move from denotation to the social [8: 70]. The two-faced nature of Ohala's theory triggered diverse reactions in the linguistics community: while sociolinguists formalized a recent trend equating social meaning to meaning in general [12] through the assumption that the «boundary between the […] affective and the referential […] is fluid» [13: 30], psycholinguists questioned the inclusion of indexical phenomena in sound symbolism sensu stricto [3: 1626; 14: 3]. Be that as it may, Ohala's frequency code still remains a «bold and synthetic proposal» and in need of further inquiries [15: 2]. Crucially, experimental replications of the supposed transfer of meaning between the two aforementioned domains are still lacking. In this study, I try to fill this gap by investigating the results of a two-alternative forced choice task submitted to Italian Florentine participants. In the Florentine variety, voiceless plosives with long-lag voice onset times in non-postvocalic positions can be associated with social meanings related to manliness [16]. Through repeated co-occurrence [3: 1626-1627] of variants and indexical values, Florentine speakers might be able to extract meaning features and coherently use them in denotation. Therefore, I anticipated to find a consistent pattern of choice of aurally presented nonce stimuli containing long-lag VOTs (vs. short-lag VOTs and distractors) for the denotation of male (vs. female) living beings. The experiment failed to support this expectation, but sound-symbolic links between reference and variation were nonetheless found.

Methods: 55 (33 targets, 22 distractors) Italian pseudowords were generated in WORDEN [17] together with information on their phonological neighborhood density [18] and frequency. The targets had a ˈCaCV structure, with an equal number of initial (cf. [19]) [k t p], a variable non-voiceless-plosive second consonant, and vowel ending in either [a o e] in order to statistically control for the effect of Italian grammatical gender on participant choices (cf. [20]). The distractors presented an analogous structure while containing an initial non-plosive segment instead of [k t p] (e.g., target [ˈtave] vs. distractor [ˈsave]). Two Florentine speakers (a 26-year-old man and a 27-year-old woman) read the stimuli in an anechoic chamber. Four repetitions of the 55-item list were recorded (44.1 kHz, 24 bit): two unsupervised and two guided pronunciations of the items. In the latter case, speakers were instructed to "emit a puff of air" after the initial voiceless plosive [16; 21] of the targets: the production of an aspirated plosive was pinpointed through positive feedback during this training phase. The best repetitions per condition and speaker were selected in terms of overall clarity. The resulting unsupervised, unaspirated set had a mean VOT of 23.3 ms. (12.33 st. dev.). Following [22], the bursts and the releases of these plosives were replaced in Praat [23] with the equivalent components of their

aspirated counterparts from the guided productions. This aspirated set had a mean VOT of 127 ms. (23.41 st. dev.). The full batch of 176 recordings (((33 targets x 2 aspiration conditions) + 22 distractors) x 2 voices) were used as stimuli (mean duration 499.68 ms., 86.96 st. dev.) in a two-alternative forced choice task built with a custom Perceval [24] script. The experimental session had the following structure: firstly, instructions were textually presented, framing the experiment into a fictitious discovery of new animal species, and asking participants to match singularly presented aural stimuli with the supposed name of the male or female specimen. No time limit was coded into the experiment, but participants were invited to make their choices using the left (for male) or right (for female) keyboard arrows as quickly as possible. Then, the 176 stimuli were presented as a sequence of randomized trials with 1.5 secs. of inter-stimulus interval (i.e., a silent black screen). During each trial, one stimulus was reproduced together with an image of a blue Mars (left) and pink Venus symbol (right) on a white background (cf. [25] for the criteria of image construction). Lastly, a textual thank you message concluded the session. Both choices and response times (RT) were recorded through Perceval. 22 last- year high school students (15 boys, 7 girls) from a *liceo* in Scandicci (Florence) took part to the experiment. Each student completed an ad hoc translated and validated (five factors, all $\alpha \geq 0.69$) version of the Motivated Strategies for Learning Questionnaire (MSLQ) [26] in order to statistically control for the effects of student engagement in school activities on the experimental results. Moreover, individual interviews were recorded to gather together a set of profiling variables (see below). The experiment was administered through a Windows laptop and a pair of Sennheiser HD 25 Light closed headphones in the school library. Each individual session (questionnaire, interview, and experiment) lasted around one hour.

Analysis and results: 3,872 trials were recorded through Perceval. Given the absence of explicit RT limits, an interquartile range criterion of outlier detection was imposed to the dataset, ultimately pruning 289 observations (RT > 5,263 ms.). The results of the remaining 3,583 trials were dummy coded (1 = M, 0 = F) and inserted as response in a generalized linear mixed effect model structure including participant and item as random factors. Condition (aspirated, filler, unaspirated) was evaluated together, and in interaction with several controls (stimulus target consonant, second consonant voicing, vowel ending, voice sex, phonological neighborhood density and frequency; participant RT, sex, regionality [27], 5 MSLQ factors, Florentine dialect attitude, parents' education, musical proficiency, number of languages known, domestic animal possession). Interactions between significant controls and the other factors were also computed. Model selection was performed manually with the aim of pinpointing a good compromise between Akaike and Bayesian Information Criterion scores. Vowel ending was of paramount importance in determining participant responses (percentages of "male" per category: *-a*, 13.32%; *-e*, 51.36%; *-o*, 94.95%); nonetheless, once morphological gender was controlled for, two other small, but significant effects emerged: Condition (filler: est. -.34, std. error .14, p .012; unaspirated: est. .29, std. error .12, p .019) and Stimulus voice sex (male: est. .3, std. error .11, p .004). Moreover, vowel ending interacted with z-scored RT (*-*e*: est. -.3, std. error .1, p .002; *-*o*: est. -.78, std. error .14, p <.001), revealing that participants tended to less morphologically congruent responses as their RT increased. The working hypothesis was contradicted through the opposite directionality of the Condition predictor: aspirated stimuli were more frequently associated with the female, and not the male, fictitious specimen. This may be explained with reference to more general, less Florence-related, associations between being female and clear, hyperarticulated speech, which seem to have both broad sociophonetic and biological explanations [28]. The other significant fixed effect more closely resembles the mechanism of Ohala's frequency code, as the perceivable stimulus voice sex was mapped to analogous referential features (i.e., nonces uttered by a male voice were more frequently associated with male animal specimens). Note that, since male mammals are usually bigger than females, this can be considered as a roundabout replication of Ohala's F_0-to-size mapping; however, previous research [29] failed to induce referent size associations through simple F_0 manipulation, so that the effect observed here might be triggered by more complex acoustic co-variations characterizing gendered voice. At present, we can assume that transfers between the indexical and the referential can be experimentally prompted using central, less- ideologically elaborate properties of the former domain. Future research should experiment a) with social variants which are more salient and

systematic than Florentine long-lag VOTs and b) with languages without noun gender markers steering participant associations.

Bibliography

[1] KÖHLER, W. (1929). *Gestalt psychology*. New York: Liveright.

[2] SAPIR, E. (1929). A study in phonetic symbolism. In *Journal of Experimental Psychology*, 12(3), 225-239.

[3] SIDHU, D. M., PEXMAN, P. M. (2018). Five mechanisms of sound symbolic association. In *Psychonomic Bulletin & Review*, 25, 1619-1643.

[4] NIELSEN, A. K., DINGEMANSE, M. (2021). Iconicity in word learning and beyond: a critical review. In *Language and Speech*, 64(1), 52-72

[5] KAWAHARA, S. (2020). Sound symbolism and theoretical phonology. In *Language and Linguistic Compass*, 14(8), 1-17.

[6] ECKERT, P. (2012). Three waves of variation study. In *Annual Review of Anthropology*, 41, 87-100.

[7] OHALA, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. In *Phonetica*, 41(1), 1-16.

[8] ECKERT, P. (2010). Affect, sound symbolism, and variation. In *University of Pennsylvania Working Papers in Linguistics*, 15(2), 70-80.

[9] LEE, N. H. (2021). Style variation in the second formant. What does it mean to be "refined" in Baba Malay? In *Language Ecology*, 4(1), 115-130.

[10] FOULKES, P. (2010). Exploring social-indexical knowledge: a long past but a short history. In *Laboratory Phonology*, 1(1), 5-39.

[11] JOHANSSON, N., ZLATEV, J. (2013). Motivations for sound symbolism in spatial deixis. In *Public Journal of Semiotics*, 5(1), 3-20.

[12] JOHNSTONE, B. (2010). Locating language in identity. In LLAMAS, C., WATT, D. (Eds.), *Language and Identities*. Edinburgh: University Press, 29-36.

[13] D'ONOFRIO, A., ECKERT, P. (2021). Affect and iconicity in phonological variation. In *Language in Society*, 50(1), 29-51.

[14] AKITA, K. (2021). Phonation types matter in sound symbolism. In *Cognitive Science*, 45, 1-14.

[15] WINTER, B. et al. (2021). Rethinking the frequency code. In *Philosophical Transactions B*, 376, 1-13.

[16] PICCARDI, D. (2017). Sociophonetic factors of speakers' sex differences in Voice Onset Time: A Florentine case study. In BERTINI, C. et al. (Eds.), *Social and Biological Factors in Speech Variation. Studi AISV 3*. Milano: Officinaventuno, 83-106.

[17] ORIGLIA, A., CANGEMI, F. & CUTUGNO, F. (2015). WORDEN: a PYTHON interface to automatic (non)word generation. In VAYRA, M., AVESANI, C. & TAMBURINI, F. (Eds.), *Language Acquisition and Language Loss. Studi AISV 1*. Milano: Officinaventuno, 459-470.

[18] SCHMITZ, D. (2022). Cuteness modulates size sound symbolism at its extremes. Paper presented at Iconicity Seminar 2022 (IcoSem2022).

[19] KAWAHARA S., SHINOHARA K & UCHIMOTO Y. (2008). A positional effect in sound symbolism: an experimental study. In *Proceedings of the Japan Cognitive Linguistics Association*, 8, 417-427.

[20] DANESI, M. (1998). Gender assignment, markedness, and indexicality: Results of a pilot study. In *Semiotica*, 121(3-4), 213-240.

[21] KIM, C. (1970). A theory of aspiration. In *Phonetica*, 21(2), 107-116.

[22] NIELSEN, K. (2011). Specificity and abstractness of VOT imitation. In *Journal of Phonetics*, 39(2), 132-142.

[23] BOERSMA, P., WEENINK, D. (2022). PRAAT: doing phonetics by computer. [software] Version 6.2.23. https://www.praat.org/

[24] GHIO, A. et al. (2005). PERCeption EVALuation Auditive & Visuelle. [software] Version 3.0.3. http://www2.lpl-aix.fr/~lpldev/perceval/

[25] FARNAND, S. P., FAIRCHILD, M. D. (2014). Designing pictorial stimuli for perceptual experiments. In *Applied Optics*, 53(3), 72-82.

[26] PINTRICH P. R., DEGROOT E. (1990). Motivational and self-regulated learning components of classroom academic performance. In *Journal of Educational Psychology*, 82(1), 33-40.

[27] CHAMBERS J. K., HEISLER T. (1999). Dialect topography of Quebéc City English. In *Canadian Journal of Linguistics*, 44(1), 23-48. [28] SIMPSON, A. P. (2009). Phonetic differences between male and female speech. In *Language and Linguistics Compass*, 3(2), 621-640. [29] ROJCZYK, A. (2011). Sound symbolism in vowels. In *Poznań Studies in Contemporary Linguistics*, 47(3), 602-615.

Prosodic markers of narrow focus in Italian L1 and L2 (French L1): an experimental study

Bianca Maria De Paolis*°, Antonio Romano *, Cecilia Andorno*
**Università degli studi di Torino, °Université Paris 8 / CNRS*

Il nostro studio mira a confrontare la realizzazione prosodica del *focus* stretto da parte di parlanti nativi di italiano e di apprendenti francofoni di italiano L2. L'italiano e il francese, pur appartenendo alla stessa famiglia di lingue romanze, si differenziano dal punto di vista intonativo e accentuale; per quanto riguarda la focalizzazione, è stato osservato che i parlanti delle due lingue utilizzano strategie diverse per dare prominenza alla porzione di enunciato interessata dal *focus* (Michelas & German 2020, Delais-Roussarie et al. 2015, Féry 2001 per il francese; Gili Fivela et al. 2015, Bocci & Avesani 2005 per l'italiano, Romano & Interlandi 2002 in particolare per la varietà di Torino, nostro punto d'inchiesta). La struttura prosodica - ma anche sintattica - di queste due lingue porta a manifestazioni diverse del rilievo focale all'interno dell'enunciato: per esempio, il fatto che l'italiano sia una lingua ad accento lessicale, mentre il francese no; la maggior flessibilità dell'italiano nell'organizzazione dei costituenti all'interno della frase rispetto al francese, e così via. Se già la descrizione dei fenomeni di focalizzazione prosodica da parte dei parlanti nativi suscita vivo dibattito tra gli studiosi, si può dire che lo studio dello stesso fenomeno in contesto di L2 sia un terreno ancora da scoprire: nessuno, a nostra conoscenza, si è occupato per il momento di questo in maniera sistematica dell'italiano L2 - francese L1. Studi esistenti su fenomeni vicini (Andorno & Turco 2015 per il contrasto di polarità, ad esempio) o su altre combinazioni L1-L2 (Zubizarreta & Nava 2011) evidenziano che la marcatura prosodica della struttura informativa dell'enunciato si rivela di difficile gestione per gli apprendenti, e può dar luogo a fenomeni di interferenza dalla L1 anche nei parlanti di livello più avanzato. L'organizzazione del piano informativo dell'enunciato, inoltre, risulta per i bilingui adulti particolarmente complessa anche in relazione allo stretto legame intrattenuto con la sintassi (v. "ipotesi dell'interfaccia", Belletti et al. 2007).

L'obiettivo di questo nostro studio è quindi quello di descrivere, in un primo momento, le strategie adottate dai parlanti nativi di francese e italiano in specifici contesti di focalizzazione; in un secondo momento, di confrontare questo repertorio con quello osservabile nel gruppo di apprendenti. In questo modo, è possibile formulare un'ipotesi sul ruolo dell'influenza cross-linguistica e dei principi propri all'acquisizione L2 nelle produzioni dei parlanti non-nativi. Le nostre domande di ricerca sono, quindi, le seguenti:

- Q1. Quali sono le strategie prosodiche utilizzate dai parlanti nativi di francese e italiano per esprimere diverse tipologie di focus (largo, stretto identificativo, stretto correttivo…)?
- Q2. Gli apprendenti sfruttano le stesse strategie dei parlanti della lingua *target*?
- Q3. Le divergenze tra le produzioni dei nativi e degli apprendenti sono attribuibili all'interferenza della L1 o, invece, a dei principi universali di acquisizione (v. Archibald 1993)?

Il corpus raccolto per condurre l'analisi è costituito dalle produzioni di tre gruppi di parlanti: francofoni nativi (15 p.), italofoni nativi (15 p.), francofoni residenti a Torino, apprendenti di italiano L2 in contesto non guidato (15 p.), per un totale di 45 partecipanti. Il livello di competenza lessicale e morfo-sintattica dei parlanti L2 si attesta intorno ai livelli B2/C1 del QECR ed è stato controllato con la somministrazione di un assessment test al momento della raccolta dati. Le registrazioni sono state effettuate in due punti d'inchiesta, Torino e Parigi, per ridurre l'impatto della variazione diatopica sulle produzioni dei parlanti. Il *task* usato per elicitare le produzioni dei partecipanti è adattato dal modello di Gabriel (2010): al parlante viene mostrata una presentazione *Power Point* contenente due brevi storie a fumetti, corredate da una didascalia, e a seguire alcune domande a proposito delle scene illustrate nei fumetti. Le domande sono formulate in modo da elicitare risposte con focus largo, identificativo e correttivo su diverse componenti sintattiche della frase: soggetto, verbo, argomenti del verbo, circostanziali. Al partecipante viene richiesto di rispondere ad alta voce a queste domande; le sue risposte vengono registrate e in seguito trascritte e segmentate per l'analisi. Per il gruppo L1 francese il *task* viene proposto in lingua francese, mentre per i

gruppi ITL1 e ITL2 viene proposto in italiano.

Fig.1 Esempi di diapositive dal *task* nella versione in italiano.

L'analisi prosodica è stata condotta tramite lo script Praat Polytonia (Mertens 2014), e integrata da un'annotazione e ricategorizzazione manuale da parte degli autori e di esperti madrelingua italiani e francesi. La scelta di utilizzare Polytonia è dettata dall'esigenza di adottare, in presenza di parlato non-nativo, un sistema di etichettatura semi-automatica che non utilizzi parametri language-dependent, e che non assuma a priori la presenza di categorie prosodiche specifiche. I risultati mostrano che i parlanti nativi di italiano e francese impiegano un repertorio simile di strategie prosodiche e, in stretta relazione a quelle prosodiche, specifiche costruzioni sintattiche; tuttavia, queste strategie non vengono sfruttate con la stessa frequenza nei due gruppi. In francese, i parlanti nativi producono sistematicamente frasi scisse (c'est X qui/que…), solitamente accompagnate da profili prosodici marcati; in italiano, i parlanti producono meno strutture scisse (è X che…), ma realizzano sistematicamente curve prosodiche marcate per i costituenti a focus stretto, identificativo o correttivo (per una descrizione dettagliata di ciò che consideriamo marcato e non marcato, rimandiamo a De Paolis et al. 2022). Nei due gruppi L1, inoltre, si può dire che la prosodia vada di pari passo con la sintassi: laddove si utilizza una struttura sintattica marcata (nel nostro corpus maggioritariamente frasi scisse), aumenta anche la frequenza dei pattern di marcatura prosodica. I risultati confermano dunque le descrizioni dei sistemi nativi presenti in letteratura (Mertens 2012, Avesani & Vayra 2001). Per quanto riguarda l'italiano L2, i risultati mostrano che gli apprendenti hanno acquisito le funzioni pragmatiche della frase scissa nella lingua target: l'espressione del focus identificativo e correttivo è spesso garantita dall'uso di queste. Tuttavia, i mezzi prosodici utilizzati dagli apprendenti non sono gli stessi osservati nei due gruppi L1: gli apprendenti, infatti, realizzano un pattern intonativo neutro, dettato più che altro dalla posizione del costituente focalizzato all'interno dell'enunciato: profilo di continuazione se il costituente si trova a inizio enunciato, dichiarativa neutra se il costituente si trova a fine enunciato; questo accade indipendentemente dal fatto che il costituente si trovi in condizione di focus largo o stretto. La frequenza di strutture prosodicamente marcate è dunque molto più bassa nel gruppo L1, con una differenza a livello statistico che si rivela altamente significativa ($p<0.005$). Questo risultato sembra suggerire che, nell'espressione della struttura informativa dell'enunciato, gli apprendenti non si affidino alla prosodia nella stessa misura dei parlanti nativi: nel gruppo L2, l'articolazione focus/background avviene per lo più attraverso la sintassi. Questo aspetto, che non è attribuibile all'influenza della L1, potrebbe far pensare a una caratteristica tipica dell'interlingua: il primo livello di acquisizione sarebbe quello della sintassi, e la prosodia arriverebbe solo in una fase avanzata dell'acquisizione, contrariamente a quanto si osserva in contesto di acquisizione L1.

Bibliografia

Andorno, C., Turco, G. (2015). Embedding additive particles in the sentence information structure: How L2 learners find their way through positional and prosodic patterns, *Linguistik Online*, 71, 2.

Archibald, J. (1993). *Language Learnability and L2 Phonology*. Dordrecht: Klower.

Avesani, C., Vayra, M. (2001). Costruzioni marcate e non marcate in italiano: il ruolo dell'intonazione. Quaderni dell'Istituto di Fonetica di Padova, 3.

Belletti, A., Bennati, E., Sorace, A. (2007) Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian, *Natural Language & Linguistic Theory*, 25.

Bocci, G., Avesani, C. (2005). Focus Contrastivo nella periferia sinistra: un accento ma non solo un accento. In Savy,

R., Crocco, C. (Eds.), Analisi Prosodica. Teorie, modelli e sistemi di annotazione. Atti del secondo convegno AISV- Associazione Italiana di Scienze della Voce.

Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S. A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R., Yoo, H. Y. (2015). Intonational phonology of French: Developing a ToBI system for French. In: Frota, S., Prieto, P. (eds), *Intonation in Romance*. Oxford: Oxford: Oxford University Press.

De Paolis, B. M., Santiago, F., Andorno, C. (2022) Syntaxe ou prosodie? Une étude préliminaire sur l'expression de la focalisation étroite par les apprenants italophones de français L2. *Proc. XXXIVe Journées d'Études sur la Parole - JEP 2022.*

Féry, C. (2001). Focus and phrasing in French. Féry C., Sternefeld, W. (eds.), *Audiatur vox sapientiae*. Berlin: Akademie-Verlag.

Gabriel, C. (2010). On focus, prosody, and word order in Argentinean Spanish: A minimalist OT account. *Revista virtual de estudos da linguagem*, 4.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. & Sorianello P. (2015). Varieties of Italian and their intonational phonology. In Frota, S., Prieto, P. (Eds), Intonation in Romance. Oxford: Oxford University Press.

Mertens, P. (2012). La prosodie des clivées. Caddeso S., Roubaud,M. N., Rouquier, M., Sabio, F. (eds), *Penser les langues avec Claire Blanche Benveniste*. Aix-en-Provence : Presses Universitaires de Provence.

Mertens, P. (2014). Polytonia: a system for the automatic transcription of tonal aspects in speech corpora. *Journal of Speech Sciences*, 4, 2.

Michelas, A. & German, J. (2020). Focus Marking and Prosodic Boundary Strength in French. Phonetica, 77(4).

Roggia, C. E. (2008). "Frasi Scisse in italiano e in francese orale: evidenze dal C-ORAL-ROM.", *Cuadernos de Filología Italiana*, 15: 21.

Romano, A., Interlandi, G. (2002). Quale intonazione per il torinese? Regnicoli, A. (eds), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia: Atti delle 12/e Giornate di studio del Gruppo di fonetica sperimentale*. Roma: Il Calamo.

Zubizarreta, M. L., Nava, E. (2011). Encoding discourse-based meaning : Prosody vs. syntax. Implications for second language acquisition. *Lingua*, 121.

Prosodic features of *quello* in Informal Neapolitan Italian: data from spontaneous speech

Claudia Crocco, Kim Groothuis, Giovanni Leo & Giuseppe Magistro
Ghent University

Introduction: Campanian dialects show a pragmatic use of the third-person pronoun *chillo,* originally a distal demonstrative (*chillo*, *chella*, *chello* 'that.M/F/N.SG'). This pronoun seems to function as an expletive subject of impersonal predicates and seems also involved in certain topic constructions [6,7,9,10]. In both cases, it has a clear – though hard to pin down exactly – function at the discourse level. Due to contact, a similar pragmatic use of the demonstrative *quello* (*quello/a,* 'that.M/F.SG', *quelli/e* 'that.M/F.PL'; expletive: ex. (1); topic: ex. (2)) percolates in markedly informal or socially low varieties of Campanian Italian [2]. The present study aims at exploring the prosodic properties of pragmatic *quello* in one of these varieties, i.e. informal Neapolitan Italian (henceforth: INI).

(1) *Quello piove*
that.N rain.3SG
'(that is because/the fact is that) It rains.' [2]

(2) *E certo, quello il brodo di gallina è sostanzioso.*
and sure that.M the broth.M of chicken be.3SG substantial
'And certainly, chicken broth is substantial.' (De Filippo, *Natale in Casa Cupiello*)

Although pragmatic *quello* has received some attention within the syntactic literature, its prosodic properties are totally uncharted, despite their most likely relevance for syntactic and informational analysis. For instance, it has been claimed that *quello* in utterances such as (2) has "the intonational contour […] of a canonical topic-comment structure" ([7], 262), in which, however, *quello* and the following NP would not belong to the same prosodic phrase. This is taken as evidence that the two do not form one syntactic constituent [7,10]. Such intriguing observations, however, are regrettably only based on impressionistic evaluations.

Aims: The main goal of the study is to contribute to fill this gap by exploring the prosodic properties of *quello* in a corpus of real-world INI. In doing so, we also aim at elaborating the above-mentioned observations concerning (a) the presence/absence of an overarching contour including pragmatic *quello* and the following NP (cf. ex. (2)); (b) the presence/absence of detectable differences when pragmatic *quello* is co-referent with a focalized or topicalized constituent. The study thereby contributes to the description of the prosody of the Italian linguistic space and especially to the study of diaphasically-marked regional varieties. Since pragmatic *quello* is involved in the expression of topic and focus [5], the study of its prosodic properties will also contribute to our understanding of the syntax-prosody interface.

Methodology: To collect a corpus of INI, we relied on the *Pyrlato* pipeline (developed for a parallel work [8]) to extract natural data from YouTube videos, mainly movies, TV shows, and comedy sketches. The corpus therefore includes scripted and spontaneous productions. First, *Pyrlato* uses the module *PyTube* to search for specific keywords (in our case, names of local shows and public figures), resulting in a list of relevant videos. Second, it checks that the videos are in Italian. *Pyrlato* then generates an object scraping YouTube's automatic subtitles, which it subsequently searches for specific words (here, 'quello/a/i/e'). The sentence containing the word is extracted in .mp4 and trimmed observing a time-window of 7s. Finally, .mp4 videos are converted to .mp3. The script will be soon available on GitHub.

Data in lossy formats and/or recorded in noisy conditions are usually discarded in prosodic studies as they do not meet the sound quality standards expected for most acoustic analyses. Such data, however, do represent an invaluable source of information concerning the speakers' actual linguistic uses, exactly for their natural character and in spite of their lesser acoustic quality. Of course, this choice has consequences for the type of analyses that can be carried out on the data. In this paper we focus on the analysis of pitch, which is a relatively robust prosodic feature [4]. We aim at detecting through auditory and visual inspection the presence and (when possible) the type of pitch accent occurring on *quello* and on the co-referent NP.

Preliminary results: Data collection and analysis are still in progress. At present, a total of 1397 videos

(ca. 550h) were scraped, creating a corpus of 1775 instances of *quello/a/i/e*. The occurrences of *quello* as a free relative or a deictic, dialectal examples, and false starts were discarded (1527 occurrences). The remaining instances of *quello* (N=248) were subdivided into different groups according to: i) the presence of a co- referent NP in the sentence; ii) the linear syntactic position of *quello* and the co-referential NP; iii) the type of verb (copular/lexical; Table1: col. "Type").

Results: Table 1 presents the types and numbers of constructions with *quello* found in the data:

EX	NON-PA		SYNTACTIC TYPE			N	PA
			?	LOW-Q			
(3)	*Quello*$_i$ DP$_i$ Pred	*Quella mammà sta in paradiso* (*La lettera di mammà*) 'Mom is in heaven'	30	16	1	5	8
(4)	*Quello*$_i$ Pred DP$_i$	*Quelli rimangono impressionati i bambini* (*Bello di papà*) 'The children get creeped out'	6	5	0	1	0
(5)	*Quello*$_{subj}$	*Quella a te ti ascolta* (*Giuro che ti amo*)	99	62	2	15	20 'Sl
(6)	*Quello*$_i$ Cop DP$_i$	*E quell'è l'umidità* (*Totò, Peppino e le fanatiche*)	102	61	5	11	25 'A
(7)	*Quello*$_i$ cleft DP$_i$	*Quello è lui ... fra Martino.. che* (*I 4 monaci*) 'That's him.. Friar Martin... who'	6	4	0	2	0
(8)	*Quello*$_i$ DP$_i$ Cop	*Quello un bottone è* (*The Jackal*) 'It's just ONE button'	1	1	0	0	0
(9)	*Quello*$_{Expl}$	*Quello eravamo prima un gruppo più ampio* (TGR) 'At first we were a larger group'	4	3	0	1	0
		TOTAL	248	151	8	35	
			53				

The 248 utterances containing a *quello* were auditorily and visually inspected in Praat [1], and manually annotated for the presence of pitch accents on *quello* and the co-referent NP. Due to low signal quality, 53 files could not be analysed and were excluded. The results are schematized in Table 1.

Quello appears marked by a high/rising pitch accent in 151 out of 248 occurrences across the different syntactic-pragmatic contexts (Figure 1-2)[1]. As for the presence of an overarching contour including *quello* and the following NP, our preliminary results indicate that both elements are pitch accented (Fig.1) but still included in the same phrase since no perceivable boundary separates them. As for the possible differences when pragmatic *quello* is co-referent with a focalized or topicalized constituent, our results indicate that *quello* is marked by a high/rising accent, while the realization of the NP varies according to its (pre-, post-, or focal) position in the utterance (Fig.2), in line with the preceding results about Neapolitan Italian [2].[2]

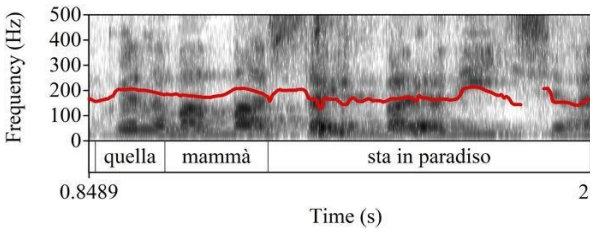


Figure 1: Intonation contour of ex. (3), Table 1

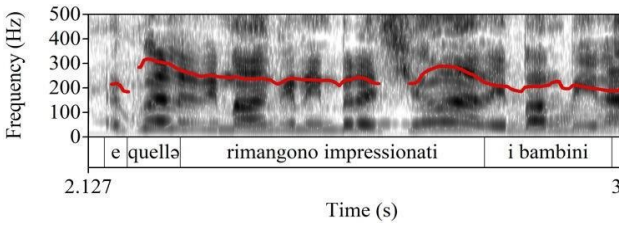


Figure 2: Intonation contour of ex. (4), Table 1

References

[1] P. Boersma, D. Weenink, Praat: doing phonetics by computer, (2021).

[2] N. De Blasi, F. Fanciullo, La Campania, in: M. Cortelazzo, C. Marcato (Eds.), I dialetti italiani. Storia, struttura, uso, UTET, Turin, 2002: pp. 628–678.

[3] M. D'Imperio, Italian intonation: an overview and some questions, Probus. 14 (2002) 37–69.

[4] R. Fuchs, O. Maxwell, The Effects of mp3 compression on acoustic measurements of fundamental frequency and pitch range, in: Speech Prosody 2016, ISCA, 2016: pp. 523–527.

[5] K. Groothuis, Reconsidering the pragmatic use of *chillo* in Campania, Paper presented at Going Romance, Barcelona, 2022. [6] A.N. Ledgeway, Grammatica diacronica del napoletano, Niemeyer, Tübingen, 2009.

[7] A.N. Ledgeway, Subject licensing in CP: the Neapolitan double-subject construction, in: P. Benincà, N. Munaro (Eds.), Mapping the left periphery, Oxford University Press, Oxford, 2010: pp. 257–296.

[8] G. Magistro & C. Crocco, Scraping YouTube audio data with *Pyrlato*: an example with Italian *mica*, (Ms.).

[9] E. Radtke, I dialetti della Campania, edizione italiana a cura dell'autore, con la collaborazione di P. Di Giovine e F. Fanciullo, Il Calamo, Roma, 1997.
[10] R. Sornicola, Alcune strutture con pronome espletivo nei dialetti italiani meridionali, in: P. Benincà, G. Cinque, T. De Mauro, N. Vincent (Eds.), Italiano e dialetti nel tempo. Saggi di grammatica per Giulio C Lepschy, Bulzoni editore, Rome, 1996: pp. 323–340.

[1] We excluded that the observed high pitch is mainly due to microprosodic or phrase-initial resetting because of both the frequent realization of the peak in later syllables and the presence of the tone after preceding segmental material (e.g., *e*, *ma*) within the same phrase.
[2] We cannot provide a systematic description of the pitch accent and its differentiation across the different pragmatic types because of the uncontrolled nature of the data (e.g., different speakers each time, differences in speech rate, different recordings, etc.).

Ethnic and gender stereotypes: linguistic discrimination through the intersectional lens

Elena Clemente[1], Rosalba Nodari[2]
[1]Università degli Studi di Pavia, [2]Università degli Studi di Siena

Molta ricerca bibliografica si è concentrata sulla valutazione sociale degli accenti non nativi, dimostrando come essi sono, in genere, valutati più negativamente lungo le dimensioni del calore e della competenza (Rakić et al. 2011, Pantos

& Perkins 2013). L'accento è però un contrassegno identitario a tutti gli effetti che permette di veicolare non solo l'eventuale retroterra migratorio o provenienza geografica, ma anche l'identità di genere, l'orientamento sessuale, la classe ecc. Secondo questa prospettiva risulta quindi cruciale assumere una prospettiva intersezionale che tenga conto di come una eventuale discriminazione o svalutazione linguistica è una discriminazione che riguarda il complesso di identità che vengono veicolate con la lingua stessa (Levon 2015, Nodari 2021).

Il seguente studio esplorativo punta proprio a colmare questa lacuna, esplorando in che modo gli eventuali stereotipi legati agli accenti non nativi si intersecano con eventuali stereotipi di genere. A partire dal Modello del Contenuto degli Stereotipi (Fiske et al., 2002, MacFarlane & Stuart-Smith 2012 per un'applicazione in sociolinguistica) e sulla base di ricerche precedenti (Masullo & Meluzzi 2020, Calamai et al. 2020), è stato svolto un esperimento percettivo tramite questionario. L'esperimento è stato creato a partire dalla registrazione di dodici parlanti distinti per genere e per provenienza, ossia sei coppie composte da un uomo e una donna, con sei tipologie diverse di accenti, sia nativi che non nativi. Per i due accenti nativi sono stati registrati quattro parlanti di italiano regionale: due con accento settentrionale, provenienti dalla Lombardia, e due con accento meridionale, provenienti dalla Sicilia. Per gli accenti non nativi si sono selezionate quattro coppie di parlanti: due parlanti con accento inglese, provenienti da Gran Bretagna e Stati Uniti, due parlanti francesi, due parlanti di arabo, originari di Marocco e Libia e, infine, due parlanti di lingue slave, provenienti da Bosnia e Serbia. Gli accenti stranieri sono stati selezionati in modo da rappresentare due gruppi, uno associato ad una condizione socio-economica e socio-culturale media o elevata (accento inglese e francese) e uno associato ad una posizione più svantaggiata e numericamente significativa sul territorio italiano (accento slavo e arabo). Per la metodologia di indagine è stata scelta la tecnica del *verbal guise* (cf. Dragojevic & Goatley-Soan 2022): le dodici voci hanno cioè letto uno stesso testo, ossia una breve previsione meteo dalla quale era stato espunto ogni possibile riferimento geografico. Il questionario constava di diverse sezioni. Nella prima sezione del questionario veniva descritto brevemente l'esperimento in modo da non rivelare apertamente l'obiettivo dello studio e rassicurare i e le rispondenti circa l'anonimato dei dati e delle risposte rilasciati; una seconda parte conteneva una serie di domande volte a raccogliere i dati relativi ad alcune caratteristiche dei e delle informanti; infine, l'ultima sezione consisteva nell'esperimento vero e proprio, con l'ascolto delle dodici voci. Utilizzando le domande del Modello del Contenuto degli Stereotipi è stato possibile ottenere delle valutazioni rispetto alle quattro dimensioni di competenza, calore, status e competizione, basate su una scala Likert da 1 a 5. Nell'ultima parte del questionario, era presente anche una domanda a risposta chiusa in cui si chiedeva di individuare la provenienza dei e delle parlanti. I questionari sono stati distribuiti online, attraverso piattaforme di messaggistica istantanea e social network, dal 24/07/2022 al 02/08/2022.

In totale sono state ottenute 97 risposte provenienti da rispondenti di un'età compresa tra i 18 e i 69 anni, provenienti da tutte le parti d'Italia e dall'estero, e i dati sono stati analizzati con R (R Core Team, 2019). A partire dalle medie dei punteggi attribuiti alle dimensioni di competenza e calore, è stata creata una matrice 12 x 2, in modo da ricavare uno spazio bidimensionale, e sono stati quindi svolti un *clustering* gerarchico con metodo di Ward e il calcolo dei valori di *Silhouette*, grazie ai quali è stato possibile individuare il numero di *cluster* ottimali per due campioni ovvero il campione totale e un sottocampione composto dai e dalle rispondenti di età tra i 60 e i 69 anni, in quanto questa particolare fascia di età si è rivelata essere statisticamente significativa sulla base di un'analisi di regressione lineare e ANOVA. Per ottenere la

rappresentazione dei gruppi nello spazio bidimensionale, è stato invece svolto un *clustering* non gerarchico di tipo *K-Means*.

Per quanto riguarda il primo campione, sono stati individuati cinque raggruppamenti: un primo *cluster* composto da entrambe le voci con accento slavo, che si caratterizzava per punteggi più bassi su entrambe le dimensioni e si configurava come un possibile giudizio di tipo sprezzante (Fiske et al., 2002); un secondo *cluster* formato invece dalle voci femminili inglese, francese e italiana meridionale e dalla voce inglese maschile, con punteggi medi per entrambe le dimensioni; un terzo *cluster*, composto solo dalle due voci femminili araba e italiana settentrionale, con punteggi più elevati per quanto riguarda la competenza e simili per la dimensione del calore rispetto al secondo *cluster*; un quarto *cluster*, caratterizzato unicamente dalla presenza della voce araba maschile, con punteggi leggermente più alti rispetto al primo raggruppamento e, infine, un quinto *cluster* composto unicamente dalle voci maschili francese e italiane, caratterizzate da punteggi più alti in assoluto per entrambe le dimensioni e rispetto alle quali è possibile ipotizzare un giudizio di ammirazione. Per quanto riguarda il secondo campione con le valutazioni provenienti dai e dalle rispondenti di età superiore ai 60 anni, sono stati rilevati tre raggruppamenti: il primo, uguale a quello ottenuto per il campione totale, conteneva le voci con accento slavo; il secondo, anch'esso caratterizzato da bassi punteggi per il calore, ma leggermente più alti per quanto riguarda la competenza, conteneva invece la voce con accento arabo maschile e le due voci di italiano regionale meridionale, mentre il terzo *cluster* raggruppava le restanti voci, caratterizzate da punteggi più elevati per la dimensione del calore.

Purtuttavia, la risposta all'ultima domanda ha dimostrato la difficoltà da parte dei e delle rispondenti di classificare correttamente le voci: nonostante la quasi totalità del campione abbia riconosciuto correttamente le voci con accento di italiano regionale, ognuna di queste viene identificata con almeno una provenienza errata; in particolare, la voce maschile con accento meridionale è quella meno riconosciuta (74,8%) all'interno di questo gruppo. Per quanto riguarda la classificazione degli altri accenti, le voci con accento inglese sono quelle identificate correttamente da un numero maggiore di rispondenti (almeno il 70% del totale) al contrario delle restanti voci straniere, che presentano invece percentuali minori di identificazioni corrette: qui, infatti, la maggioranza delle risposte classifica in modo errato la provenienza di chi parla ad eccezione unicamente della voce femminile francese, classificata correttamente dal 67,8% delle risposte.

In complesso, i dati confermano l'esistenza di un pregiudizio ancora diffuso nei confronti degli accenti non nativi dell'est Europa, indipendente dal genere di chi parla (Calamai, 2015), ma anche la difficoltà nell'individuare correttamente la provenienza di chi parla (Bianchi & Calamai, 2012; Calamai, 2015). Infine, si nota una differenza generazionale per quanto riguarda la percezione dei parlanti di italiano regionale meridionale (Masullo & Meluzzi 2020), in quanto questo viene associato a più bassi livelli di calore per i e le rispondenti di età più avanzata. Più complessi risultano invece gli stereotipi legati al genere. Sembra infatti che il pregiudizio legato ad alcuni accenti non nativi abbia traiettorie diverse a seconda del genere del parlante: è il caso dell'accento arabo, il quale viene giudicato come più competente quando associato a una voce maschile. I risultati mostrano, in genere, una differenza nei punteggi rispetto alla dimensione della competenza tra uomini e donne di una stessa coppia, per cui ai primi sono attribuiti punteggi più elevati rispetto alla propria controparte femminile. Il dato può essere letto alla luce di una generale percezione della categoria sociale delle donne come meno capaci in ambito lavorativo e può anche confermare una interiorizzazione dello stesso pregiudizio da parte delle stesse rispondenti donne. Per la sua natura esplorativa, lo studio presenta sicuramente il limite di considerare un campioni ridotto di voci e accenti, elemento che ha tuttavia permesso di evitare una durata eccessiva dell'esperimento. I risultati sottolineano la necessità di ampliare il campione di voci così da poter considerare e controllare eventuali caratteristiche individuali dell'eloquio dei e delle parlanti, che possono aver guidato i giudizi.

Riferimenti bibliografici

Calamai, S. (2015). Between linguistics and social psychology of language: the perception of non-native accents. *Studi e Saggi Linguistici*, 3, 289–308.

Calamai, S., Nodari, R., & Galatà, V. (2020). "Fregati dall'accento!". Lo stereotipo etnico e linguistico nei contesti scolastici. *Italiano LinguaDue*, 12 (1), 430-458.

Dragojevic, M. & Goatley-Soan, S. (2022). The Verbal-Guise Technique. In R. Kircher & L. Zipp (eds.), *Research Methods in Language Attitudes* (pp. 203-218). Cambridge: Cambridge University Press.

Fiske S.T., Cuddy A.J.C., Glick P. & Xu J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, 82(6), 878-902.

Levon, E. (2015). Integrating intersectionality in language, gender, and sexuality research. *Language and Linguistics Compass*, 9(7), 295-308.

MacFarlane A.E. & Stuart-Smith J. (2012). 'One of them sounds sort of Glasgow Uni-ish'. Social judgements and fine phonetic variation in Glasgow. *Lingua*, 122(7), 764-778.

Masullo, C., & Meluzzi, C. (2020). Stereotipi, lingua e società. Un'analisi dei pregiudizi legati ad accenti italiani e stranieri. In A. De Meo & F. M. Dovetto (Eds*.), La comunicazione parlata. Spoken Communication* (pp. 153– 170). Roma: Aracne.

Nodari, R. (2021), È possibile una linguistica intersezionale in Italia? Breve storia di un termine militante all'interno degli studi linguistici italiani. *gender / sexuality / Italy*, 8, 33-51.

Pantos, A.J. & Perkins, A.W. (2013), Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology*, 32, 3-20.

Rakić, T., Steffens, M.C. & Mummendey, A. (2011). When it matters how you pronounce it. The influence of regional accents on job interview outcome. *British Journal of Psychology*. 102, 868–883.

Forensic Automatic Speaker Recognition based on Emphasized Channel Attention, Propagation and Aggregation in Time Delay Neural Network

Francesco Sigona, Giuseppe Vitolo and Mirko Grimaldi
Università del Salento

La comparazione forense della voce (o riconoscimento forense del parlante, *FSR*) gioca un ruolo estremamente importante nelle indagini e nella risoluzione di casi giudiziari, quando bisogna attribuire una registrazione vocale anonima ad un determinato parlante di identità nota, con un certo grado di affidabilità. Il riconoscimento vocale automatico del parlante trova numerose applicazioni anche in ambiti diversi da quello forense, in particolare nei contesti in cui è necessaria una qualche forma di autenticazione utente per l'accesso a servizi info-telematici, o a dispositivi locali (ad esempio, *smartphone*), o ambienti fisici riservati e così via, in diversi settori applicativi (come quello bancario- finanziario e sanitario [1]). In tali applicazioni, generalmente il sistema calcola un modello matematico del campione vocale da autenticare e confronta tale modello con uno o più modelli pre-arruolati di voci di utenti autorizzati. Il risultato di tale confronto è generalmente un valore numerico (*score*) che esprime la similarità tra il modello in ingresso e quello arruolato, e che viene poi confrontato con un valore di soglia opportunamente pre-calcolato per rispondere a determinati requisiti in termini di probabilità di errore della decisione, a loro volta scelti a seconda della criticità dell'applicazione: ovvero, falsa identificazione di un campione vocale estraneo, e falso rifiuto di un campione vocale legittimo.

Il *FSR* è un campo di applicazione particolarmente delicato, con requisiti più stringenti rispetto alle summenzionate applicazioni: infatti, è stata più volte evidenziata la necessità che la metodologia di comparazione si basi su misurazioni quantitative, conduca a tassi di errore accettabili, e soprattutto che sia riproducibile e validata dalla comunità scientifica [2, 3]. La metodologia comunemente accettata in letteratura si basa su approccio di tipo Bayesiano (recentemente raccomandato anche dall'*European Network of Forensic Science Institutes* [4]. Pertanto, l'*FSR* non si configura come test di decisione binario (identificato/rifiutato), ma si esprime attraverso un rapporto di verosimiglianza (*Likelihood Ratio, LR* [5, 6]) tra l'ipotesi accusatoria—in cui prevale il grado di similarità tra i due campioni posti in comparazione— e l'ipotesi difensiva—in cui prevale il grado di tipicità delle voci, cioè di similarità con un insieme di campioni tratti da una opportuna popolazione di riferimento. Le *performances* di un siffatto sistema di *FSR* possono essere sinteticamente rappresentate da metriche come l'*Equal Error Rate* (*EER),* rappresentativo della capacità del sistema di discriminare tra due campioni vocali, e il *Log-Likelihood-Ratio-Cost* (*CLLR*), rappresentativo del grado di accuratezza che il sistema può fornire in generale in determinate condizioni operative. Molteplici sono i fattori che possono influenzare le *performances* di un sistema di *FSR*. A parità di scelta degli algoritmi e relativi parametri, un ruolo decisivo è svolto dalle condizioni operative della comparazione, tra cui: (1) l'effetto del canale di trasmissione/registrazione del campione vocale, e quindi di alterazione dello spettro di frequenze della voce; (2) il livello e il tipo di rumore presente nelle registrazioni; (3) la durata delle stesse; (4) alcuni fattori socio-linguistici quali il sesso e la varietà linguistica dei parlanti, Si tratta di condizioni che si riferiscono alla correttezza logico-metodologica della comparazione (prima ancora che ai risultati delle metriche di *performances*).

In questo contributo, presentiamo una tra le prime implementazioni, e sperimentazioni, di un sistema di *FSR* automatico (*Forensic Automatic Speaker Recognition*, *FASR*) basato su *Emphasized Channel Attention, Propagation and Aggregation in Time Delay Neural Network* (*ECAPA-TDNN*). Si tratta di un recentissimo modello di rete neurale che, nelle applicazioni non forensi ha già mostrato migliori performance in termini di tassi di errore di riconoscimento, rispetto ai modelli precedenti [7]. Il modello di rete neurale utilizzato è già pre-addestrato con i dataset *Voxceleb1* e *Voxceleb2* [8, 9] e, nell'ambito della sperimentazione condotta, è essenzialmente utilizzato per trasformare i campioni vocali da comparare e quelli del campione della popolazione di riferimento in opportuni vettori di *features* (*embeddings*). Una adeguata metrica di similarità (*cosine-distance*), e un algoritmo di normalizzazione (*adaptive simmetric*

norm) consentono poi di ricavare un valore di *score* per ogni coppia di *embeddings*: il *LR* si ottiene infine come rapporto tra il valore della funzione densità di probabilità (*pdf*) che modella lo *score* nell'ipotesi accusatoria—stimata attraverso confronti multipli di campioni vocali provenienti da uno stesso parlante, per vari parlanti del campione della popolazione di riferimento—e il valore della *pdf* che modella lo *score* nell'ipotesi difensiva—stimata attraverso confronti multipli tra il campione vocale anonimo e i campioni vocali dei parlanti nel campione della popolazione di riferimento.

In questo caso, il sistema è sottoposto a vari test di comparazione, utilizzando un database di 127 voci adulte maschili in lingua italiana (estratte da narrazioni di *audiobook*), al fine di determinarne le prestazioni al variare dei parametri legati alla qualità dei campioni. Fissate le condizioni operative (in particolare, durata e rumorosità dei campioni vocali della voce anonima e della voce nota, di volta in volta estratti dal database), ciascun campione adattato alle condizioni della voce anonima è comparato con gli altri campioni adattati alle condizioni della voce nota. Sono state così ottenute le metriche di *performances* per le condizioni operative prefissate, per le quali ci si aspetta un progressivo miglioramento all'aumentare della durata dei campioni vocali e al diminuire della loro rumorosità. Nella Fig. 1 sono riportati alcuni risultati preliminari, sull'andamento dell'*EER* e del *CLLR*, relativamente ad uno scenario di riferimento in cui vari campioni vocali sono stati convertiti in segnali di qualità telefonica (banda compresa tra 300-3400 Hz, campionamento a 8KHz): quelli rappresentativi della voce nota hanno durata variabile tra 10 e 60 secondi e sono debolmente contaminati con rumore additivo gaussiano bianco (*AWGN*), con rapporto segnale/rumore (*SNR)* pari a 24 dB, mentre i vari campioni vocali di identità anonima hanno durate variabili comprese tra 5 e 60 secondi, ed *SNR* anch'esso variabile tra 6 dB (condizioni molto critiche) e 24 dB.

Le sperimentazioni condotte in [7] con il dataset *Voxceleb1* attestano un *EER* minimo tra 0,87% e 1,01%, a seconda della complessità della rete (non sono riportati valori di *CLLR*). La sperimentazione del presente lavoro mira ad ottenere valori analoghi in situazioni di bassa rumorosità: in effetti, i risultati preliminari confermano ottime prestazioni in condizioni operative non critiche, con *EER* dell'ordine dell'1%. Risultano preliminarmente confermate anche le ipotesi sull'andamento delle *performances* in varie condizioni di test. Infatti, vi è un tendenziale miglioramento di *EER* e *CLLR* all'aumentare della durata del campione anonimo, passando da 5 secondi a 60 secondi. Quando il campione anonimo ha elevati *SNR*, tale miglioramento avviene repentinamente, già a 10 secondi; oppure a 20 secondi quando la durata del campione di voce nota è di soli 10 secondi. In generale, l'aumento della durata del campione di voce nota tende a migliorare le *performance* quando le condizioni del campione anonimo non sono ottimali (bassa durata, basso *SNR*).

Analisi più approfondite, attualmente in corso, consentiranno di caratterizzare il sistema in ulteriori condizioni operative, nonché in funzione della scelta del campione della popolazione di riferimento, che si rivela cruciale per validare l'affidabilità del sistema [10, 11]. In lavori futuri saranno considerate anche sperimentazioni con voci di sesso femminile, nonché varietà dialettali e lingue straniere.

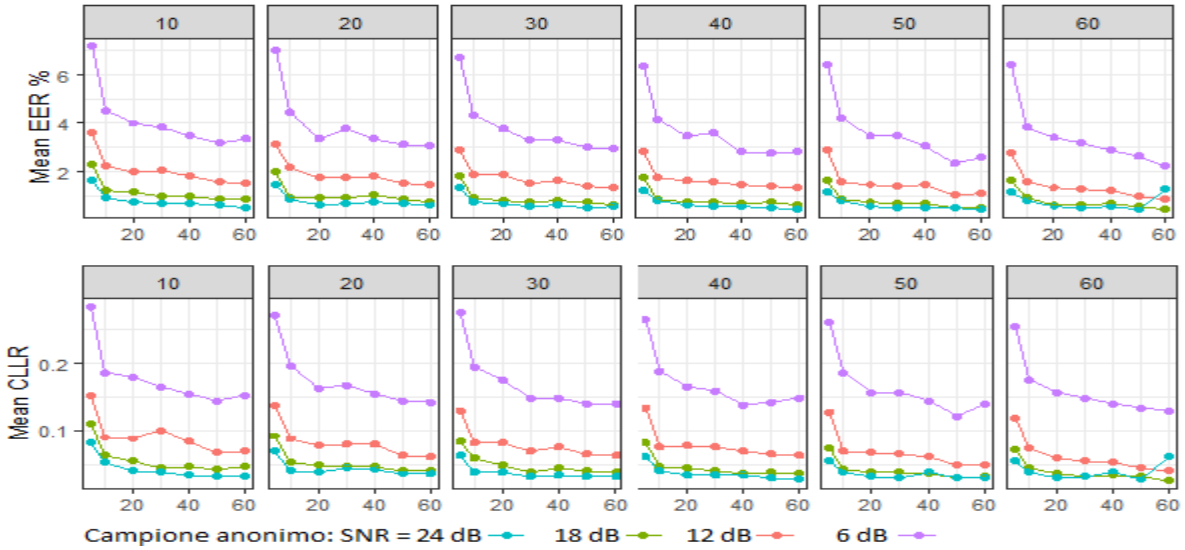


Figura 1: andamenti dei valori medi di EER *(in alto) e di* CLLR *(in basso), nelle comparazioni tra il generico campione di voce nota, avente* SNR *= 24 dB (contaminato con* AWGN*) e durata di 10, 20, … 60 secondi (in testa ad ogni grafico), ed il generico campione di voce anonima, avente* SNR *di 24 (in celeste), 18 (in verde), 12 (in rosso) e 6 dB (in lilla), e durate di 5, 10, 20, …, 60 secondi (sull'asse orizzontale di ogni grafico). I campioni vocali sono di qualità telefonica.*

Riferimenti bibliografici

[1] Sigona F., *Voice Biometrics Technologies and Applications for Healthcare: an overview*, in *JDREAM. Journal of interDisciplinary REsearch Applied to Medicine*, Vol. 2, Issue 1, 2018, ESE - Salento University Publishing: 5-15.

[2] Enzinger E., Morrison G.S. & Ochoa F., *A demonstration of the application of the new paradigm for the evaluation of forensic evidence under conditions reflecting those of a real forensic voice-comparison case*, in *Science & Justice*, 56, 1 (2016): 42–57.

[3] Morrison G.S., *Distinguishing between forensic science and forensic pseudoscience: testing of validity and reliability, and approaches to forensic voice comparison*, *Science & Justice*, 54 (2014): 245–256.

[4] Drygajlo,A., Jessen M., Gfroerer S., Wagner I., Vermeulen J., & Niemi, T. (2016). *Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition including Guidance on the Conduct of Proficiency Testing and Collaborative Exercises*. https://enfsi.eu/

[5] Morrison G.S., *Forensic voice comparison and the paradigm shift*, in *Science & Justice*, 49 (2009): 298–308

[6] Morrison G.S., *Measuring the validity and reliability of forensic likelihood-ratio systems*, in *Science & Justice*, 51 (2011):91-98 [7] Desplanques, B., Thienpondt, J., & Demuynck, K. (2020). *ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification*. Interspeech.

[8] Nagrani A., Chung J.S., Xie W. & Zisserman A., *Voxceleb: Large-scale speaker verification in the wild*, in *Comput. Speech Lang.*, 60, 2020

[9] Speechbrain. Pretrained ECAPA-TDNN using voxceleb1 and voxceleb2. https://huggingface.co/speechbrain/spkrec-ecapa- voxceleb

[10] Robertson B., & Vignaux, G. A., (1995), *Interpreting evidence*, Chichester: Wiley.

[11] Rose, P., (2006), Technical Forensic Speaker Recognition: Evaluation, types and testing of evidence, *Computer Speech and Language* 20, pp. 159–191.

Automatic speech recognition for low-resource languages: exploring transfer learning from related languages

Eleonora Rodegher[1], Alessio Brutti[2], Daniele Falavigna[2]
[1]*Libera Università di Bolzano,* [2]*Fondazione Bruno Kessler*

Introduction

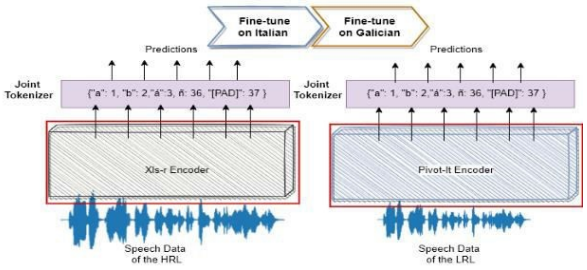
Automatic speech recognition (ASR) systems typically require huge amounts of data, which are not always available for languages lacking digital resources. Recently, new approaches have been investigated to overcome the scarcity of labelled data [1]. One method is to adapt models pre-trained on vast quantities of data from well-resourced languages to poorly resourced ones. In this work we applied different transfer learning techniques by fine-tuning the pre-trained model wav2vec2-xls-r [2] on different languages. In particular, we considered Italian, Portuguese, Spanish, and two minority languages: Galician and Romansh Vallader.

Proposed Solutions

Transfer learning techniques for ASR aim at transferring the knowledge from high-resource languages to low- resourced ones. These methods consist in firstly pre-training a model on speech from high-resource languages (pivot language) and then fine-tuning all or parts of the model on speech from the low-resourced ones [3]. This permits to reuse the speech representations shared among languages when fine-tuning on the target language. The pre-trained model wav2vec2-xls-r [2] is the large-scale cross-lingual version of wav2vec2.0 [4]. The model was pre-trained on half a million hours of speech from 128 languages, using the self-supervised learning approach. For ASR tasks, a linear layer is added on top of the pre-trained model to predict the output vocabulary using the connectionist temporal classification (CTC) [5]. In our work, we used graphemes as speech units to be predicted by the model. In the experiments we used the open crowd-sourced multi-language dataset Mozilla Common Voice [6]. The corpora include recordings of read sentences and their transcriptions. Common Voice employs crowdsourcing also for data validation, therefore some inaccuracies can be found.

Experiments

The experiments considered different amounts of training data. Results were measured in Character and Word Error Rates. As baselines, we created monolingual models for Italian, Galician, and Romansh Vallader, considering for each language a tokenizer containing all the characters plus special tokens for decoding. Starting from the assumption that training on closely related languages improves the performance of the low-resourced ones [8], we tried alternative approaches involving related languages which are better resourced. The approach adopted implied training in cascade the pre-trained model: firstly, by fine- tuning on the better-resourced language and after, by fine-tuning on the minority one. The union of the characters of the two languages was used to decode the output. Italian was selected as the pivot or bridge language for both Galician and Romansh Vallader. For Galician, we performed the same experiments using Portuguese and Spanish as pivot languages.



Results

Similar trends were observed for the Italian, Galician, and Romansh monolingual models. For Italian, in particular, after fine-tuning the pre-trained model on 8 hours of speech the amount of data had to be doubled to obtain significant improvements. The models derived from related languages displayed data efficiency compared to those trained on the same amount of data but without a pivot language. A significant reduction in CER and WER was observed when fine-tuning on an amount of data in the range of

9 to 30 minutes. Clearly, increasing the number of hours reduces the advantages, performing similarly to the related monolingual models. Figure 1 and 2 show CER and WER for Galician using Italian, Spanish, Portuguese and German as pivot.

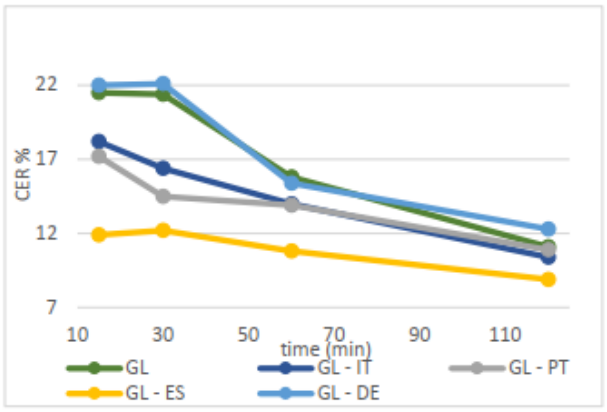


Figure 1: CER: Galician without pivot, with Italian (IT), Portuguese (PT), and Spanish (ES) as pivot, on 100% Galician test set

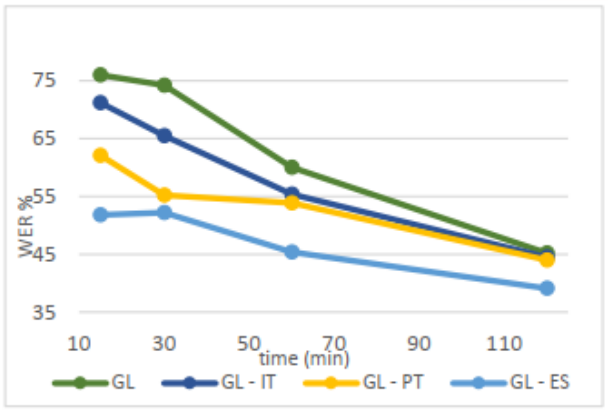


Figure 2: WER: Galician without pivot, with Italian (IT), Portuguese (PT), and Spanish (ES) as pivot, on 100% Galician test set

The best performance is achieved with Spanish, as it shares a large number of graphemes and have high overlap of phonemes. Conversely, German, a language from a different family with different phonology and orthography, brought no advantage, leading to similar performance as the monolingual model. Figure 3 shows the CER for Romansh – Vallader with and without Italian as pivot

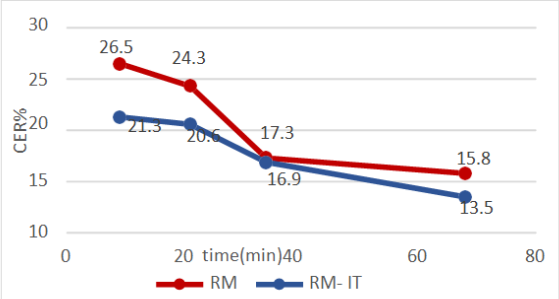


Figure3: CERforRomanshVallader.

Conclusions

This study investigated the use of a pivot language for fine-tuning a large-scale pre-trained model on low- resourced languages. In the case of Galician, we observed the effect of using different romance languages as pivot. The results were confirmed in the case of Romansh Vallader with Italian.

References

[1] Besacier, L., Barnard, E., Karpov, A., & Schultz, T., "Automatic speech recognition for under-resourced languages: A survey." in Speech Communication 2014, 2014, 56, 85–100.

[2] Babu, A., et al.., "XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale", ArXiv:2111.09296, 2021.

[3] Khare, S., et al. "Low Resource ASR: The Surprising Effectiveness of High Resource Transliteration", in Interspeech 2021, 1529–1533.

[4] Schneider, S., Baevski, A., Collobert, R., & Auli, M., "wav2vec: Unsupervised Pre-training for Speech Recognition", ArXiv:1904.05862, 2019

[5] Graves, A., Fernandez, S., Gomez, F., & Schmidhuber, J., "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks", 369–376, 2006.

[7] Koehn, P. "Europarl: A Parallel Corpus for Statistical Machine Translation", Proceedings of Machine Translation Summit X: Papers, 2005, pp. 79–86. 2005.

[6] Ardila, R., et al., "Common Voice: A Massively-Multilingual Speech Corpus", ArXiv:1912.06670, 2020,

[8] Jacobs, C., & Kamper, H., "Multilingual Transfer of Acoustic Word Embeddings Improves When Training on Languages Related to the Target Zero-Resource Language", in Interspeech 2021, 1549–1553.

[9] Chen, S., et al., "WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing", arXiv:2110.13900,202.

Amelia Rosselli and the anguish of breathing: style or pathological speech?

Valentina Colonna[1], Federico Lo Iacono[2], Antonio Romano[3]
[1, 3] LFSAG-Laboratorio di Fonetica Sperimentale "Arturo Genre", Dip. di Lingue e Letterature Straniere e Culture Moderne dell'Università di Torino; [2] Università degli Studi di Bologna

Con questa ricerca si intende discutere un caso tanto particolare quanto interessante di parlato poetico, che presenta alcuni tratti caratteristici che sembrerebbero accomunarlo alla sfera del parlato patologico. Oggetto della nostra analisi sono infatti otto letture poetiche di Amelia Rosselli (1930-1996), una delle poetesse più note della nostra tradizione letteraria. Le interpretazioni selezionate per la ricerca coprono un arco temporale che va dal 1985 (con un primo nucleo di tre poesie) al 1990 (con la lettura integrale del poemetto Impromptu): si tratta, dunque, di registrazioni relative all'ultima fase della breve vita di Rosselli.

L'interesse per uno studio sul parlato patologico della poetessa nasce innanzitutto dalla ricostruzione delle sue vicende cliniche, testimoniate da lei stessa in uno scritto del 1977, e trova negli indici del dato orale analizzati un suo riflesso (Caputo, 2004; Giovannuzzi, 2012; Venturini & De March, 2010; Paris, 2020). Le prime tappe significative nella storia della sua patologia risalgono al 1954, anno in cui venne sottoposta a Roma al primo elettroshock e a vari shock insulinici per curare quello che inizialmente sembrava un esaurimento nervoso o un disturbo del sonno. Sempre nel 1954 ricevette la prima diagnosi clinica di schizofrenia paranoide e numerosi sono gli anni di terapia psicanalitica con diversi periodi di ricovero clinico. Una svolta cruciale nella storia della patologia della poetessa risale al 1969, quando le viene diagnosticato il morbo di Parkinson, o con più precisione "una lesione al sistema extrapiramidale", in seguito a cui è curata con farmaci antispastici, che Rosselli avrebbe continuato a prendere fino alla sua tragica morte per suicido nel 1996. Diversi dubbi sono stati sollevati da critici letterari sulla vera natura della sua malattia, tra cui l'ipotesi di una copertura della sua schizofrenia paranoide con il morbo di Parkinson: l'assunzione dei farmaci antispastici prescritti regolarmente alla poetessa potrebbe spiegarsi infatti come strumento curativo per i probabili effetti dannosi lasciati dai vari elettroshock sul suo sistema muscolare. Nonostante l'oggettiva difficoltà nel ricostruire l'effettiva eziologia della sua disartria, l'analisi della sua lettura ci offre interessanti informazioni sulla natura del suo eloquio e ci consente di valutare anche, nella peculiarità all'interno del panorama italiano, i possibili riflessi prosodici di un sostrato patologico. Particolarmente affascinante è, inoltre, la poliedricità di quest'autrice, in grado di conciliare la dimensione poetica con quella musicologica e una natura plurilinguistica: oltre al suo interesse per la sostanza sonora della poesia, è riconoscibile una convergenza nel suo stile compositivo delle tecniche della musica atonale a lei contemporanea (Rosselli, 1962). Il nostro interesse per questo caso di studio si è sviluppato e ampliato in questi anni, fornendo un primo approccio di indagine in Colonna (2021), da cui sono emersi alcuni primi dati stimolanti, che hanno messo in luce tratti stilistici caratteristici, quali l'impiego di allungamenti di contoidi e un uso sperimentale della voce in relazione al testo, tale da consentirne l'annoveramento tra le voci "sperimentali" nella storia della lettura del Novecento italiano della poesia. Il suo modo di leggere, che assorbe le influenze dello sperimentalismo musicale, si caratterizza per una tendenza alla frammentazione tramite silenzio e rumore enfatizzati dalla microprosodia e da un uso peculiare delle pause. In seguito a questo primo lavoro, in cui è stata impiegata la metodologia VIP-Voices of Italian Poets, basata sull'analisi di una selezione di indici e grazie a un'annotazione manuale su 4 livelli, sono emersi, più dettagliatamente, alcuni tratti di particolare rilievo. Nello specifico, il VIP-Radar (VIP-R), con al suo interno 20 indici rappresentativi dello stile di lettura poetico, ha offerto un primo riassunto descrittivo del suo stile distinto (v. Colonna, 2022).

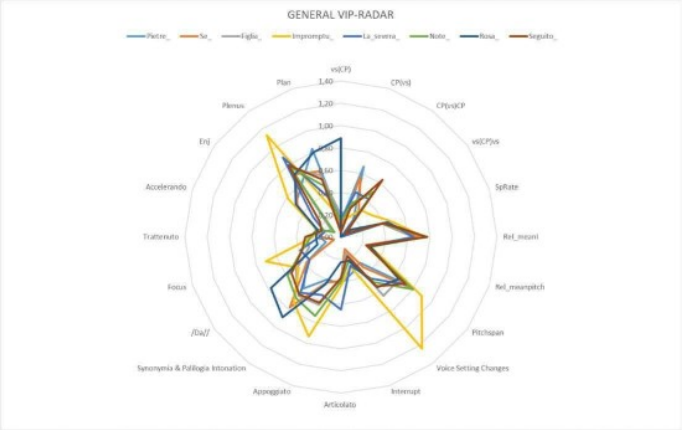


Figura 1 General-VIP-Radar di otto letture di Amelia Rosselli

A partire da questo studio, nel nostro lavoro sono state confrontate 8 registrazioni tramite un *General-VIP- Radar* (Fig. 1), realizzato in seguito ad annotazione, estrazione e percezione dei dati a cura di due degli autori. Il grafico consente di notare un'ampia variazione tra le letture della poetessa, con uno scostamento di un caso rispetto agli altri (*Impromptu*), ma è possibile al contempo individuare diversi punti di convergenza nelle scelte acustiche, stilistiche e, in parte, organizzative. Soffermandoci su alcuni aspetti caratteristici, si nota la presenza costante in tutte le letture di *Interrupt* (indice di uno stile tendente alla frammentazione prosodica), un ampio *Pitchspan*, mentre molto mutevole risulta l'uso delle pause, come mostra il *Plenus* (rapporto tra durata totale del parlato e pausale) e l'eterogenea resa dell'inarcatura. Soffermandoci più dettagliatamente sull'aspetto pausale, alla luce di studi rivolti al silenzio in ambito patologico (cfr. Dovetto, 2013 e Nodari & Calamai, 2021), e considerata inoltre la particolarità stilistica di Rosselli, si è successivamente condotta una ricerca sulle pause impiegate dall'autrice nelle sue letture e, dopo, si è adottato un approccio comparativo sull'uso pausale in 10 diverse interpretazioni di uno stesso testo poetico a cura di diversi poeti contemporanei e dell'autrice.

Alla luce dei protocolli di annotazione in uso nel progetto CLIPS (cf. Savy, 2006; Savy, 2007) e degli studi condotti presso il LFSAG (cf. Romano, 2008; De Iacovo *et al.*, 2020) e, in particolare, della distinzione tra pause lunghe e pause brevi, la categorizzazione su cui ci siamo basati è la seguente: pause con laringalizzazione (<?pbb> <pbb?> <?pbb?>), pause brevi (<pb>), pause lunghe (<pl>), pause molto lunghe (<pl>), inspirazioni (<insp>). Emerge dalla prima parte di studio un frequente uso di colpi di glottide e <pb> che si alternano a <pl>: nel dettaglio, le <(?)pbb(?)> e <pb> si individuano con una media intorno a 370 ms e una deviazione standard media di circa 200, mentre le <pl> con una media superiore a 1000 ms e una deviazione standard media inferiore a 170. Questi dati ci confermano che il nostro sistema di annotazione può risultare affidabile, con una soglia che, intorno ai 400 ms, determina uno scollamento nell'inventario pausale dell'autrice. La non trascurabile presenza di colpi di glottide su <pbb>, assimilati in questa quantificazione al calcolo delle <pb>, così come l'ulteriore presenza di <insp> (da cui è minima la variazione), oltre a caratterizzarsi come tratto stilistico di questo stile di lettura poetica lascia aperta un'ipotesi di riflesso di una patologia presente. Emerge, oltre alla netta variazione, con usi maggiori o minori di pause, la possibilità di margini di ulteriore classificazione, in cui la soglia si colloca su un livello più alto e, in generale, nel caso di tutti gli altri, con valori simili ma grado di accuratezza minore.

In conclusione, possiamo dire che la descrizione tramite il VIP-R e i dati emersi dall'analisi pausale, unitamente al confronto con le voci contemporanee, mostrano la prosodia di Rosselli con tratti specifici, nettamente distinti da quelli di altri interpreti, le cui produzioni andrebbero valutate anche in considerazione di una diversa velocità d'eloquio. Con questo studio non desideriamo fornire una diagnosi di parlato patologico o un suo avvaloramento, bensì aprire una questione ancora inesplorata, approfondendo con l'indagine fonetica il dato acustico di una delle voci più rappresentative del Novecento italiano, da una nuova prospettiva, che lascia supporre uno stile prosodico originale, caratterizzato da indici riflessi di tratti patologici in un parlato poetico inconfondibile.

Bibliografia

Caputo F. (a cura di). (2004). *Una scrittura plurale. Saggi e interventi critici*. Novara: Interlinea.

Colonna V. (2021). *"Voices of Italian Poets". Analisi fonetica e storia della lettura della poesia italiana dagli anni Sessanta a oggi.* Tesi di Dottorato inedita. Università di Genova-Università di Torino. Tutor: Prof. Antonio Romano.

Colonna V. (2022). *Voices of Italian Poets. Storia e analisi fonetica della lettura della poesia italiana del Novecento*. Alessandria: Edizioni Dell'Orso.

De Iacovo V., Colonna V., Romano A., (2020). "La pausazione". *Bollettino del Laboratorio di Fonetica Sperimentale «Arturo Genre»*, n. 5, 41-48.

Dovetto F. M. & Gemelli M. (a cura di). (2013). *Il parlar matto. Schizofrenia tra fenomenologia e linguistica: il corpus CIPPS*. Roma: Aracne.

Giovannuzzi S. (a cura di). (2012). *Rosselli. L'opera poetica.* Milano: Mondadori.

Nodari R. & Calami S. "I silenzi dei matti. Gli spazi 'vuoti' del parlato nell'archivio sonoro di Anna Maria Bruzzone" [Università di Zurigo, 04/02/2021 – 05/02/2021] XVII Convegno AISV, presentazione orale accettata.

Paris R. (2020). *Miss Rosselli*. Vicenza: Neri Pozza.

Romano, A. (2008). *Inventari sonori delle lingue: elementi descrittivi di sistemi e processi di variazione segmentali e sovrasegmentali*. Torino: Edizioni dell'Orso.

Rosselli A. (1962). "Spazi metrici". In: Rosselli A. *Le poesie*. Milano: Garzanti, 1997, 337-342.

Savy R. (2006). "Specifiche per la trascrizione ortografica annotata dei testi raccolti". In: Albano Leoni F. & Giordano S. (a cura di). *Italiano parlato. Analisi di un dialogo*. Napoli: Liguori,1-37.

Savy R. *et alii* (2006). "Specifiche per la trascrizione e l'etichettatura dei livelli segmentali in CLIPS», in CLIPS – Corpora e Lessici di Italiano Parlato e Scritto (www.clips.unina.it).

Venturini M. & De March S. (2010). *Amelia Rosselli. È vostra la vita che ho perso. Conversazioni e interviste 1964-1995*. Firenze: Le Lettere.

IL PARLATO IN AMBITO MEDICO:
Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD:
Linguistic analysis, technological applications and clinical tools

FRIDAY 17 FEBRUARY
KEYNOTE SPEECS AND ORAL PRESENTATIONS

Speech and quality of life in patients with Parkinson's disease

Antonio Schindler
University of Milan

Background: Dysarthria occurs in most patients with Idiopathic Parkinson Disease (IPD). Patients with IPD often have depression and reduced quality of life (QOL). While dysarthria, depression and QOL in patients with IPD have been studied, the association among these variables has never been investigated. The study aims to investigate the association among measures of dysarthria, IPD disease severity, depression, and QOL in patients with IPD, and to identify existing clusters among the investigated variables.

Methods: Seventy-five consecutive patients with IPD were recruited in this cross-sectional study. Each patient was assessed through Robertson Dysarthria Profile, Therapy Outcome Measure Scale (TOM), Movement Disorders Society-Unified Parkinson's Disease Rating Scale (UPDRS), QOL for the dysarthric speaker (QOL-DyS) questionnaire, Short Form-36 (SF-36) questionnaire, and Beck Depression Inventory-II (BDI), to measure respectively severity of dysarthria, IPD, dysarthria-related and generic QOL, and depression. Patients with a TOM score ≤ 4 were considered dysarthric. Principal Component Analysis, using a Biplot, was used to visualize the pattern of association of the total scores of UPDRS, QOL-DyS, SF-36, and BDI together with the clusters of patients. Spearman's correlation coefficient was used to quantify the correlation between the questionnaires' score. Cluster analysis was performed using all the 4 principal components and a hierarchical algorithm.

Results: Dysarthria occurred in 47/75 (63%) patients. The UPDRS and the SF-36 were negatively associated ($r=-0.42$, $P<0.0001$), while there was no significant correlation between the QOL-Dys and the UPDRS ($r=0.108$, $P=0.355$) or the SF-36 ($r=-0.151$, $P=0.195$). Four clusters of patients were identified: 1st) patients with severe IPD, dysarthria in 50% of the cases, poor SF-36, high QOL-DyS; 2nd) dysarthric patients with severe IPD, depression, reduced QOL (SF-36 and QOL-DyS); 3rd) dysarthric patients with mild-moderate IPD, good SF-36, poor QOL-DyS; 4th) non-dysarthric patients with mild IPD and satisfactory QOL (SF-36 and QOL-DyS).

Conclusions: Dysarthria is a common disorder in patients with IPD, and it is characterized by different profiles of QOL and depression. The QOL-DyS might be useful to identify IPD patients with dysarthria and reduced dysarthria related QOL.

Disfluencies in Parkinson's Disease. A study on Italian early-stage patients

Marta Maffia[1], Loredana Schettino[2], Rosa De Micco[3], Alessandro Tessitore[3]
[1]Università degli Studi di Napoli "L'Orientale" [2]Università degli Studi di Napoli "Federico II"
[3]Università degli Studi della Campania "Luigi Vanvitelli"

Parkinson's Disease (PD) is a neurodegenerative disorder affecting more than 2-3% of the population over 65 years of age and consisting in a deterioration of dopamine-producing neurons in the basal ganglia (de Lau & Breteler, 2006). One of its early symptoms is hypokinetic dysarthria, caused by poor activation and coordination of the muscles that control speech production and including a range of speech and voice impairments: reduced voice intensity, significantly narrower tonal range (monopitch), monoloudness, increased voice nasality, increased acoustic noise, imprecise consonantal articulation and reduced vowel space area, vocal tremor, harsh and breathy voice quality, impaired speech rate and rhythm, longer silent pauses (see, among others, Darley et al. 1969, Goberman & Coelho, 2005, Skodda et al. 2011, Gili Fivela et al. 2014).

Although parkinsonian speech is usually defined as "disfluent", still the specific characteristics of disrupted PD speech have not been well documented (Goberman & Blomgren, 2003). Recent studies have focused on stuttering-like disfluencies in mild-to-severe PD patients, examined in different speech styles and compared to those produced by healthy speakers and by individuals with developmental stuttering (Juste et al. 2008; Goberman et al. 2010). Results report significantly greater disfluency percentages in PD patients than in the healthy control group, especially in the case of within-words disfluencies (stuttering occurring on part of a word; syllable repetitions, incomplete syllable repetitions, inaudible and audible fixed postures) and in the monologue speech task, thus supporting the relationship between stuttering and the basal ganglia/dopamine system. In another research, no differentiation in the speech disfluency severity between different speech styles produced by PD patients was found but a positive correlation between the frequency of disfluencies and the duration of the disease (Półrola & Góral-Półrola, 2015). The effect of Levodopa medication on disfluencies has also been observed, trying to support the so-called *excess dopamine hypothesis,* according to which increased levels of dopamine should lead to the development of stuttering in PD, with controversial results (Im et al. 2019).

More generally, speech errors and disfluencies have been demonstrated to be strong predictors of cognitive impairment in other pathologies, such as dementia and Alzheimer's disease (König et al. 2015; Wang et al. 2019). In fact, the very first classifications of the phenomena ascribable to the category of "disfluencies" (Johnson, 1961) emerged in the strive to distinguish between disfluencies in typical and atypical speech (Wingate, 1987), given the acknowledgement that typical speech also commonly includes phenomena like repetitions, hesitant pauses, self-repairs as useful tool for speakers to manage and monitor the own speech production.

In this light, the analysis of differences between the patterns of disfluency phenomena in PD and in typical speech can offer a unique opportunity to distinguish and examine acquired disfluencies of known neurogenic origin and associated with cognitive, linguistic, and motor deficits resulting from the damages in central nervous system. So, this study aims at describing the occurrences of speech disfluency phenomena that characterize the spoken production of Italian subjects with early-stage PD.

The data for the present research were collected from 40 Italian native speakers residing in Campania region: 20 participants with idiopathic non-demented PD were examined (12 males, 8 females; 51–81 years of age, M=64), along with 20 age-matched controls (8 males, 12 females; 54–77 years of age, M=65). The Patients have been recruited from a cohort of early PD with no history of previous language and speech disorder. These patients are prospectively followed and underwent an extensive motor and non-motor clinical assessments every 12 months, with a clinical follow-up every 6 months. They were all recorded while performing two tasks: a reading task and a monologue. Sociolinguistic information on each speaker were also collected. This study is based on the analysis of disfluency patterns annotated in the collected monologic speech using the ELAN software (Sloetjes & Wittenburg, 2008). Disfluencies are defined as the

linguistic tools at speakers' disposal to monitor and effectively manage the online processes of speech planning, coding, articulation, and reception (Levelt, 1989). According to the context of occurrence, each disfluent phenomenon is identified and annotated on three main levels (Schettino et al., 2021). On the first level, disfluencies' macro-structure is labeled: the region to be repaired (Reparandum, RM), the repaired one (Reparans, RS), the one where the delay occurs (Interregnum, IM). The second level is for the micro-structure, here disfluent items are categorized as Insertion, Deletion, Substitution, Repetition, Silent Pause, Lengthening, Filled Pause, Lexicalized Filled Pause. On the third level, each item is assigned its macro-function, Backward-Looking, when retracing and altering already uttered speech, or Forward-Looking, when gaining time for speech planning processes. The analysis concerns the comparison of PD and Health Control (HC) speech with reference to the following parameters: the frequency of disfluent items and of their main contextual function; the syntactic positioning of the items (within words, within phrases, between phrases, between clauses); the duration of silent pauses, filled pauses and lengthenings. The statistical significance of the results is tested with Generalised Linear Mixed Models, considering the health condition (PD or HC) as dependent variable, the subjects' sociolinguistic and clinical data as independent variables, and subjects as random effect.

Preliminary results from a pilot based on 8 speakers (4 PD and 4 HP) show that the occurrence of disfluencies is on average higher in HC speech (30 per minute) rather than in PD speech (22 per minute). This difference specifically concerns forward looking disfluencies (HC: 25 items/minute vs. PD: 17 items/minute) rather than backward looking disfluencies (about 5 items/minute in both conditions). As for the disfluent item positioning, a prominent trend concerns the higher number of disfluent items occurring after word fragments, that is interrupting words, in patients' speech. Considering forward looking disfluencies, it is worth noticing that, in PD speech, lengthenings are more frequent and silent pauses are less frequent than in HC speech. Finally, as for duration, filled pauses, silent pauses and lengthenings are on average longer in PD speech than in HC (mean durations in PD vs. HC speech: filled pauses, 506.6 ms vs. 333.2 ms; silent pauses, 616.5 ms vs. 241.8 ms; lengthenings, 366.6 ms vs. 265.3 ms). However, the durations of prolongations and silent pauses in PD show greater variability.

These preliminary results highlight the relevance of considering specific patterns of disfluency phenomena rather than just frequency of occurrence for a better understanding of the features characterizing PD speech.

References

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Cluster of deviant speech dimension in the dysarthrias. Journal of Speech and Hearing Research 12(3), 462-469.

de Lau, L. M. & Breteler, M. M. B. (2006). Epidemiology of Parkinson's disease. The Lancet. Neurology 5(6), 525-535.

Fivela, B. G., Iraci, M., Sallustio, V., Grimaldi, M., Zmarich, C., & Patrocinio, D. (2014). Italian vowel and consonant (co) articulation in Parkinson's Disease: Extreme or reduced articulatory variability?. 10th ISSP, May 5-8, Cologne, Germany, Proceedings, 146-149.

Goberman, A. M., & Blomgren, M. (2003). Parkinsonian speech disfluencies: effects of L-dopa-related fluctuations. Journal of fluency disorders, 28(1), 55-70.

Goberman, A. M., Blomgren, M., & Metzger, E. (2010). Characteristics of speech disfluency in Parkinson disease. Journal of Neurolinguistics, 23(5), 470-478.

Goberman, A. M. & Coelho, C. A. (2005). Prosodic characteristics of Parkinsonian speech: The effect of levodopa-based medication. Journal of Medical Speech-Language Pathology 13, 51-68.

Im, H., Adams, S., Abeyesekera, A., Pieterman, M., Gilmore, G., & Jog, M. (2019). Effect of Levodopa on speech dysfluency in Parkinson's disease. Movement disorders clinical practice, 6(2), 150-154.

Johnson, W. (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. The Journal of speech and hearing disorders, (Suppl 7), 1–20.

Juste, F. S., Sassi, F. C., Costa, J. B., & de Andrade, C. R. F. (2018). Frequency of speech disruptions in Parkinson's Disease and developmental stuttering: A comparison among speech tasks. Plos one, 13(6), e0199054.

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Der-reumaux, A. et al., (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's Disease, Alzheimer's & Dementia:Diagnosis, Assessment & Disease Monitoring, 1 (1), 112–124.

Levelt, W.J. (1989). Speaking: From intention to articulation. Cambridge, MIT Press.
Półrola, P. J., & Góral-Półrola, J. (2015). Speech disfluencies in Parkinson's disease. Medical Studies/Studia Medyczne, 31(4), 267-270.
Schettino, L., Betz, S., Cutugno, F., Wagner, P. (2021). Hesitations and individual variability in Italian tourist guides' speech. In C. Bernardasci, D. Dipino, D. Garassino, S. Negrinelli, E. Pellegrino e S. Schmid (Eds.), XVII Convegno Nazionale AISV. Studi AISV 8, 243-262.
Sloetjes, H., & Wittenburg, P. 2008. Annotation by category-ELAN and ISO DCR. Proceeding of the 6th international Conference LREC 2008. 816-820.
Skodda, S., Visser, W. & Schlegel, U. (2011). Vowel articulation in Parkinson's disease. Journal of voice: official journal of the Voice Foundation 25(4). 467-472.
Wang, T., Lian, C., Pan, J., Yan, Q., Zhu, F., Ng, M. L. et al. (2019). Towards the speech features of mild cognitive impairment: Universal evidence from structured and unstructured connected speech of Chinese. INTERSPEECH, 3880–3884.
Wingate, M. E. (1987). Fluency and disfluency; illusion and identification. Journal of Fluency Disorders, 12 (2), 79–101.

Linguistic proficiency, communicative attitude and severity in a sample of pre-school children who stutter

Claudio Zmarich Italy[1], Simona Bernardini[2], Sabrina Bonichini[3], Mirko Bonotto[3], Federica Chiari[3], Erika Ganassin[3], Caterina Pisciotta[4], Fabiola Vullo[3]
[1]CNR-ISTC, [2]ABC BALBUZIE, [3]University of Padova, [4]Logopedista Freelance

Introduction: Although the International Classification of Diseases has currently arrived at the 11th edition (ICD-WHO, 2021), we still prefer the definition of stuttering of the 9th edition: "disorders in the rhythm of speech, in which the individual knows precisely what he wishes to say, but at the time is unable to say it because of an involuntary, repetitive prolongation or cessation of a sound" (ICD-WHO, 1977, p. 202).

Research has established that stuttering typically starts in preschool years (Yairi & Ambrose, 2015) and from the beginning children who stutter (CWS) can develop a negative speech-related attitude, including viewing themselves as a bad communicator, evaluating speech as difficult, and being apprehensive in speaking situations. Consequently, communication attitude (CA) tends to be more negative in CWS compared to children who do not stutter (CWNS) (Guttormsen, Kefalianos & Naess, 2015). We know also that a negative CA can be associated with other linguistic disorders in CWNS (Havstam et al. 2011; McCormack, McLeod & Crowe, 2019). It is also known that most CWS show subtle differences in linguistic skills compared to CWNS (Ntourou, Conture & Lipsey, 2011; Bernstein-Ratner & Brundage, 2022, but see Hollister, Van Horn & Zebrowski, 2017 and the review by Nippold, 2019, for a contrary position). Adding to this, surveys based on questionnaires compiled by Speech and Language Pathologists (SLPs) showed that CWS suffer from a higher rate of comorbidity with other speech and language disorders than CWNS (Arndt

& Healey, 2001; Blood, Ridenour, Quallas & Hammer, 2003). Finally, according to some authors, a negative CA increases together with both severity (Beilby, Byrnes & Yaruss, 2012; Kawai, Healey, Nagasawa & Vanryckeghem, 2012) and time elapsed from the onset (Guttormsen et al. 2015).

Aims: This study aims at investigating possible correlations among stuttering severity, CA, and linguistic skills, in preschool children. Therefore, it is assumed that CWS could develop a negative CA not only because of fluency problems but as a result of other speech and language difficulties they might have, although not necessarily of clinical relevance, and that negative CA could increase together with age and the amount of time elapsed from the stuttering onset.

Method: The study involved 18 children (15 males, 3 females) aged between 4 years, zero months, and 6 years, 11 months, diagnosed as stutterers but not yet treated. They were Italian monolinguals, living in Veneto, without other certified concomitant disorders. The SSI-4 (Riley, 2009) was used in order to assess stuttering severity, the BVL 4-12 (a battery with wide coverage, Marini et al., 2015), for language and speech skills assessment in production and perception, and the KiddyCAT (Vanryckeghem & Brutten, 2007; for the Italian adaptation, see Vanryckeghem & Brutten, 2022), for the evaluation of the CA. In addition, parents completed a questionnaire about the subject's medical history and family information. The tests were administered to each child through two meetings lasting about 2 hours, at the children's home or at the study of the speech therapist. All meetings were audio- and video-recorded. As a term of comparison, we used the normative samples contained in the Italian adaptation of the KiddyCAT (Vanryckeghem et al., 2022) and in the BVL. As to severity level, there were 3 severe, 9 moderate 4 mild, and 2 very mild subjects.

Results: The average score of the CWS subjects on the KiddyCAT test is 2.82 (DS=2.39; CWS' mean in Vanryckeghem & Brutten, 2022: 4.37; SD: 2.56) and it is not significantly different from the mean score of CWNS (0.72; SD: 1.19). Only 4 children out of 18 have obtained a score that is indicative of a negative CA. Regarding speech and language skills, all cautions were taken in order to exclude fluency difficulties from the calculation of the scoring. In all the tests of the BVL battery, the majority of the participants (around 70%) have obtained scores that fall within the average expected for their age (between -1 SD and

+1 SD), while around 20 % scored lower and 10 % scored higher than those of normative sample. The sections of BVL in which a substantial percentage of children (>30%) have ranked below 1 SD are Naming, Narrative Fluency (a composite assessment involving the description of a short story elicited through a series of illustrations), Grammatical decision, and Repetition (we summed up the results of the three repetition tasks of the word, non-word, and sentences) and a mean of around half of these subjects scored below two SD, which is considered of clinical interest. Finally, the separate by-pair correlation analyses between CA, the time elapsed from the onset, stuttering severity, and speech and language skills did not produce any significant results.

Discussion and Conclusions: Our results turn out to be in line with the current literature, showing a substantial similarity between the speech and language skills of stuttering and nonstuttering peers, with the possible exception of the very bad performances (<-2 SD) in the articulation task and the word and non-word repetition tasks, which are consistent with the evidence of the greater comorbidity of stuttering with Speech Sound Disorders (SSD). Interestingly, Gerwin et al. (2022) found that CWS with SSD performs significantly lower than CWS without SSD in a task of non-word repetition (NWR). As to correlations, no significant results were found neither between CA and stuttering severity (as in Winter and Byrd, 2021), between CA and speech and language skills nor between CA and time elapsed from the stuttering onset. In our opinion, the absence of correlations could be explained by the unusual positive communicative attitude held by most of the subjects, with a mean score well below the mean for CWS in Vanryckeghem et al. (2022), and by the absence of any certified comorbidity with other speech and language disorders. We conclude by putting under the lens the deficient performance of the subjects in the tasks involving repetition, especially non-words repetition (NWR). This result is in line with the similar weakness of CWS in NWR (Pelczarski & Yaruss, 2016; Sasisekeran & Weathers, 2019). NWR has been claimed by Graf-Estes et al. (2007) and Coady & Evans (2008) as a powerful index for SLI. Regrettably, bad performance on NWR could be due to different causes (Krishnan et al., 2017), because NWR reflects a variety of Speech & Language skills, including lexical access (e.g. Rubenstein et al. 1970), motor planning (e.g. Yoss and Darley 1974), phonological processing (e.g. Snowling 1981), phonological memory (e.g. Gathercole & Baddeley, 1989; Bowers et al. 2018), and the task also intercepts the processes of linguistic perception, phonological assembly and articulatory behavior (see the review of Anderson et al., 2019 pointing to CWS' weakness in executive functions).

However, new light has been shed on NWR by Krishnan et al. (2017), claiming this task in children to be the reflection of the ability to compute and perform the sensorimotor conversions necessary for the production of new words. This explanation matches the theories which interpret stuttering as a consequence of sensory-motor integration deficits (Speech Motor Control models based on efferent copy or hybrid feedback: Max et al., 2004; Hickock et al., 2011; Tian & Poeppel, 2012; Civier et al., 2013; for a review, see Bradshaw e al., 2021). Last, but not least, these results could have a precious clinical value: Walsh et al. (2021) argued that prognostic index of chronic stuttering are: familiarity, low scores on standardized phonetic-phonological assessment tests, high frequency of SLD in spontaneous speech, and low scores on RNP tasks (italics are of the present authors; see also Leech et al., 2017).

Selected References:

Brundage, S. B., & Bernstein Ratner, N. (2022). Linguistic Aspects of Stuttering: Re-search Updates on the Language– Fluency Interface. Topics in Language Disorders, 42, 5–23.

Guttormsen, L.S., Kefalianos, E., Naess, K.A.B., (2015). Communication attitudes in children who stutter: A meta-analytic review. Journal of Fluency Disorders, 46-1-14.

Marini, A., Marotta, L., Bulgheroni, S., Fabbro, F. (2015). Batteria per la Valutazione del Linguaggio in Bambini dai 4 ai 12 anni. GIUNTI O.S. Organizzazioni Speciali, Firenze.

Yairi, E., & Seery, C. H. (2015). Stuttering. Foundation and Clinical Applications (2nd ed.). Pearson.

Vanryckeghem, M. & Brutten G. J. (2022), KiddyCAT - Communication Attitude Test for Preschool and Kindergarten Children Who Stutter, Adattamento italiano di S. Bernardini, L. Cocco e C. Zmarich, Hoegrefe.

The asymmetrical communication in doctor-patient relationships: the particular case of General Practitioners

Barbara Gangi
Università Ca' Foscari

Questo intervento ha l'obiettivo di presentare la ricerca sociolinguistica condotta negli studi di nove Medici di Medicina Generale (MMG o GP dall'inglese *General Practitioners*) operanti nell'ex provincia di Udine (ora Azienda sanitaria universitaria Friuli Centrale, ASUFC), che ha visto l'analisi conversazionale di oltre 18 ore di registrazioni audio, raccolte durante i colloqui degli MMG aderenti con quasi 80 dei loro pazienti. L'ipotesi iniziale è quella che dai colloqui possa emergere una relazione medico-paziente, teoricamente inserita tra le relazioni asimmetriche, ma di fatto con forti tratti di informalità e familiarità; plausibilmente, questo fenomeno è riconducibile alla continuità di una relazione che, diversamente da quanto accade nelle visite specialistiche ospedaliere o private, non è uno scambio *una tantum* ma può durare molti anni.

La letteratura di riferimento evidenzia dati quantitativi relativamente a vari aspetti, come durata delle visite (Deveugele et al. 2002), fasi della consultazione medica (Ohtaki et al. 2003), tipo e frequenza di interruzioni (Li et al. 2004, Li et al. 2008), e comprensione del gergo medico (McCarthy et al. 2012). Tuttavia, il limite più sentito della *literature review* condotta sta nel fatto che gli studi presentati coinvolgono diverse specialità di medici, tra cui si inseriscono anche i *General Practitioner* (i nostri MMG) come una delle tante specializzazioni. Per quella che è invece l'organizzazione e la logistica della relazione MMG-paziente in Italia, e in particolare nella regione FVG dove si è condotto lo studio, questo *cluster* può non rendere conto della profonda diversità tra un contesto ospedaliero e l'ambulatorio di un MMG, il che può portare a decisivi cambiamenti nella relazione e quindi nella comunicazione tra questi due interlocutori.

In particolare, uno degli elementi analizzati durante la presente ricerca al fine di rispondere alla domanda centrale, cioè se vi siano caratteristiche informali e familiari in questa relazione teoricamente asimmetrica, è quello relativo all'uso della lingua Friulana nelle visite mediche. Posta una premessa di carattere censorio (ARLeF e UniUd 2014) che fornisce dati sulle variabili anagrafiche che maggiormente correlano con l'uso del Friulano (Età > 50 anni; scolarizzazione elementare o media inferiore; genere di poco in favore dei maschi (55% vs 45%), è stato somministrato ai partecipanti un questionario sulle abitudini linguistiche per indagare l'uso del Friulano nell'interazione con varie figure professionali. I risultati sono rappresentati al Grafico 1. Al Grafico 2, si presentano i risultati dell'analisi statistica *Conditional Inference Tree* (CTree) che ricerca le variabili sociolinguistiche in grado di spiegare il fenomeno; questo evidenzia che l'abitudine ad usare il Friulano come «lingua del cuore» nei contesti familiare e amicale ("BIL_PCA_Factor1", dato derivato da una *Principal Component Analysis* condotta preliminarmente) motiva i pazienti a voler utilizzare il Friulano anche nelle consultazioni mediche col proprio MMG.

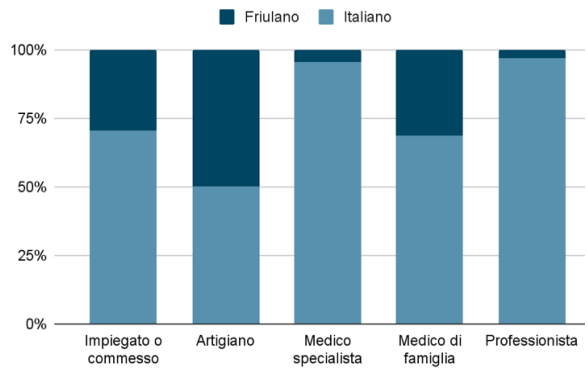


Grafico 1. Uso del Friulano con mestieri e professioni.

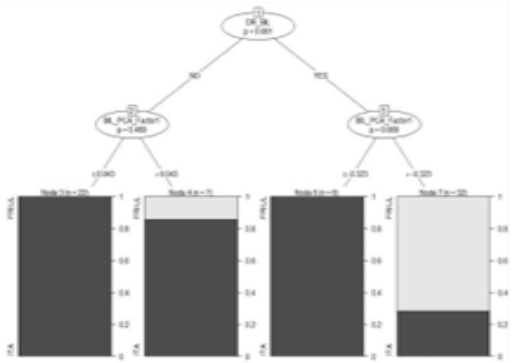


Grafico 2. CTree sull'uso di Friulano e Italiano con il MMG.

Tra gli altri aspetti indagati nella ricerca spicca poi quello relativo al pronome allocutivo (Tu/Lei) in uso tra pazienti e MMG, a confronto con altre professioni. I risultati al Grafico 3. Al Grafico 4 invece, l'analisi statistica che evidenzia al primo nodo che la fascia d'età 40-59 anni sembra dare del Tu al MMG più delle altre; questo può accadere perché questo gruppo è quello più vicino all'età media dei medici, il che ne farebbe un fatto generazionale in cui si tende a dare del Tu a persone di età simile alla nostra. Questo elemento risulta interessante perché lo sappiamo vero per le relazioni simmetriche, ma qui si potrebbe evidenziare un comportamento simile in quelle asimmetriche, ovvero, nella prospettiva opposta, questo tratto concorrerebbe a caratterizzare il basso livello di asimmetria nella relazione in esame. Il secondo nodo è invece dato dalla durata della relazione (espressa in Anni di cura / Età del paziente) che porterebbe a maggiore informalità.

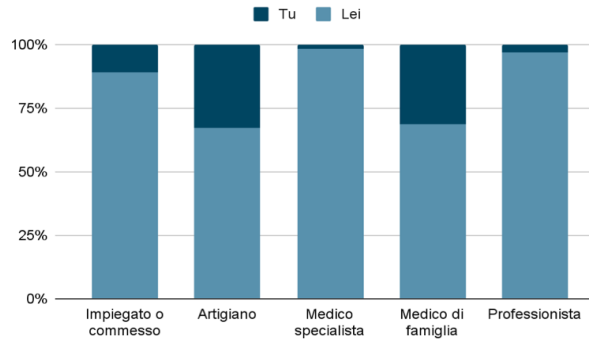


Grafico 3. Allocuzione Tu/Lei con mestieri e professioni.

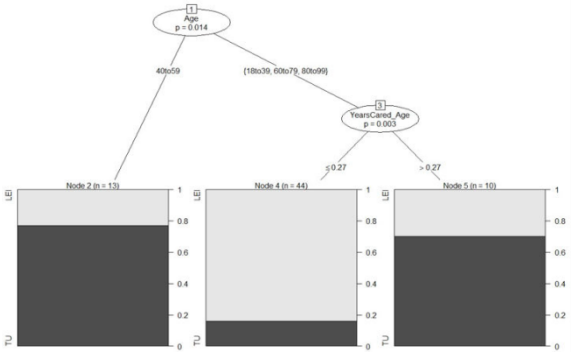


Grafico 4. CTree sull'uso degli allocutivi Tu/Lei con il MMG.

Un ultimo aspetto è legato l'analisi qualitativa di variazione linguistica in termini di tono della conversazione (in particolare, dei registri bassi come l'informale e il volgare). Questo aspetto sembra evidenziare un'intimità particolare nelle relazioni medico-paziente che durano da più tempo.

In conclusione, lo studio condotto mostra elementi intrinsechi alla relazione MMG-paziente che ne farebbero un tipo particolare di relazione asimmetrica, ove la durata della relazione e il grado di familiarità che si può creare genererebbero dei comportamenti (socio)linguistici che di discostano di molto rispetto a ciò che dalla teoria e dalla letteratura è noto sulla variazione linguistica nei contesti formali e istituzionali, in particolare rispetto a ciò che il paziente dichiara in relazione ai medici specialisti. Un limite del presente studio è di non fornire dati riguardo questa controparte ospedaliera, ma ciò potrebbe risultare in uno slancio per futuri approfondimenti.

Bibliografia:

ARLeF e UniUd 2014, *Ricerca Sociolinguistica sulla Lingua Friulana (FUR-IT)*. Agenzia Regionale per la Lingua Friulana e Università di Udine. https://arlef.it/it/materiali/ricercje-sociolinguistiche-su-la-lenghe-furlane-it/

Deveugele, M., Derese, A., van den Brink-Muinen, A., Bensing, J., & De Maeseneer, J. (2002). Consultation length in general practice: cross sectional study in six European countries. BMJ (Clinical research ed.), 325(7362), 472.

Li, H. Z., Krysko, M., Desroches, N. G., & Deagle, G. (2004). Reconceptualizing interruptions in physician- patient interviews: cooperative and intrusive. Communication & medicine, 1(2), 145–157.

Li, H. Z., Zhang, Z., Yum, Y., Lundgren, J., & Pahal, J. S. (2008). Interruption and patient satisfaction in resident-patient consultations. Health Education, 108(5), 411– 427.

McCarthy, D. M., Waite, K. R., Curtis, L. M., Engel, K. G., Baker, D. W., & Wolf, M. S. (2012). What Did the Doctor Say? Health Literacy and Recall of Medical Instructions. Medical Care, 50(4), 277–282.

Ohtaki, S., Ohtaki, T., & Fetters, M. D. (2003). Doctor-patient communication: a comparison of the USA and Japan. Family practice, 20(3), 276–282.

Dimensions of pain expression in human voice

Franco Cutugno
Università di Napoli "Federico II"

Determinare con strumenti automatici o semiautomatici quanto dolore sta provando un paziente che soffra di una patologia di interesse algologico sta diventando uno dei temi più gettonati nel mondo della corpus linguistics e delle analisi di tipo paralinguistico su dati mono e multi modali. Nonostante le difficoltà connesse con gli aspetti etici, strutturali (in alcuni casi coloro che contribuiscono alla costruzione del dataset sono persone prive di patologie acute o croniche che accettano volontariamente di essere registrate mentre ricevono stimolazioni dolorose), di disponibilità di dati in differenti lingue, in letteratura e nei repository più diffusi in rete si possono trovare diversi dataset che possono essere usati per vari tipi di analisi sia qualitative che quantitative per mettere in relazione livelli di dolore provato, con valutazioni oggettive (intensità dello stimolo somministrato) o soggettive (valutazioni del soggetto stesso con le classiche scale numeriche o, in alternativa, valutazione del clinico che osserva la registrazione).

I dati sono solitamente multimodali, vale a dire videoregistrazioni, quindi audio più video, e, in molti casi, a queste dimensioni possono essere collegati altri dati fisiologici registrati tramite l'impiego di sensori che misurano frequenza cardiaca, saturazione, transconduttanza cutanea e altri parametri di natura simile che la letteratura ha dimostrato avere correlazione con lo stato del paziente che prova dolore (sia in una fase acuta che in forma cronica).

Limitandosi alla dimensione audio video, il mio intervento vuole considerare le tre dimensioni primarie espresse in queste registrazioni: la valutazione della espressione del volto, il testo che viene prodotto dal parlante che descrive la propria esperienza, le caratteristiche paralinguistiche della voce di chi descrive il proprio dolore a posteriori o mentre lo sta provando. Per la sede AISV mi soffermerò su questa ultima dimensione, descrivendo i dataset che sono disponibili in letteratura specificamente miranti allo studio della variabile fonica, descriverò un dataset che è in corso di raccolta presso l'Istituto Tumori Pascale di Napoli, e proporrò una serie di misure basate su un approccio con Deep Neural Networks, utilizzando sistemi liberamente disponibili in letteratura pre-addestrati allo scopo di riconoscere le emozioni espresse nel segnale vocale, a cui abbiamo applicato un successivo fine tuning per specializzarli alla previsione automatica dei livelli di dolore confrontata con quella indicata nei dataset secondo quanto descritto sopra.

Non mancherà un livello di riflessione più specificamente fonetico-acustico sugli aspetti più rilevanti del parlato presente nei corpora, analizzati per ora solo qualitativamente.

Very Short Utterances in clinical interviews with patients with schizophrenia

Francesco Cangemi[1], Martine Grice[1], Valeria Lucarini[2], Malin Spaniol[3], Kai Vogeley[3], Simon Wehrle[1]
[1]University of Cologne, [2]Université Paris Cité, [3]University Hospital Cologne

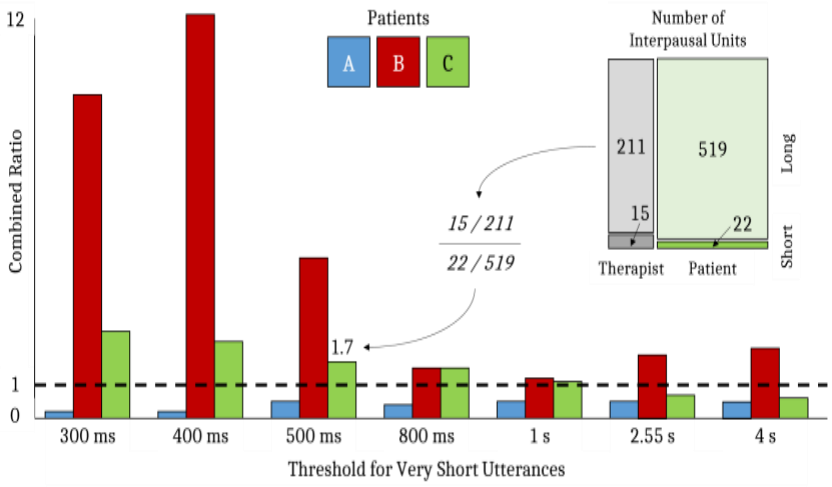
According to the concept of schizophasia (Kraepelin 1915), language provides a crucial means of expression for formal thought disorders, which are a key symptom of schizophrenia. These expressions span a continuum from alogia to logorrhea (Andreasen & Flaum 1991). In order to understand how these different speech behaviours relate to the disorder, it would be useful both for psychiatrists and linguists to get access to large corpora of speech produced by individuals with schizophrenia. However, partly due to privacy concerns, such corpora are extremely rare. For this reason, drawing on Chapple (1939), the informativeness of content-free interaction records has been explored, based exclusively on timestamps for the speech activity of the interactants (Cangemi, 2022). By expunging verbal content from audio recordings or orthographic transcriptions, such records make it possible to study the distribution of floor possession and dynamic interactional behaviour during conversation, while at the same time fully respecting the privacy of the speakers. This in turn allows researchers to share results in an open research framework.

In order to test the usefulness of content-free records, Cangemi (2022) reanalysed a subset of Dovetto & Gemelli (2013), one of the few publicly available corpora of interviews between a therapist and patients with schizophrenia. The metrics extracted from the content-free records yielded a remarkably clear separation of the three patients featured in the subset. Distinct patterns of interaction were extracted, based on the speech time of therapist/patients and on the amount of joint speech/silence. The patterns hinted at different modes of "direction" across the three interviews (Viaro & Leonardi 1983), consistently with the phenomenological analysis of the three patients provided in Pastore (2013). Despite their success, however, these metrics leave ample room for improvement. For example, compared to other cases of speech activity, listener signals, such as backchannels like "yes" and "uh-huh" (Schegloff 1982), are known to play a different role in shaping the interaction between therapist and patient. Of particular importance is the interplay between the listener signals produced by the two interactants, since the fundamental unit of interaction is the dyad, and not the individual speaker. The content-free records used in Cangemi (2022), however, are blind to such distinctions, since they do not make use of the verbal content of the interaction.

To address this shortcoming, we analysed content-free records for a two-hours subset of Dovetto & Gemelli (2013) after establishing an automatic identification of listener signals. The reduced nature of content-free records only allows us to use temporal data. For this reason, as a first approximation, we followed Edlund, Heldner, Moubayed, Gravano & Hirschberg (2010) in classifying stretches of speech activity shorter than 500ms as Very Short Utterances (VSU). This category has been shown to contain most if not all instances of backchannel signals, as well as cases of other signals (such as acknowledgements, answers to yes-no questions and repair initiations), which however cannot be filtered out due to the content-free nature of our approach. We then used this classification to reanalyse the interactions, by counting for each patient-therapist dyad the instances of VSU and long units, separately for the therapist and the patient. For example, while in conversation with patient C, the therapist produced 15 VSU and 211 long units, whereas patient C produced 22 VSU and 519 long units; see Figure. We then calculated the individual ratio of VSU to long units for each interactant, and subsequently their combined ratio for each dyad (e.g. (15/211) / (22/519) ≈ 1.7). This combined ratio provides a rough estimation of the degree of symmetry of the interaction. High values suggest that the therapist is mainly engaged in active listening, while values below 1 suggest that the therapist leads the interaction by occupying the conversational floor more substantially than the patient.

These findings are consistent with the phenomenological analysis in Pastore (2013), suggesting that this metric can be employed as an empirical indicator for the structure of the individual conversation. Compatibly with the expectations for an interview in a medical setting, the therapist takes an active listening role (combined ratio = 1.7) during the interaction with patient C, a subject who seems to have

regained the ability to communicate with clarity and confidence after having developed a long-term familiarity with psychotic episodes. Patient B, on the other hand, shares an unstructured delusion in a tumultuous and logorrheic flight of ideas; for this interaction, the combined ratio is 4.8, indicating that most of the therapist's speech activity takes place in the back channel. The opposite is true for the interaction with patient A, who exhibits stark negative symptoms. Here the combined ratio is 0.5, indicating that the therapist engages in comparatively many more content-heavy utterances, probably in order to counter the patient's tendency to alogia.



To test whether VSU extracted with a durational threshold of 0.5s can serve as a proxy for the degree of symmetry in the interaction, we tested six additional values (see Figure). Thresholds around 1s fail to capture differences between patients B and C. Thresholds around 3s fail to capture differences between patients A and C (2.55s is the mean duration of all interpausal units in the subcorpus). These results suggest that the short utterances indeed play an important role in extracting measures of symmetry from content-free interaction records. This interpretation is reinforced by the fact that, for this corpus, shorter values yield even clearer results than the 500ms threshold suggested by Edlund et al. (2010).

Undoubtedly, a fully-fledged understanding of communicative behaviour in individuals diagnosed with schizophrenia requires a close examination of many more dimensions (Fatouros Bergman et al. 2006). These could include lexical choices, discourse coherence and non-verbal components such as gaze, gesture and posture. The study of these dimensions, however, would require costly multimodal measures and would pose a challenge to the privacy of individuals. The speech activity measures employed in this study, despite their substantial reductionism, yield results in line with independent, detailed phenomenological analyses, respect the privacy of the interactants, and are easy to extract. As such, we believe that they hold promise for a variety of applications in clinical settings, such as the real-time profiling of patients during an interview and/or the training of medical personnel.

References

Andreasen, N. C., & Flaum, M. (1991). Schizophrenia: The Characteristic Symptoms. *Schizophrenia Bulletin*, 17(1), 27–49.

Cangemi (2022). Schizophrenia at the crossroads of thought, language and communication. *Proceedings of the Thought and Language meeting* (Institute of Electronics, Information and Communication Engineers), November 2022, Yamagata, Japan. 1-5.

Chapple, E. D. (1939). Quantitative Analysis of the Interaction of Individuals. *Proceedings of the National Academy of Sciences of the United States of America*, 25(2), 58–67.

Dovetto, F. M., & Gemelli, M. (2013). *Il parlar matto. Schizofrenia tra fenomenologia e linguistica. Il corpus CIPPS.* (Second edition with DVD-ROM). Aracne.

Edlund, J., Heldner, M., Moubayed, S. A., Gravano, A., & Hirschberg, J. (2010). Very short utterances in conversation. *Working Papers, Lund University, Department of Linguistics and Phonetics*, 54, 11–16.

Fatouros Bergman, H., Preisler, G., & Werbart, A. (2006). Communicating with patients with schizophrenia: Characteristics of well functioning and poorly functioning communication. *Qualitative Research in Psychology*, 3(2), 121–146.

Kraepelin, E. (1915). *Psychiatrie. Ein Lehrbuch für Studierende und Ärzte* (Eighth edition). Barth.

Pastore, C. (2013). Tre modi dell'esperire schizofrenico. Mondo congelato, Mondo frammentato, Mondo delirato. In Dovetto & Gemelli (2013), 19–44.

Schegloff, E. (1982). Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In Tannen, D. (Ed.), *Analyzing Discourse: Text and Talk*, 71–93. Georgetown University Press.

Viaro, M., & Leonardi, P. (1983). Getting and giving information: Analysis of a family-interview strategy. *Family Process*, 22, 27–42.

Pragmatic characteristics of speech in Autism Spectrum Disorder: a pilot study on Italian-speaking children living in Campania

Syria Cira Imparato[1], Maria Izzo[2], Olga Liguori[2], Ernesto Orsino[2], Donata Sensale[2], Tina Santarpia[2], Giulia Perfetto[2], Giovanna Gison[2] & Gloria Gagliardi[1]

[1]Alma Mater Studiorum – Università di Bologna [2]Centro Medico Riabilitativo – Pompei

Background

Lo studio si propone di indagare le caratteristiche pragmatiche dell'eloquio di bambini con Disturbo dello Spettro dell'Autismo (ASD – Autism Spectrum Disorders), etichetta con cui nella letteratura scientifica si indica un insieme eterogeneo di disturbi del neurosviluppo che esordiscono nei primi anni di vita e sono caratterizzati sotto il profilo clinico dalla presenza di difficoltà nell'area socio-emotiva e comunicativa e di pattern di comportamento, interessi e attività ristretti e ripetitivi (APA, 2013). L'eterogeneità che contraddistingue tale tipologia di neurodiversità sul piano eziopatogenetico e sintomatologico si estende anche alla dimensione linguistica, al punto che è possibile individuare 'fenotipi comunicativi' (Brandi, 2005; Biancalani et al., in stampa) e, per descrivere il quadro generale, ricorrere all'immagine di un continuum lungo cui si dispongono, a un estremo, i soggetti che non hanno mai acquisito il linguaggio ('non-verbali') e, all'altro, coloro che, pur sviluppando abilità linguistiche, manifestano difficoltà di natura comunicativa (Vivanti, 2021).

Anche in virtù del suo stretto legame con la sfera intersoggettiva, che risulta essere altamente compromessa nell'autismo (Vicari, 2012), la competenza pragmatica è stata identificata come il nucleo centrale delle difficoltà nei soggetti autistici, inclusi coloro che abbiano sviluppato competenze linguistiche fonetico-fonologiche, morfo-sintattiche e lessicali nella norma (Arciuli & Brock, 2014). Molti resoconti clinici sull'eloquio di individui con ASD concordano infatti nell'osservare che, di frequente, la loro iniziativa comunicativa spontanea è scarsa o del tutto assente; quando sembrano voler avviare una conversazione – il che avviene sporadicamente – usano il linguaggio in modo limitato: raramente fanno commenti, chiedono informazioni o chiarimenti all'ascoltatore (Tager-Flusberg et al., 2005). Altre difficoltà si manifestano nella forma di violazioni delle regole non scritte della conversazione (Vivanti, 2021): per esempio, faticano a mantenere la reciprocità e l'alternanza dei turni dialogici (Biancalani et al., in stampa), a portare avanti il topic proposto dall'interlocutore (Wilkinson, 1998), intervenendo spesso con commenti irrilevanti e poco adeguati allo scambio comunicativo. Hanno difficoltà a decifrare espressioni metaforiche, ironiche e richieste indirette (Kissine et al., 2016), ossia tutte quelle espressioni che, per essere interpretate, richiedono di andare "oltre" il significato letterale; inoltre risultano spesso incapaci di compiere inferenze, cioè di cogliere i contenuti non asseriti esplicitamente e di dedurre il collegamento implicito tra gli avvenimenti di una storia (Biancalani et al., in stampa).

Anche le abilità narrative risultano spesso compromesse: i dati ottenuti da compiti di storytelling rivelano, invero, che i soggetti autistici non sono capaci di riportare il nucleo di un racconto, né di considerare le necessità dell'ascoltatore; di conseguenza, raccontano gli eventi in maniera confusa e disorganizzata (Pfanner et al., 2008), spesso tralasciando passaggi fondamentali e concentrandosi su dettagli poco utili ai fini della comprensione del racconto da parte dell'interlocutore.

Obiettivo dello studio

Lo studio si pone l'obiettivo di individuare e descrivere le caratteristiche pragmatiche e conversazionali che contraddistinguono l'eloquio dei soggetti di età scolare con Disturbo dello Spettro Autistico a partire dall'analisi di sessioni di parlato semi-spontaneo in cui i bambini interagiscono con il proprio partner comunicativo durante l'esecuzione di compiti narrativi (i.e., generazione e rievocazione di storie) e descrittivi (i.e., descrizione di figura complessa).

Il progetto di ricerca è stato approvato dal Comitato di Bioetica dell'Alma Mater Studiorum - Università di Bologna (n. 0173455/2022). I caregiver di tutti i bambini coinvolti nello studio hanno espresso il loro consenso alla partecipazione del minore alla ricerca e al trattamento dei suoi dati personali.

Metodo

Il campione reclutato per lo studio si compone di 34 bambini in totale (26M, 8F) di età compresa tra i 6 e i 13 anni (9;9 ± 2;5), divisi in due gruppi bilanciati per sesso ed età:

ASD: 17 soggetti che hanno ricevuto diagnosi di ASD ad alto funzionamento;

GC – Gruppo di Controllo: 17 pari età con sviluppo linguistico e cognitivo tipico

I bambini con ASD sono stati reclutati presso il Centro Medico Riabilitativo di Pompei (NA), mentre quelli a sviluppo tipico hanno partecipato allo studio su volontà dei genitori, interessati all'argomento e agli esiti della ricerca. Per ridurre gli effetti della variazione diatopica, sono stati considerati unicamente bambini provenienti dalla regione Campania e aventi l'italiano come lingua materna. Sono state somministrate, in forma di gioco, le seguenti prove:

1. *Descrizione di figura complessa*: "Cookie Theft" (Goodglass *et al.*, 2001) (fig. 1);
2. *Generazione di storia* ("*telling"*): "Una festa di compleanno" (tratta da "Sequenze da raccontare", Shubi, fig. 2);
3. *Rievocazione di storia* ("*retelling*"): "Bus Story Test" (adattamento italiano a cura di Mozzanica *et al*., 2016) (fig. 3).

Fig.1: Cookie Theft

Fig. 2: Una festa di compleanno

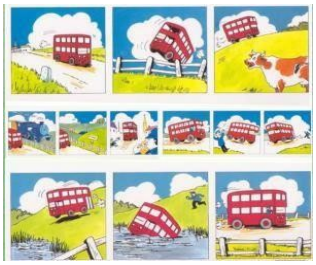


Fig. 3: Bus Story Test

Le sessioni sono state audioregistrate in formato WAV, trascritte ortograficamente in formato CHAT-LABLITA (MacWhinney, 1991; Cresti & Moneglia, 2005) in ambiente ELAN e annotate manualmente. Gli indici considerati nello studio riguardano: a) la fluenza dell'eloquio: durata del task, total phonation time (TPT), velocità di articolazione e velocità di eloquio, numero e durata di pause piene e pause vuote; b) aspetti di natura pragmatica e narrativa: coesione testuale (i.e., tipologia e appropriatezza delle espressioni referenziali, numero di connettivi testuali), capacità di fare inferenze, e grammatica delle storie (Stein & Glenn, 1979).

Risultati

I risultati ottenuti dalla somministrazione delle tre prove, seppur preliminari, hanno evidenziato una serie di differenze tra i gruppi ASD e GC. In particolare si è osservato che, nel corso dello svolgimento dei task narrativi, i bambini con ASD spesso si sono limitati a fornire una mera descrizione delle immagini, trascurando i collegamenti (impliciti) tra gli eventi in esse raffigurati, focalizzandosi su elementi accessori, e tralasciando spesso passaggi fondamentali nella prospettiva dell'ascoltatore. Alle sollecitazioni dello sperimentatore o del/della terapista sono seguite di frequente pause lunghe – fatte perlopiù di silenzi, vocalizzi, sbadigli o lamenti – oppure risposte irrilevanti rispetto all'argomento di discussione. Sebbene anche nel GC siano state riscontrate, in diversi casi, pause più o meno lunghe tra un turno e il successivo, esse sono state impiegate dai bambini nella quasi totalità dei casi per prendere tempo così da formulare una risposta più adeguata allo scambio. Al contrario, nei soggetti autistici, le pause censite sono nella maggior parte dei casi vuote, e comunque non finalizzate all'elaborazione di una risposta per l'interlocutore. Inoltre, talvolta, i bambini del gruppo ASD hanno mostrato difficoltà nella comprensione delle istruzioni fornite dall'intervistatore.

Infine, sono state riscontrate differenze significative tra i due gruppi anche a livello di pianificazione del discorso: i soggetti con autismo hanno mostrato maggiori difficoltà nella costruzione di un testo orale coeso e coerente, riflesso nell'uso di un numero limitato di connettivi testuali e segnali discorsivi.

Riferimenti bibliografici

APA (2013). *Diagnostic and statistical manual of mental disorders*, *Fifth edition*. Washington DC-London: America Psychiatric Publishing. Arciuli J. & Brock J. (Eds.) (2014). *Communication in autism*, *11*. Amsterdam: John Benjamins Publishing Company.

Biancalani S. *et al*. (in press). Aspetti soprasegmentali e pragmatici dell'eloquio di bambini in età scolare con Disturbo dello Spettro Autistico: uno studio pilota. In: M. Castegneto & M. Ravetto (Eds.), *La comunicazione parlata. Spoken communication*. *2020*. Roma: Aracne.

Brandi L. (2005) Linguaggio e comunicazione: dis/giunzioni autistiche. *Quaderni del Dipartimento di Linguistica - Università di Firenze*, 15: 169-192. Cresti E. & Moneglia M. (2005). *C-ORAL-ROM. Integrated Reference Corpora for spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Goodglass H. *et al.* (2001). *The Boston Diagnostic Aphasia Examination (BDAE)*.

Kissine M. *et al*. (2016). La pragmatique dans les troubles du spectre autistique. Développements récents. *Médecine/sciences*, 3*2*(10), 874-878. MacWhinney B. (1995). *The CHILDES Project: Tools for analyzing talk*. Mahwah: Lowrence Erlbaum Associates.

Mozzanica F. *et al*. (2016). Reliability, validity and normative data of the Italian version of the Bus Story test. *International Journal of Pediatric Otorhinolaryngology*, 89: 17-24.

Pfanner L. *et al.* (2008). Comunicazione e linguaggio nei disturbi pervasivi dello sviluppo. *Giornale di Neuropsichiatria dell'Età Evolutiva*, 28, 59–74.

Stein N.L. & Glenn C.G. (1979). An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.), *New directions in discourse processing*. Hillsdale, NJ: Ablex Publishing Corporation.

Tager-Flusberg H. *et al*. (2005). Language and Communication in Autism. In F.R.Volkmar *et al.* (Eds.), *Handbook of autism and pervasive developmental disorders: Diagnosis, development, neurobiology, and behavior*. Hoboken: John Wiley & Sons Inc., pp. 335–364.

Vicari S. *et al.* (2012). *L'autismo. Dalla diagnosi al trattamento*. Bologna: il Mulino.

Vivanti G. (2021). *La mente autistica. Le risposte della ricerca scientifica all'enigma dell'autismo*. Firenze: Hogrefe Editore (2a ed.).

Wilkinson K.M. (1998). Profiles of language and communication skills in autism. *Mental retardation and developmental disabilities research reviews*, 4(2), 73-79.

Pragmatic and prosodic features in Autism Spectrum Disorder: analysis of the speech of school-aged children.

Virginia Chiti, Valentina Saccone, Alessandro Panunzi
Università degli Studi di Firenze

Il presente contributo costituisce un primo tentativo di indagare le alterazioni della componente prosodica nel parlato di 9 bambini in età scolare con ASD (Disturbo dello Spettro Autistico), nella prospettiva più generale di delineare un quadro di variazione del fenomeno attraverso la raccolta di un corpus più ampio di parlanti.

Nell'inquadramento fornito dal DSM 5 (2013), L'ASD è identificato come una sottocategoria dei Disturbi del neurosviluppo. Le aree di deficit funzionale che portano alla diagnosi sono principalmente due: comunicazione/interazione sociale e stereotipie comportamentali/ristrettezza di interessi.

Per quanto riguarda la prima area, di interesse in questa sede, gli studi sottolineano come un approccio sociale anomalo comporti ad esempio il fallimento nella reciprocità della conversazione, l'incapacità di iniziare un turno – sia per prendere autonomamente parola che per rispondere – e di interagire in contesti sociali; sono inoltre comuni anomalie del contatto visivo e del linguaggio del corpo, deficit di comprensione e uso dei gesti, mancanza totale di espressività facciale e di comunicazione non verbale (ISAAC 2017). Nell'analisi dei deficit della comunicazione, lo studio clinico rivela come costante un'alterazione dell'andamento prosodico rispetto al parlato non patologico: un'insolita prosodia espressiva che va di pari passo con difficoltà pragmatiche (già in Kanner 1943; Eigisti et al. 2012; McAlpine et al. 2014), con un'alta variabilità di caso in caso che non permette di definire un profilo patologico caratteristico del disturbo. Data l'eterogeneità dei fenomeni riscontrati, la descrizione della prosodia resta sempre alquanto generica; viene descritta come «monotona, robotica e pedante, non modulata nell'intonazione (piatta o al contrario esagerata) e nella rapidità di eloquio (troppo lenta o, viceversa, eccessivamente accelerata)» (Barbi et al. 2015; Eigisti et al. 2012). Inoltre, si nota come i deficit prosodici generalmente permangano anche quando altre aree linguistiche migliorano (Eigisti et al. 2012 anche sui deficit nella comprensione della prosodia affettiva; McCann & Peppé 2003 anche sul deficit ricettivo nell'espressione prosodica).

Per un approfondimento sull'aspetto prosodico-pragmatico, è stato raccolto un campione di audio-registrazioni di parlato di bambini con ASD durante i loro incontri di terapia settimanale presso la Fondazione MAiC di Pistoia a luglio-agosto 2022, per un totale di circa 150 minuti di registrazioni. Con l'aiuto degli operatori della Fondazione, sono stati selezionati 9 soggetti (7 maschi e 2 femmine, di età compresa fra gli 8 e i 12 anni) con le seguenti caratteristiche: diagnosticati ASD; che presentano linguaggio intelligibile; di lingua madre italiana; frequentanti la scuola primaria o secondaria di primo grado. La selezione è stata quindi effettuata cercando di rendere il campione omogeneo in termini di sviluppo linguistico, età e funzionamento. Il dislivello fra maschi e femmine del campione segue l'andamento delle diagnosi: in Italia, per esempio, dove a 1 bambino su 77 è diagnosticato il Disturbo dello Spettro Autistico, i maschi risultano 4.4 volte più a rischio rispetto alle femmine (Ministero della Salute 2022).

Durante le registrazioni, ai bambini è stato chiesto di presentarsi e raccontare spontaneamente qualcosa di sé; in caso di difficoltà, i partecipanti sono stati incoraggiati con dei suggerimenti e domande generiche relative alla classe frequentata, allo sport praticato e alle vacanze estive. Successivamente, è stato loro proposto il test elaborato a partire dalla storia *Frog, Where are you?* (Mayer 1969) che costituisce, nella pratica clinica logopedica, un test per la valutazione delle competenze narrative, rientra tra i test di *telling* e ha permesso di raccogliere esempi di parlato di tipo narrativo.

Le registrazioni sono state interamente trascritte e segmentate in unità prosodiche ed enunciati secondo i principi della Teoria della Lingua in Atto tramite l'individuazione percettiva di break prosodici terminali e non terminali (L-AcT, Cresti 2000; Cresti & Moneglia 2010, Izre el et al 2020). Per ogni singolo

partecipante è stato delineato un profilo linguistico che comprende informazioni sia qualitative che quantitative, sulle abilità comunicative, le competenze narrative e i tempi di elaborazione, la presenza di segnali discorsivi e riempitivi, il rapporto fra enunciati semplici e complessi (ovvero formati da una o più unità tonali). I dati forniscono informazioni pragmatiche sulla struttura interna dei turni e sull'elaborazione dell'informazione. Le registrazioni sono poi state analizzate manualmente tramite il software Praat (Boersma & Weenik 2005); per ognuna, differenziando le parti di presentazione personale e racconto della *Frog story*, sono stati annotati f0 media, range di f0, velocità di eloquio (tokens/stretch of speech) e intensità media. Un confronto fra i dati ottenuti permette di mettere in relazione proprietà prosodiche e spinta comunicativa dei bambini e aiuta a sottolineare le tipologie di alterazioni che l'ASD comporta a livello prosodico. Verranno presentati i dati nel loro complesso per individuare caratteristiche comuni e range di variazione fra i casi analizzati.

Le analisi svolte permettono quindi di sottolineare le caratteristiche pragmatico-prosodiche del parlato autistico in scambi semi-spontanei durante gli incontri di terapia. Tali caratteristiche sono fondamentali dal punto di vista comunicativo-pragmatico e dell'integrazione sociale, aspetti generalmente non inclusi nella valutazione e nel trattamento logopedico ordinario dei soggetti con ASD.

Il lavoro presenta attualmente il limite di non considerare un confronto con dati di controllo. La costruzione di un corpus di parlanti di età scolare fa parte degli obiettivi futuri che il lavoro si propone.

Riferimenti bibliografici principali

Boersma P., Weenink D. (2005), *Praat: Doing phonetics by computer*, http://www.praat.org/

Cresti, E. (2000), *Corpus di italiano parlato. Vol. I-II* (Studi di grammatica italiana pubblicati dall'Academia della Crusca), Firenze, Accademia della Crusca.

Cresti, E. & Moneglia, M. (2010), *Informational patterning theory and the corpus based description of spoken language*, Firenze, Firenze University Press.

DSM 5 *The Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM–5); American Psychiatric Association, 2013.

Eigsti, I., Schuh, J., Mencl, E., Schultz, R., Paul, R. (2012). *The neural underpinnings of prosody in autism*, in «Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence», Psychology Press Taylor & Francis group.

ISAAC (2017). *Documento preliminare ISAAC Italy su CAA e disturbo dello spettro dell'autismo*, Roma. Kanner, L. (1943). *Autistic disturbances of affective contact. Nervous child, 2*(3), 217-250.

Izre´el S., Mello H., Panunzi A., Raso T. (2020), *In search of basic units of spoken language: A corpus-driven approach*, Amsterdam, John Benjamins.

Mayer, M. (1969), *Frog, Where are you?*

McAlpine,A., Plexico, L. W., Plumb, A. M., Cleary, J. (2014). *Prosody in Young Verbal Children With Autism Spectrum Disorder*, in «Contemporary Issues in Communication Science and Disorders», 41.

McCann, J., Peppé, S., (2003), *Prosody in autism spectrum disorders: a critical review*, in «International Journal of Language & Communication Disorders», 38 (4), Edinburgh, Taylor & Francis.

Ministero della Salute (20.01.2022) https://www.salute.gov.it/portale/saluteMentale/dettaglioContenutiSaluteMentale.jsp?lingua=italiano&id=5613&area=salute%20mentale&menu=vuoto

IL PARLATO IN AMBITO MEDICO:
Analisi linguistica, applicazioni tecnologiche e strumenti clinici

SPOKEN LANGUAGE IN THE MEDICAL FIELD:
Linguistic analysis, technological applications and clinical tools

FRIDAY 17 FEBRUARY
POSTER SESSION III

Analysis of spelling skills in subjects with Developmental Dyslexia

Rosso Manuel Senesi
Università di Pisa

Nell'analisi della fenomenologia del disturbo di Dislessia Evolutiva si riscontra un'estrema eterogeneità dei *deficit* riportati in letteratura, che spaziano da *deficit* di natura fonologica (Snowling 2001 e Ramus et al. 2003) e morfosintattica (Antón-Méndez et al., 2019) a *deficit* di natura motoria (Capellini et al. 2010) o di percezione visiva (Giofrè et al., 2019) e uditiva (Tallal & Stark, 1982), fino ai problemi di organizzazione temporale (Thomson & Goswami, 2008 e Pagliarini et al., 2020). Come notano Pennington et al. (2012), nonostante questa eterogeneità, la maggior parte delle proposte eziologiche dominanti suppongono che il disturbo possa essere fatto risalire comunque ad un'unica causa, vale a dire un *deficit* di natura fonologica. Come già osservato da Spencer (1991) e Castles & Friedmann (2014), le teorie che pongono alla base del disturbo di Dislessia Evolutiva un *deficit* di natura fonologica soffrono di circolarità argomentativa, in quanto si basano sulla manipolazione esplicita di segmenti fonologici (*Phonological Awareness*). Come osservato da Adrián et al. (1995) e Morais et al. (1979), i risultati dei test di *Phonological Awareness*, utilizzati per comprovare la teoria fonologica, sono condizionati dall'acquisizione di un sistema alfabetico segmentale. Morais (2021) conclude che l'idea di un segmento fonologico non è conciliabile con la natura continua dell'onda sonora. L'autore suppone che concetti quali fono e, di conseguenza, fonema derivino da un errore di analisi ed in particolare dalle lenti alfabetizzate che il ricercatore indossa. Il dubbio ricorda, in parte, la *querelle* tra Marotta (2010) e Albano Leoni (2011) sulla presunta morte del fonema. In quest'ottica il disturbo di natura fonologica in soggetti con Dislessia sarebbe un sintomo secondario, derivante dalla acquisizione deficitaria del sistema di scrittura segmentale. Questa supposizione è in linea con l'ipotesi di ristrutturazione del sistema fonologico presentata da Perre et al. (2009) e Brewer (2008).

Basandosi sulle osservazioni di Turk & Shattuck-Hufnagel (2020) e Nakai et al. (2012), si osserva, però, la necessità di distinguere un modulo fonologico simbolico dai meccanismi dominio-generali della componente fonetica del linguaggio. Fowler (2016) e Kazanina et al. (2018) osservano che la discretizzazione in segmenti del segnale acustico avviene tramite l'applicazione di operazioni cognitive di base che condividono basi neurofisiologiche simili a quelle dell'organizzazione ritmica degli stimoli visivi (Poeppel et al. 2008). Quest'ottica è compatibile con il quadro teorico della fonologia *Substance-free* (Scheer *in stampa*), in cui le strutture del modulo fonologico, considerato come amodale e incapsulato, si interfacciano con le rappresentazioni fonetiche tramite un processo di trasduzione (Volenec & Reiss 2017). Questo quadro teorico è stato recentemente utilizzato da Pathi & Mondal (2021) per analizzare e classificare le disfluenze in soggetti con *Speech Sound Disorders* (*SSD*). Un'impostazione teorica simile è stata adottata da Berent et al. (2016) per delineare l'eziologia del Disturbo di Dislessia Evolutiva. Dai dati raccolti emerge che l'eziologia più probabile per il disturbo di Dislessia sia un deficit nella trasduzione delle strutture fonologiche in rappresentazioni fonetiche. È possibile, inoltre, ipotizzare che lo stesso modulo di manipolazione simbolica sia sotteso alla formazione del modulo grafematico, che estrarrebbe a partire da stimoli visivi dei tratti grafetici distintivi (un'intuizione simile si trova già in Gibson & Levin 1975 e Althaus 1973).

Si ritiene, dunque, necessario reinterpretare le categorie analitiche utilizzate per classificare gli errori di spelling proposte in letteratura (Angelelli et al. 2004, 2010). I casi di sostituzione e omissione verranno analizzati non solo sulla base dei tratti fonologici e delle norme ortografiche ma anche sulla base dei tratti grafetici distintivi, così come sulla base della complessità articolatoria fonetica e grafetica della stringa. Apel et al. (2019) evidenziano l'importanza di studi volti all'analisi linguistica delle produzioni spontanee delle abilità di scrittura dei soggetti con Dislessia Evolutiva in contesto naturale. Questa impostazione è stata già seguita da Rello et al. (2014), per lo spagnolo, e Rauschenberger et al. (2016), per il tedesco, al fine di creare e annotare un *Corpus* di errori commessi da bambini con Dislessia Evolutiva in lingue differenti. Nel presente lavoro, in vista di una creazione di un *Corpus* di errori, si sono raccolti e analizzati gli errori commessi da 3 bambini di Gavi (AL) con Dislessia evolutiva di età compresa tra gli 8 e i 9 anni in

compiti in classe svolti in contesto naturale. Nell'analisi si seguirà il modello di scrittura di parole proposto da Kandel (*in stampa*).

Bibliografia:

ADRIÁN, J. A., ALEGRIA, J., & MORAIS, J. (1995). Metaphonological abilities of Spanish illiterate adults. *International Journal of Psychology,* 30(3), 329-351.

ALBANO LEONI, F. (2011). Discutendo sulla (presunta) morte del fonema. *Linguistic Studies and Essays,* 49, 205-220.

ALTHAUS, H. P. (1973): Graphetik. In H. P. ALTHAUS,

H. HENNE & H. E. WIEGAND (*eds*.), *Lexikon der germanistischen Linguistik*, 105–118. Tubingen: Niemeyer

ANGELELLI, P., JUDICA, A., SPINELLI, D., ZOCCOLOTTI, P., & LUZZATTI, C. (2004). Characteristics of writing disorders in Italian dyslexic children. *Cognitive and Behavioral Neurology*, 17(1), 18-31.

ANGELELLI, P., NOTARNICOLA, A., JUDICA, A., ZOCCOLOTTI, P., & LUZZATTI, C. (2010). Spelling impairments in Italian dyslexic children: Phenomenological changes in primary school. *Cortex*, 46(10), 1299-1311.

ANTÓN-MÉNDEZ, I., CUETOS, F., & SUÁREZ-COALLA, P. (2019). Independence of syntactic and phonological deficits in dyslexia: A study using the attraction error paradigm. *Dyslexia,* 25(1), 38-56.

APEL, K., HENBEST, V. S., & MASTERSON, J. (2019). Orthographic knowledge: Clarifications, challenges, and future directions. *Reading and Writing*, 32, 873-889.

BERENT, I., ZHAO, X., BALABAN, E., & GALABURDA, A. (2016). Phonology and phonetics dissociate in dyslexia: evidence from adult English speakers. *Language, Cognition and Neuroscience,* 31(9), 1178-1192

BREWER, J. B. (2008). *Phonetic reflexes of orthographic characteristics in lexical representation*. Ph.D. dissertation, The University of Arizona.

CAPELLINI, S. A., COPPEDE, A. C., & VALLE, T. R. (2010). Fine motor function of school-aged children with dyslexia, learning disability and learning difficulties. *Prò-Fono Revista de Atualizacao Cientìfica,* 22, 201-208.

CASTLES, A., & FRIEDMANN, N. (2014). Developmental dyslexia and the phonological deficit hypothesis. *Mind & Language*, 29(3), 270-285.

FOWLER, C. A., SHANKWEILER, D., & STUDDERT-KENNEDY, M. (2016). "Perception of the speech code" revisited: Speech is alphabetic after all. *Psychological Review,* 123(2), 125-150.

GIBSON, E. J.,& LEVIN, H. (1975). *The psychology of reading*. Cambridge: The MIT press.

GIOFRÈ, D., TOFFALINI, E., PROVAZZA, S., CALCAGNÌ, A., ALTOÈ, G., & ROBERTS, D. J. (2019). Are children with developmental dyslexia all the same? A cluster analysis with more than 300 cases. *Dyslexia*, 25(3), 284-295.

KANDEL, S. (*in stampa*). The APOMI model of word writing: Anticipatory Processing of Orthographic and Motor Information. In HARTSUIKER R. & STIJKERS K. (*eds*.), *Current issues in the Psychology of Language.* London: Routledge press.

KAZANINA, N., BOWERS, J. S., & IDSARDI, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic bulletin & review,* 25(2), 560-585.

MAROTTA, G. (2010). Sulla (presunta) morte del fonema. *Linguistic Studies and Essays,* 48, 283-304.

MORAIS, J., CARY, L., ALEGRIA, J., & BERTELSON, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7(4), 323-331

MORAIS, J. (2021). The phoneme: A conceptual heritage from alphabetic literacy. *Cognition*, 213(104740).

NAKAI, S., TURK, A. E., SUOMI, K., GRANLUND, S., YLITALO, R., & KUNNARI, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish, *Journal of Phonetics,* 40(6), 796-807.

PAGLIARINI, E., SCOCCHIA, L., GRANOCCHIO, E., SARTI, D., STUCCHI, N., & GUASTI, M. T. (2020). Timing anticipation in adults and children with Developmental Dyslexia: evidence of an inefficient mechanism. *Scientific Reports*, 10(1), 17519.

PATHI, S., & MONDAL, P. (2021). The mental representation of sounds in speech sound disorders. *Humanities and Social Sciences Communications*, 8(1), 1-12

PENNINGTON, B. F., SANTERRE-LEMMON, L., ROSENBERG, J., MACDONALD, B., BOADA, R., FRIEND, A., LEOPOLD, D.. SAMUELSSON, S., BYRNE, B., WILLCUTT, E. G., & OLSON, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of abnormal psychology*, 121(1), 212-224.

PERRE, L., PATTAMADILOK, C., MONTANT, M., & ZIEGLER, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain research*, 1275, 73-80.

POEPPEL, D., IDSARDI, W. J., & VAN WASSENHOVE, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363(1493), 1071-1086.

RAMUS, F., ROSEN, S., DAKIN, S. C., DAY, B. L., CASTELLOBETE, J. M., WHITE,, & FRITH, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain,* 126(4), 841-865.

RAUSCHENBERGER, M., RELLO, L., FÜCHSEL, S., & THOMASCHEWSKI, J. (2016). A Language Resource of

German Errors Written by Children with Dyslexia. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016) (Paris, France: European Language Resources Association)*. 83-87.

RELLO, L., BAEZA-YATES, R., & LLISTERRI, J. (2014). DysList: An annotated resource of dyslexic errors. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, May 2014*. 1289–1296

SCHEER, T. (*in stampa*). 3xPhonology. *Canadian Journal of Linguistics*

SNOWLING, M. J. (2001). From language to reading and dyslexia. *Dyslexia*, 7(1), 37-46.

SPENCER, A. (1991). A Linguist's Reflections on 'Phonological Awareness' and Literacy. *Mind & Language,* 6(2), 146- 155.

TALLAL, P., & STARK, R. E. (1982). Perceptual/motor profiles of reading impaired children with or without concomitant oral language deficits. *Annals of Dyslexia,* 32(1), 163-176.

THOMSON, J. M., & GOSWAMI, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris, 102*, 120-129.

TURK, A., & SHATTUCK-HUFNAGEL, S. (2020). Timing evidence for symbolic phonological representations and phonology-extrinsic timing in speech production. *Frontiers in Psychology*, 10(2952).

VOLENEC, V., & REISS, C. (2017). Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface. *Biolinguistics*, *11*, 251-294.

Social distance and intonation in interrogative requests: A study on Lecce and Rome Italian

Mayara Da Silva Neto, Barbara Gili Fivela*, Elisabetta Santoro, Flaviane Romani Fernandes Svartman
*Universidade de São Paulo - São Paulo (Brazil), *Università del Salento – Lecce (Italy)*

Requests are speech acts through which a speaker tries to get hearers to do something (Searle, 1979) and are quite frequent in our everyday life. There is an extensive literature on how pragmatic aspects, such as the social distance between interlocutors (Brown and Levinson, 1987), can influence a speaker in their choice of lexical and morphosyntactic elements when performing a request, mitigating or not its illocutionary force (e.g. Takimoto, 2007; Nuzzo, 2013; Spadotto & Santoro, 2019; Santoro, Kulikowski & Silva, 2017; Silva Neto, 2018). Studies that focus on the impact of social distance or even mitigation on the prosody of requests, however, are still rare, although there are some studies that make reference to this possibility (cf. Albano Leoni, 2003).

In this paper, which is part of a larger project on the production of orders and requests in the case of different social distances between speakers in Brazilian Portuguese and Italian, we aim to investigate whether social distance influences the intonational patterns of requests, especially those performed by means of an interrogative form, produced by Italian speakers from Lecce and Rome. More specifically, we investigate the impact of social distance on intonation patterns in (simulated) interactions by considering the distance as modulated both by the low/high familiarity between speakers and by the lack/presence of explicit mitigators. The main hypothesis is that, independently of the variety considered, social distance might affect intonation patterns, with high distance favoring the use of less marked intonation patterns (more neutral for the specific sentence type under analysis), especially when the sentence includes no explicit mitigators. Further, as the requests under analysis are performed in an interrogative form, we expect that they show similarities with information-seeking questions (see Silva Neto et al., 2020).

The speech dataset was obtained by means of a Discourse Completion Test - DCT (Levenston & Blum, 1978; Blum-Kulka, 1982; Blum-Kulka & Olshtain, 1984; Blum-Kulka, House & Kasper, 1989), which is a widely used methodology in the field of Pragmatics, but that is increasingly used in research that investigate prosodic aspects (Vanrell, Feldhausen & Astruc, 2018). Ten informants from Lecce and Rome (five speakers for each variety, both male and female, aged between 20-35 years) were asked to read written contexts that emulated everyday life situations, with either high or low social distance between hypothetical speakers. They stimulated a specific pragmatic interpretation as well as the production of certain sentence modalities and keywords (Table 1). In line with the modified DCT used in Gili Fivela et al.'s (2015), speakers were asked, first, to spontaneously react to the proposed context, and, second, to read a sentence that fit the same context, where the target word, the number of syllables and their structure and segmental composition were controlled. Speakers were audio recorded while performing the task, wearing a microphone (headset) connected to a computer equipped with an integrated Realtek audio card. Both recordings and analysis were performed in Praat (Boersma & Weenink, 2022) (sampling: 22,050 Hz). More specifically, the analysis was made by identifying and annotating tonal events with reference to the Autosegmental-metrical framework (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988; Ladd, 2008 [1996]).

Data discussed in this paper only regard read sentences including the target words dividi 'split it' and lava 'wash it' as produced by three speakers both from Lecce and Rome. Speech data on the intonation pattern found in requests will be discussed, with specific attention to pattern variation depending on the social distance between speakers and the presence of mitigators, as well as on the comparison between intonation strategies observed in the two varieties under investigation.

Table 1. Sample of target sentences and contexts (for each speech act, there is a context followed by a target phrase).

	Imperative Request (interrogative form)			
	High social distance context		**Low social distance context**	
	No mitigation	**Explicit mitigation or utterance with verb + object**	**No mitigation**	**Explicit mitigation or utterance with verb + object**
Dividi	*You are in a pizzeria and give your order to the waiter. Since you want to eat the pizza with a friend, ask the waiter, who is your age, to divide it.*		*During a dinner with friends, the dessert is a cake. You, therefore, ask the friend who brought it to divide it.*	
	La dividi? Do you divide it?	*La dividi, per favore?* Do you divide, please?	*La dividi?* Do you divide it?	*La dividi, per favore?* Do you divide, please?
Lava	*You are in a pizzeria with a friend who has a baby just a few months old. While the mother is busy with the baby, you have a dirty baby bottle in your hand and ask the waiter, who is your age, to wash it.*		*You just finished having lunch with your roommate and you have to go out together short after. You don't want to leave the dishes dirty and then, while you're clearing away the table, you ask him/her to wash them.*	
	Lo lavi? Do you wash it?	(after new contextual element) *Lavi la lampada?* Do you wash the lightbulb?	*Li lavi?* Do you wash it?	(after new contextual element) *Ah, dimenticavo, lavi la lampada?* Oh, I forgot, do you wash the lightbulb?

Main references

ALBANO LEONI, F. (2003). Tre progetti per l'italiano parlato: AVIP, API, CLIPS. In MARASCHIO, N. & POGGI-SALANI, T. (eds.). Atti del XXXIV Congresso Internazionale di Studi della Società di linguistica italiana (SLI), 19-21 ottobre 2000. Firenze: Bulzoni, 675-683.

BOERSMA, P.; WEENINK, D. (2022). Praat – Doing Phonetics by Computer, 6.3.06 version http://www.fon.hum.uva.nl/praat.

BROWN, P. & LEVINSON, S. C. (1987). Politeness: some universals in language use. Cambridge: Cambridge University Press.

LADD, R. (2008 [1996]). Intonational Phonology. Cambridge: Cambridge University Press.

LEVENSTON, E.; BLUM, S. (1978). Discourse completion as a technique for studying lexical features of interlanguage. Working Papers in Bilingualism 15, 13-21.

NUZZO, E. (2013). La pragmatica nei manuali d'Italiano L2: una prima indagine sull'atto linguistico del ringraziare. Revista de Italianística, XXVI(1), 5-29.

PIERREHUMBERT, J. (1980). The phonology and phonetics of English intonation. 1980. Tese de doutorado – MIT, Cambridge.

SEARLE, J. (1979). Expression and Meaning. Studies in the Theory of Speech Acts. Cambridge: Cambridge University Press.

SILVA NETO, M.; GILI FIVELA, B.; SANTORO, E.; FERNANDES-SVARTMAN, F. (2020). Do mitigation strategies affect prosodic correlates? An investigation on orders and requests in Italian. *Studi AISV 7*.

L2 Speech Perception with Auditory and Audiovisual Training

Federica Cavicchio, Mirko Grimaldi
CRIL, Università del Salento

Una delle principali cause che determinano l'accento straniero dei parlanti di una seconda lingua (L2) è da ascrivere alle difficoltà nel percepire le strutture fonologiche della L2. Studi percettivi hanno infatti dimostrato che gli studenti di L2 hanno difficoltà non solo con la produzione di segmenti fonologici non nativi, ma anche con la percezione di consonanti e vocali che sono foneticamente diverse nella lingua materna (L1, cfr. Flege, 2003 per una discussione). Studi comportamentali e neurofisiologici condotti su studenti universitari frequentanti il primo e il quinto anno del curriculum di lingua inglese e su un gruppo di ascoltatori non iscritti a corsi di lingua inglese, hanno dimostrato che le capacità di categorizzazione e discriminazione fonetica della L2 degli studenti non differivano significativamente da quelle dei soggetti naïve. L'istruzione L2 in classe nell'età adulta porta quindi all'assimilazione dei fonemi L2 alle categorie di fonemi della L1 degli studenti e non ad un miglioramento della discriminazione fonetica della L2 stessa (Grimaldi et. al., 2014).

Ricerche sull'apprendimento della L2 hanno suggerito numerose metodologie per facilitare la discriminazione e identificazione dei fonemi contrastivi L2 in studenti adulti. Molti studi hanno riportato che il metodo del training fonetico ad alta variabilità (*High Variability Phonetic Training*, HVPT) può sviluppare la discriminazione e identificazione delle vocali e delle consonanti L2 da parte degli studenti (cfr. Thomson 2018 per una revisione). Il metodo HVPT di solito consiste in più sessioni di training che consistono nella somministrazione di vocali e consonanti L2 incorporate in più contesti linguistici (ad esempio, CVC) e prodotte da più parlanti madrelingua della lingua target. Agli studenti L2 è chiesto di identificare o discriminare le vocali o le consonanti che ascoltano in ognuna delle sessioni di training. Viene inoltre fornito un *feedback* immediato ad ogni discriminazione o identificazione di coppie minime di fonemi in modo che gli studenti prestino attenzione alle proprietà acustiche che distinguono i suoni. Sebbene sia stato stabilito che l'uso del training audio è efficace nell'apprendimento fonetico della L2, è ancora poco studiato se l'uso di stimoli audiovisivi inseriti in un training HVPT possa aiutare l'identificazione o la discriminazione dei fonemi target in L2. È risaputo che nella L1 la componente visiva è correlata con le caratteristiche uditive e contiene segnali fonemici che aiutano ad apprendere le distinzioni dei fonemi (cfr. Bernstein et al. 2013).

In questo studio abbiamo utilizzato l'HVPT con stimoli audio e audiovisivi e abbiamo testato se le informazioni articolatorie visive possano aiutare a identificare i suoni L2 meglio del solo audio. Infatti, poiché è noto che le aree senso-motorie e uditive sono attivate congiuntamente durante l'acquisizione dei suoni del linguaggio L2 negli adulti (Kuhl et al. 2014), ipotizziamo che il training audiovisivo migliori l'accuratezza percettiva dei suoni della L2. Per verificare questa ipotesi, abbiamo scelto due contrasti vocalici del Standard Southern British English (SSBE): /e-æ/ e /ʌ-ɒ/. Uno studio uditivo precedente ha stabilito che i parlanti L1 dell'italiano parlato nel Salento (SI, Puglia meridionale) tendono a confondere la vocale SSBE /æ/ con le vocali L1 /a/ e /ɛ/ e le vocali SSBE /ʌ/ e /ɒ/ con le vocali L1 /a/ e /ɔ/ (Escudero, Sisinni e Grimaldi 2014). Inoltre, dal punto di vista articolatorio, i fonemi /e-æ/ si distinguono per l'apertura della bocca, mentre i fonemi /ʌ-ɒ/ si distinguono per l'arrotondamento delle labbra.

Abbiamo videoregistrato 40 pseudoparole (20 per la coppia minima /e/-/æ/ e 20 per la coppia minima /ʌ/-/ɒ/) prodotte da sei madrelingua di SSBE (tre maschi e tre femmine) di età compresa tra 35 e 56 anni. L'audio è stato registrato con microfono Speedlink collegato esternamente a una telecamera Sony HDR. La telecamera riprendeva la parte inferiore del viso del parlante, concentrandosi sui movimenti delle labbra e della lingua (cfr. Fig. 1a, *audiovisual*). Allo studio hanno preso parte 66 parlanti di SI (20 maschi, età media=19,5, livello CEFR inglese B1=52, A2=14). Tutti i partecipanti allo studio sono studenti del primo anno del corso di Lingue, culture e letterature straniere dell'Università del Salento e hanno ricevuto istruzioni riguardanti la fonetica acustica e articolatoria durante il corso di Linguistica generale. I parlanti selezionati sono tutti residenti in Salento da almeno 15 anni. Nessuno di loro è bilingue, e tutti hanno

studiato inglese solo in contesto scolastico. I partecipanti sono stati assegnati in modo casuale a tre gruppi sperimentali: training audio, audiovisivo o gruppo di controllo. I tre gruppi erano comunque bilanciati per genere.

Prima del training, gli studenti hanno eseguito un *pre-test* di identificazione di 160 stimoli audio (40 pseudoparole x 2 parlanti nativi x2 ripetizioni). I partecipanti assegnati ai gruppi di training audio e audiovisivo hanno preso parte a quattro sessioni di training ad intervalli settimanali durante i quali identificavano 160 pseudoparole prodotte da 2 nuovi parlanti e ricevendo *feedback* immediato. I partecipanti del gruppo di controllo hanno invece sostenuto a intervalli settimanali un test di comprensione del testo seguito da una serie di risposte a scelta multipla e *feedback* (cfr. Fig. 1a). Una settimana dopo la fine del training tutti i partecipanti hanno fatto un *post-test* di identificazione di stimoli audio, con la stessa procedura del *pre-test* e stimoli registrati da nuovi parlanti madrelingua.

I dati raccolti sono stati analizzati con una regressione logistica *mixed-effect*, con l'accuratezza dell'identificazione delle vocali come variabile dipendente e tipo di test (pre-test vs post-test), metodo di training (audiovisivo, audio o gruppo di controllo) e vocali (/e/, /æ/, /ʌ/ e /ɒ/) come variabili indipendenti. Sono stati inoltre introdotti effetti casuali per partecipanti e item (pseudoparole). I risultati del modello statistico hanno mostrato una differenza significativa tra il pre e il post test nell'identificazione delle vocali SSBE per i partecipanti al training audio rispetto ai partecipanti del gruppo di controllo (Est. = 0,77, SE=0,1, p<0,001) e tra il gruppo di training audiovisivo e il gruppo di controllo (Est. =0,97, SE=0,1, p<0,001) ma non tra i gruppi di training audiovisivo e audio (Est. = 0,2, SE=0,1, p=0,87, cfr. Fig. 1b). Esiste inoltre una differenza significativa tra l'accuratezza percettiva di /ʌ/ vs. /ɒ/ sia per il gruppo del training audiovisivo (Est. = 0,96, SE=0,14, p<0,001) che per il gruppo che ha seguito il training audio (Est.=0,9, SE=0,14, p <0,001), suggerendo che per chi parla SI la vocale SSBE più difficile da percepire è /ʌ/.

Concludiamo quindi che il training percettivo audiovisivo non migliora significativamente la percezione delle vocali L2 rispetto all'allenamento uditivo e che il modulo uditivo svolge un ruolo cruciale durante l'acquisizione delle vocali L2 negli studenti adulti. Indagini future si concentreranno sul ruolo dei training audio e audiovisivo sul miglioramento della produzione dei fonemi L2.

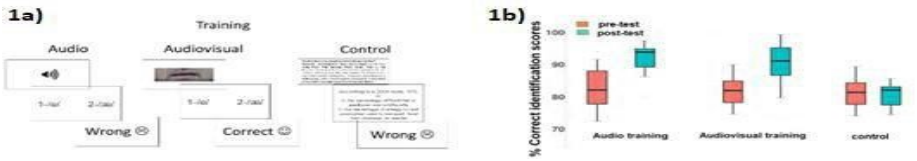


Figura 1a: Esempio della procedura sperimentale dei training audio e video e del gruppo di controllo. **Figura 1b:** Percentuale di risposte corrette nel *pre-test* di identificazione delle vocali SSBE a confronto con le percentuali di risposte corrette ottenute nel *post-test* dai partecipanti dei training audio, audiovisivo e del gruppo di controllo.

Bibliografia

Bernstein L. E., Auer E. T. Jr., Eberhardt S. P., Jiang J. (2013). Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Frontiers in Neuroscience* 7-34. DOI: 10.3389/fnins.2013.00034.

Escudero P., Sisinni B., Grimaldi M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *Journal of Acoustic Society of America*,135(3), pp. 1577-84. DOI: 10.1121/1.4864477.

Flege J. E. (2003). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production*, eds A. Meyer and N. Schiller (Berlin: Mouton de Gruyter), 319–357.

Grimaldi M., Sisinni B., Gili Fivela B., Invitto S., Resta D., Alku P., and Brattico E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. *Frontiers in Human Neuroscience*, 8. DOI: 10.3389/fnhum.2014.00279

Kuhl P. K., Ramirez R. R., Bosseler A., and Toshiaka I. (2014), Infants' brain responses to speech suggest Analysis by Synthesis, in *Proceedings of the National Academy of Science*, 5, 111(31), pp. 11238- 11245. DOI: 10.1073/pnas.141096311

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT). A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, 4(2), pp. 208 – 231.

Scraping YouTube audio data with Pyrlato: an example with Italian mica

Giuseppe Magistro, Claudia Crocco
Ghent University

Prosodic studies largely rely on speech elicited with a range of laboratory techniques aimed at collecting data of the highest acoustical quality while at the same time preserving the ecology of data as much as possible. However, the naturalness of the data is often sacrificed in the name of the acoustic standards, which are considered crucial for reliable analyses. Natural data are, however, fundamental for linguistic analysis as they represent the actual speakers' usage and are therefore vital to explore such usage and/or corroborate laboratory results. In this study we propose a pipeline to extract natural data from social media that are subsequently used to test the validity of previous observations on laboratory speech. The illustration case is offered by the negative particle *mica* in Italian.

The Italian negative particle *mica* covers a specific pragmatic function, inasmuch as it denies a predicate which is activated previously in the context (ex.1) (Cinque 1976).

(1) A: *Non andarci se è pericoloso!* 'Don't go there if it is dangerous!'
B: *Non è mica pericoloso.* 'It is not dangerous at all'

Recent research has defined more formally *mica* as Falsum operator (the opposite of Verum, remarking that a previous uttered or implied proposition is false) (Frana and Rawlins 2016 and subsequent), although a wider range of functions can be covered by *mica* such as the denial of the speakers' expectations or implicatures (Magistro 2022). Magistro and Crocco (2022) conducted a reading task experiment to explore the prosodic properties of *miga* in Venetan dialects and found that *miga* is pitch accented, as expected by literature on Verum/Falsum operators (Hoehle 1992, Gutzmann et al. 2020 for a criticism). This study, however, focused on the dialects and did not explore the prosodic properties of *mica* in Italian. Moreover, the ecological validity of the data collected through reading tasks has been questioned (Del Mar Vanrell et al. 2018), especially when it comes to the elicitation of Verum operators (Andorno and Crocco 2018). With this in mind, the aim of our paper is twofolds: we present a novel methodology to create a spoken corpus by scraping audio data on social media and use it to explore the prosodic properties of Italian *mica*.

We developed an executable program in Python to get audio files from YouTube, named *Pyrlato*. Pyrlato pipeline follows a series of steps: first, it uses the module PyTube to search for videos according to the user's input, so that the user can narrow the field of their research (e.g. a specific topic, speaker, context). In our case, we decided to focus on interviews and tv shows. Secondly, Pyrlato makes sure that the videos are in Italian and generates an object which possesses a method to obtain YouTube's automatically generated subtitles. Pyrlato then looks for specific words or sentences in the subtitles: in our case we searched for '*mica*'. If an instance is found, Pyrlato extracts and stores the relevant sentence(s) from the video using MoviePy, otherwise the video is discarded from the corpus. The program extracts a portion lasting 7 seconds, which is also customizable by the user. In case of longer or shorter duration, the sentences from our corpus were manually adjusted. The entire code will be soon available on GitHub.

At the time of writing, the corpus obtained through Pyrlato contains a total amount of 43 hours, where 434 instances containing *mica* were found (The computational time to run the program and obtain the total of 434 *mica* was almost 2 hours on a MacBook Pro M1). Among these, we removed all the examples of *mica* working as denial of constituent, which is beyond the scope of the paper. In addition, we removed all the extracts with bad sound quality (e.g. background music and bad microphones), obtaining a total of 290 usable instances. The sentences were visually and auditorily checked in Praat (Boersma & Weenink 2022).

The data partly confirm for Italian the results concerning Venetan *miga.* The syllable *mi* of Italian *mica* is stressed and pitch accented in the 93% of the corpus (fig.1). The tonal event usually takes the shape of a rising pitch accent, but specific details such as peak alignment and scaling are difficult to ascertain on uncontrolled material. However, in 7% of the occurrences *mica* is not pitch accented; the nuclear pitch accent is to be found elsewhere in the sentence (fig.2). While this effect may be explained as the result of phonetic undershooting, the contexts where these latter instances of *mica* appeared did not evoke any

corrective function of an activated proposition, but encoded a contrastive denial, without rejecting a proposition activated in the previous turn (e.g. the denial of an expectation, implicature or the speaker's own belief) (ex.2). This result corroborates the hypothesis that *mica* falls within a complex taxonomy of Falsum functions that are prosodically interpreted: in correction, it is pitch-accented, but it can also be used in less contrastive functions where prominence is not found anymore.

(2) Ora parliamo dell'uomo della luce [elettrica], che non è mica un fisico.
Now let's talk about the power man, who is not a physicist.

In conclusion, the use of acoustic data present in social media promises to offer a fertile test ground for exploration and corroboration of hypotheses. For the specific example of *mica,* the data offered empirical support to results on Venetan indicating that *mica* is a prosodically strong element. However, the corpus also unveiled other possible realisations, depending possibly on the type of denial or phonetic underspecification. We can conclude that the use of Pyrlato can enhance triangulation strategies across laboratory-like and more spontaneous data (cf. Mereu and Vietti 2021) aimed at strengthening the validity of analyses.

Figures

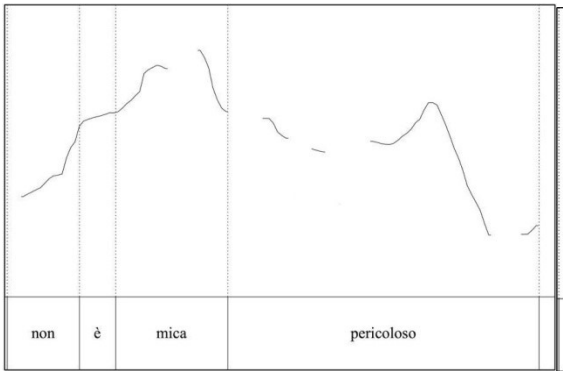


Fig.1 Intonation of statement (1)

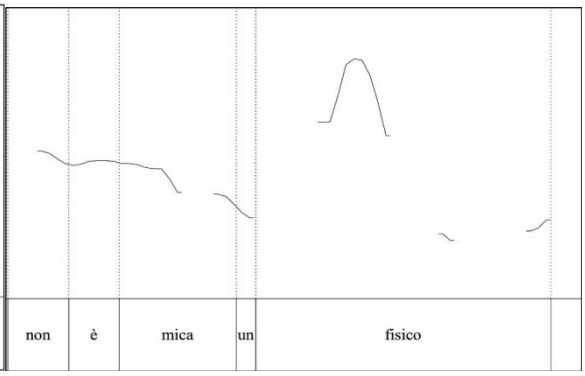


Fig. 2 Intonation of statement (2)

Selected References

Andorno, C., & Crocco, C. (2018). In search for polarity contrast marking in Italian. *The Grammatical Realization of Polarity Contrast: Theoretical, empirical, and typological approaches*, *249*, 255.

Cinque, G. (1976). Mica. In *Annali della Facoltà di Lettere e Filosofia dell'Università di Padova*, 101–12

del Mar Vanrell, M., Feldhausen, I., & Astruc, L. (2018). The Discourse Completion Task in Romance prosody research: Status quo and outlook. *Methods in prosody: A Romance*, 191

Frana, I., & Rawlins, K. (2016). Italian 'mica'in assertions and questions. In *Proceedings of Sinn und Bedeutung* (Vol. 20, pp. 234-251).

Gutzmann, D., Hartmann, K., & Matthewson, L. (2020). Verum focus is verum, not focus: Cross-linguistic evidence. *Glossa: a journal of general linguistics*, *5*(1)

Höhle, T. N. (1992). Über Verum-Fokus im Deutschen. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, 112–141. Opladen: Westdeutscher Verlag.

Magistro, G., & Crocco,C. (2022). Phonetic erosion and information structure in function words: the case of mia. In *The 2022 Conference of the International Speech Communication Association*.

Magistro, G. (2022). Mica preposing as focus fronting. Glossa- A Journal of General Linguistics, 7(1).

Mereu, D., & Vietti, A. (2021). Dialogic ItAlian: the creation of a corpus of Italian spontaneous speech. *Speech Communication*, *130*, 1-14.

Approximating and lateral palatals in Italian: an acoustic and perceptual study of young speakers in Turin

Anna Anastaseni[1,2] Antonio Romano[1]
[1] LFSAG, Università di Torino; [2] Gipsa-lab, CNRS/Università Grenoble-Alpes

L'inventario fonologico dell'italiano distingue la laterale palatale /ʎ(ʎ)/(da qui /ʎ/) e l'approssimante palatale /j/; tuttavia tale distinzione è a basso rendimento funzionale. La laterale ha un complesso meccanismo articolatorio (Bladon & Carbonaro 1978; Canepari 1992; Quilis & Fernández 1969; Recasens 1984), e compare tardi nell'inventario fonologico dei bambini italofoni, per questo viene spesso semplificata (Galatà et al., 2012; Zampaulo & Haug, 2015).

Il fonema /ʎ/ è raro nelle lingue del mondo; è presente solo in 20 delle 451 considerate in UPSID (UCLA Phonological Segment Inventory Database). Inoltre ricorre solo in 16 lingue nel cui inventario compaia anche l'approssimante palatale /j/ (presente, al contrario in ben 378 lingue del database): una perdita di contrasto è infatti favorita dalla somiglianza acustica, come mostra l'evoluzione dei sistemi fonologici di altre aree romanze in cui è endemico (Porras 2013). Diversamente dallo spazio accordato alle proprietà articolatorie di questi suoni (Recasens et al. 1993, Calamai & Bertinetto 2006), nella maggior parte della manualistica sperimentale è data una descrizione acustica solo di [ʎ], spesso senza esplicitare i parametri che lo distinguerebbero [j] (in molti casi non trattato per via del suo incerto statuto fonologico). Data la debole possibilità di una distinta caratterizzazione acustica tra /ʎ/ e /j/ e la generale tendenza alla neutralizzione tra i due fonemi in certe aree (storicamente alcuni focolai nel sud- Italia; cfr. Canepari 1999), è stato avviato un progetto di studio acustico e percettivo. L'obiettivo principale è indagare i confini categoriali dell'approssimante palatale e la laterale palatale e le caratteristiche acustiche (sia in produzione che in percezione) che aiutano a distinguerle. Si è scelto di focalizzare la nostra attenzione su parlanti piemontesi, per verificare se in una regione nella quale non ci si aspetterebbe la neutralizzazione di questi suoni si attesti tra i parlanti un certo grado di confusione e se questo abbia delle conseguenze nella conversione fono- grafemica.

Dopo una prima raccolta di registrazioni di parole e non-parole con l'aiuto di due speaker professionisti, sono stati somministrati a un campione di 113 studenti che frequentano l'Università di Torino, che hanno compilato un questionario sociolinguistico, due task percettivi di identificazione miranti a valutare l'incidenza del fattore durata nella distinzione di segmenti intervocalici di tipo [ʎ], [j] e [lj]. Si è scelto utilizzare una lista di pseudo-parole affinché le variabili lessicali non incidessero nel riconoscimento dei foni. Partendo dai logatomi originari, sono stati costruiti stimoli con segmenti intervocalici di durata diversa avendo cura di preservare le porzioni di transizione da un fono all'altro. Questi primi due task pilota somministrati tramite la piattaforma Folerpa (Fernández Rei et al. 2021) hanno mostrato che i partecipanti hanno avuto più difficoltà nella classificazione di /ʎ/ e /j/, rispetto a /lj/ e /l:j/. Secondariamente è risultato che il tratto durata ha un ruolo significativo nel corretto riconoscimento di /ʎ/ (r di Pearson=0,947**, p=0,01): al di sotto dei 140-160 ms il fono percepito è infatti maggioritariamente ricondotto a una risposta di tipo ‹i› (associata a /j/). Nel caso di durate progredienti di un fono originariamente di tipo [j] invece sembrerebbe esserci una soglia attorno ai 200 ms: il fono percepito se la durata è superiore a questo valore non è però sistematicamente riconosciuto come quello corrispondente alla grafia ‹gl›. Visti gli esiti di questi due primi esperimenti, si è scelto di approfondire ulteriormente, con l'obiettivo anche di fornire una descrizione acustica dei due foni.

Per questo si è deciso di proporre un task di lettura ad alta voce, un dettato di pseudo-parole con ripetizione di ciascuno stimolo dettato e un task di discriminazione uditiva. I task sono stati proposti a 60 parlanti piemontesi dai 15 ai 30 anni, divisi in classi di età (15,20](20-25](25-30]. Sono stati selezionati solo parlanti prevalentemente monolingui e che hanno indicato come regione di provenienza della propria famiglia il Piemonte. Si è scelto di proporre il task a giovani ipotizzando che la neutralizzazione tra /ʎ/ e /j/ sia un fenomeno in espansione, in particolare in Piemonte dove non ci si aspetta questo tratto come

autoctono. Ai soggetti sono stati proposti un questionario sociolinguistico e un questionario di autovalutazione volto a identificare difficoltà di letto-scrittura, poiché i DSA sono spesso associati deficit fonologici e visto che il compito di identificazione e di dettato prevedono la conversione grafo-fonematica e fono-grafemica. Tutte le registrazioni si sono svolte presso l'LSFAG ed è stato usato il software OpenSasame.

Dai dati collezionati dal task di lettura ad alta voce e dal task di dettato e ripetizione è emerso che in tutte le classi di età la percentuale di partecipanti che neutralizza le realizzazioni di /ʎ/ e /j/ è superiore al 20% e segue un andamento crescente. Le registrazioni dei due task sono state confrontate con le produzioni dei due professionisti (M e F).

L'analisi acustica delle produzioni di questi mostra realizzazioni che presentano caratteristiche diverse in termini di valori assoluti, ma movimenti formatici relativamente simili. Il contrasto tra [j] e [ʎ], pur distintamente percepibile in entrambi, è maggiormente affidato a strategie di abbassamento/convergenza tra formanti nel caso di F e a indici di durata più solidi per M. Le dinamiche sovrapponibili delle prime tre formanti della speaker suggeriscono invece che la distinzione possa essere affidata a contrasti di energia (si osservano infatti profili di ampiezza diversi) e indici di costrizione localizzati però in distinte posizioni, in particolare in fase di attacco e rilascio. Queste considerazioni sono state ritrovate *mutatis mutandis* nei confronti tra i dati acustici che caratterizzano coppie di non-parole nelle produzioni di 19 dei giovani parlanti torinesi in grado di differenziare [ʎ] e [j]. Ne è emerso che le strategie usate per differenziare le due palatali sono molto differenti tra i parlanti, e a volte anche nelle singole produzioni dei diversi stimoli. In questo campione sembrerebbe che /ʎ/ si realizzi più spesso con formanti F2, F3 e F4 leggermente più basse (il contrario di quanto riscontrato negli speaker di professione) e con rumori di frizione nell'ultima fase, con profilo di ampiezza più brusco; ma, a conferma di quanto emerso dai task, l'unico parametro che regge il contrasto per tutti i locutori è la durata. Il tratto di durata risulta anche qui un indice significativo, del resto come dimostrato da Mairano & De Iacovo (2020) anche le varietà di italiano settentrionale conoscono la geminazione. Sembra quindi che l'associazione [ʎ]~durata lunga, [j]~durata breve, sia una costante nella produzione, così come in percezione come è emerso dal task di identificazione percettiva.

Per quanto riguarda invece i risultati del task di dettato, sono fortemente in linea con i risultati del task di identificazione percettiva. I dati sull'accuratezza ortografica nel dettato sono stati incrociati con la pronuncia dei partecipanti. Ne emerge che coloro che neutralizzano i due fonemi nella realizzazione orale, hanno anche maggiori difficoltà nella conversione grafo-fonematica delle pseudo-parole. Questo dato supporta l'idea che per questi ultimi la corrispondenza tra [ʎ] e ‹gl(i)›, e tra [j] e ‹i› non sia del tutto trasparente. Inoltre, se si entra maggiormente nel dettaglio analizzando le percentuali di risposte target distribuite per fono bersaglio, si scopre che per quanto riguarda le pseudo-parole con [j] di durata più alta (in particolare 260 ms) la discrepanza tra i due gruppi (chi neutralizza e chi distingue nella pronuncia) è molto evidente (56,82% risposte target vs 8,93%). Per il gruppo di chi neutralizza sembra che non entrino in conflitto parametri diversi (la qualità del fono e la sua durata) nel riconoscimento di [j] con durata non prototipica, ma che i parlanti si affidino unicamente alla durata del fono per la sua classificazione, associando saldamente [j] 260 ms alla grafia ‹i›.

In conclusione i dati fin qui presentati confermano le difficoltà nella classificazione, discriminazione e conversione grafofonematica di /ʎ/ e /j/. La distinzione tra i due foni inoltre non troverebbe sistematici e sufficienti indici acustici. Tuttavia i dati di riproduzione e i risultati di alcuni task di percezione suggeriscono come un indice di organizzazione temporale sia un prezioso correlato percettivo, almeno nella popolazione di parlanti italiani considerata.

Riferimenti Bibliografici

Bladon, R., & Carbonaro, E. (1978). Lateral consonants in Italian. *Journal of Italian Linguistics*, 3(1), 43-54.

Calamai S. & Bertinetto P.M. (2006). "Per uno studio articolatorio dei legamenti palatale, labiovelare e labio-palatale dell'italiano". In: V. Giordani *et alii* (a cura di), *Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche*, Torriana (RN): EDK, 43-56.

Canepari, L. (1999). *Il MaPI. Manuale di pronuncia italiana*. Bologna: Zanichelli (1ª ed. 1992).

Fernández Rei E., Aguete Cajiao A., Osorio Peláez C. & Cutrín Garabal J.A. (2021). FOLERPA: Ferramenta On- Line para ExpeRimentación PerceptivA. Santiago de Compostela: Instituto da Lingua Galega.

Galatà, V., Meneguzzi, G., Conter, L., & Zmarich, C. (2012). Primi dati sull'acquisizione fonetico-fonologica dell'italiano L2 in prescolari rumeni. In: A. Paoloni & M. Falcone (a cura di), *La voce nelle applicazioni,* Roma: Bulzoni, 35-50.

Mairano, P. & De Iacovo, V. (2020), *Gemination in Northern versus Central and Southern Varieties of Italian: A Corpus- based Investigation*, Language and Speech, Vol. 63(3) 608–634

Pettorino M. & Giannini A. (1991). Indagine acustica sulle approssimanti dell'italiano. Atti del XIX convegno AIA (Napoli 10-12 aprile 1991), 441-447.

Porras, J. E. (2013). Spanish Yeísmo: A Cognitive Linguistic Approach to Phonological Change. In: R. Gómez & Molina Martos (eds.), *Variación yeísta en el mundo hispánico*, Iberoamericana Editorial Vervuert, 335-352.

Quilis A. & Fernández J.A. (1969). *Curso de fonética y fonología españolas para estudiantes angloamericanos*. 4. Vol. 2. Ma- drid. Istituto «Miguel Cervantes».

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *The Journal of the Acoustical Society of America*, 76(6), 1624-1635.

Recasens D., Farnetani E., Fontdevila J. & Pallarès M.D. (1993). "An electropalatograpic study of alveolar and palatal consonants in Catalan and Italian". *Language and Speech*, 36 (2-3), 213-234.

Zampaulo, A., & Haug, D. (2015). The evolution of the (alveolo) palatal lateral consonant in Spanish and Portuguese. In: D. Haug (ed.), *Historical Linguistics 2013*, Amsterdam: John Benjamins, 69-86.

Calibration and Fusion of Parametric and Non-parametric Statistical Methods for Forensic Semi-Automatic Speaker Recognition Systems: experimental results

Francesco Sigona, Mirko Grimaldi
University of Salento – CRIL-DReAM, Lecce, Italy

The aim of this study is to assess the performance of two Forensic Semi-Automatic Speaker Recognition (FSASR) approaches traditionally used in Italy (along with some variants), namely IDEM/SPREAD (first appeared in [1]; see [2] for a discussion and further references) and SMART ([3, 4, 5]). Both systems rely on statistical representations (parametric vs non-parametric) of the fundamental and the first three formant frequencies (F0, F1, F2, F3) of the Italian vowels /a, e, i, o/ (16 features altogether). Both the approaches have been emulated into the IMPAVIDO software ([6]), and the assessment is performed by means of score *calibration* and *fusion* with Logistic Regression (LogReg, [7]). *Calibration* is a procedure that is used to compute the actual Likelihood Ratio (LR) value (between the prosecutor and defendant hypothesis) of the casework at hand, starting from the raw score output from the last stage of the comparison. The same term, *calibration*, is often used also to indicate a property of a set of LR values, which can be measured, and it is strictly related to the accuracy of a Forensic Speaker Recognition (FSR) system ([8]). In general, if the same casework can be processed by two or more FSR systems, their output scores may be combined into a single outcome. This can be achieved by a *fusion* procedure, which aims at producing a compound system that performs better than the single ones. The metric used to assess the system accuracy was the likelihood-ratio cost (Cllr [9, 10]), which represents the total probability of error, at any possible prior, that the fact finder would obtain by using the LR value computed by the forensic expert, in the Bayesian decision framework. Thus, the less the Cllr, the more accurate is the system, and the Cllr computed on calibrated scores, i.e. on LR, values, is expected to be smaller than the Cllr computed on uncalibrated raw scores.

The factors of analysis include: the size of the features datasets (SPREAD's vs IMPAVIDO's) and different software implementations of LogReg (namely: a) the Focal toolkit [11]; b) the Bosaris toolkit [12]; c) a variant by Morrison [13] of the LogReg training function of the Focal Toolkit ([14]) and d) the *mnrfit* Matlab® function for multinomial regression). The assessment has been done with cross-validation approach, while the comparison between the different set of results has been done by means of paired samples t-test (1 tail), α =5% on the Cllr values.

The first finding is that, according to expectations, the LogReg calibration has improved the Cllr in every circumstance, suggesting that also the accuracy of the tested FSASR approaches may be improved introducing calibration. The second finding is that the tested FSASR approaches give more accurate results when the larger dataset (by IMPAVIDO) is used. This may be due to the larger number of speakers and the required number of vowel occurrences with respect to the selected IDEM/SPREAD dataset. The best result (Cllr=0.004) has been achieved with the Bosaris toolkit with an eIDEM variant, using all the 4 vowels, with the IMPAVIDO database. Regarding the fusion strategy, it has been found that for the eSMART emulator is always better to let it work on each vowel separately and then fusing the 4 resulting scores, instead of using the 4 vowels together. This is important because the same may possibly apply also to the actual SMART software. A specific study on that software is therefore highly suggested. Finally, for each fusion that has been tested, the compound systems were sometimes able to outperform their single components, but no one was able to outperform the best single system (above mentioned). This may be explained by the fact that, despite the difference in the statistical approaches between eIDEM and eSMART, both systems work on the same features, which may be somehow redundant. From this point of view, a future direction of investigation might be to fuse a FSASR system based on a different kind of features, or, even better, to fuse a semi-automatic with fully automatic approach.

The results of these experimentations suggest that a more detailed study would be necessary in order to explore the behavior of such approaches depending on other factors such as the numerosity of the feature sets to compare (i.e. the number of occurrences for each vowel) and possibly the noise conditions. Reducing the requirements on vowel quantity, at least for anonymous sample, can help to simulate those caseworks

where very limited data are available: this is a quite simple operation. On the other hand, increasing that prerequisite, or even replicate the same experiment with introducing a sufficiently large number of noisy speech recordings, is a challenging and very time-consuming task because it would require to accurately label a very large number of vowels.

Also, both the tested databases do not distinguish between different sessions of a same speaker; neither have any other meta-data such as noise level. On the other hand, calibration can be influenced by the actual conditions in which the speech recordings have been produced. In the present study we have implicitly assumed that the features values in both the tested databases where representative of typical casework conditions, where the signal-to-noise ratio (SNR) was above the threshold commonly accepted by practitioners to extract spectroacoustic measurements of vowels (10 dB). But a deeper study of the impact of calibration on the tested FSASR would probably add more details to the picture.

References

[1] FALCONE, M., PAOLONI, A., DE SARIO, N., SAVERIONE, B. (1992). IDEM: un sistema per l'analisi e la rappresentazione vocale. *Associazione Italiana di Acustica, XX Congresso Nazionale*, Rome, 20-24 Aprile, pp. 417-422.

[2] SIGONA F., GRIMALDI M. (2015). Il riconoscimento del parlante in ambito forense: uno studio indipendente sul software IDEM/SPREAD in uso ai Carabinieri. In *Sicurezza e Giustizia*, IV, 16– 21.

[3] BOVE, T., GIUA, P. E., FORTE, A. & ROSSI, C. (2002). Un metodo statistico per il riconoscimento del parlatore basato sull'analisi delle formanti. In *Statistica*, 62(3), 475-490.

[4] LASINIO, J. G., C. ROSSI & T. BOVE (2004). The Speaker Recognition problem. *Proceedings of the XLII Scientific Meeting of the Italian Statistical Society (SIS)*, Palermo, 20-22 June, 429–440.

[5] ROMITO, L., BOVE, T., DELFINO, S., ROSSI, C., LASINIO, G. J. (2009). Specifiche linguistiche del database utilizzato per lo Speaker Recognition in S.M.A.R.T. In Romito, L., Lio, R., Galatà, V. (Eds.), *La Fonetica Sperimentale: metodo e applicazioni*, Torriana: EDK Editore, p. 632–640.

[6] SIGONA F., GRIMALDI M. (2017). Tools for Forensic Speaker Recognition, in Orletti, F., Mariottini, L. (Eds.),,Forensic Communication: Theory and Practice, Cambridge: Cambridge Scholars Publishing, 127-150.

[7] MORRISON, G. S., (2013). Tutorial on logistic-regression calibration and fusion: converting a score to a likelihood ratio. In *Australian Journal of Forensic Sciences*, 45(2), 173-197.

[8] RAMOS, D. & GONZALEZ-RODRIGUEZ, J. (2013). Reliable support: Measuring calibration of likelihood ratios. In *Forensic Science International*, 230(1–3), 156–169.

[9] BRÜMMER, N. & DU PREEZ, J. (2006). Application-independent evaluation of speaker detection. In *Computer Speech and Language*, 20(2-3), 230–275.

[10] VAN LEEUWEN, D.A., BRÜMMER, N. (2007). An introduction to applicationindependent evaluation of speaker recognition systems. In Müller, C. (Ed.), *Speaker Classification I: Fundamentals, Features, and Methods*., Heidelberg: Springer-Verlag, 330–353.

[11] Niko Brummer, "Focal toolkit," available in https://sites.google.com-/site/nikobrummer/focal.

[12] E. de Villiers and N. Brummer, "BOSARIS toolkit," https://sites.google.com-/site/bosaristoolkit.

[13] http://geoff-morrison.net.

[14] MORRISON, G. S. (2011). A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model-universal background model (GMM-UBM). In *Speech Communication*, 53(2), 242–256.

Degemination in Tuscan rhotics: a preliminary investigation of a historical oral archive

Roberta Bianca Luzietti [1], Silvia Calamai[2]
[1]*Università di Pisa,* [2]*Università di Siena*

Il lavoro presenta i primi risultati di un'indagine sociofonetica con oggetto lo studio del fenomeno residuale di degeminazione delle consonanti rotiche (/rr/) nell'area toscana pratese, attraverso il riuso di un archivio orale del passato, inizialmente raccolto per altri scopi (non linguistici).

Le consonanti rotiche sono dal punto di vista fonetico caratterizzate da un'ampia e nota variabilità riguardo i modi e tipi di articolazione a livello sia intra- che inter- linguistico (Hall 1997, Solé 2002, van Hout e Van de Velde 2001, Wiese 2001, 2010). Di conseguenza, le consonanti rotiche sono difficilmente definibili o classificabili come una classe univoca di suoni (come, ad esempio, quelli occlusivi) e, rispetto ad altre consonanti, la letteratura (specialmente fonetica) a riguardo è al quanto scarsa (Maddieson 1984, Lindau 1985, Ladefoged e Maddieson 1996, Catford 2001, Foulkes e Docherty 2001, Scobbie 2006, Chatbot 2019). Nel sistema consonantico italiano il contesto scempio e geminato (in posizione intervocalica) si distinguono per il tratto di lunghezza (temporale della consonante) e nel caso delle rotiche anche per il/in base al tipo di consonante: tap o trill (Canepari 1980, Vietti et al. 2010, Romano, 2013, Celata et al 2016).

Rispetto al passato, studi recenti hanno messo in luce che il fenomeno di degeminazione è un tratto sempre meno comune sia a nord che a sud dell'isoglossa La Spezia - Rimini, a causa della progressiva standardizzazione della lingua (Mairano e De Iacovo 2020). Nelle zone dell'Italia centrale il fenomeno ha riguardato alcune consonanti, tra cui, appunto, le rotiche (Rohlfs 1937, Giannelli 1976, Loporcaro 2009, Nodari e Meluzzi 2020, Meluzzi e Celata 2020). In Toscana, la degeminazione si configura come aspetto residuale, derivante dal percolamento di questo tratto dai dialetti settentrionali verso le zone a sud dell'isoglossa (Rohlfs 1937, Giannelli 1976, Giannelli 1983). Tuttavia, il fenomeno risulta essere scarsamente indagato dal punto di vista sia della distribuzione areale (diffuso a macchia di leopardo) che sociofonetico e culturale, pertanto, ben si presta a una verifica su archivi orali del passato. I contributi che hanno recentemente indagato questo fenomeno sono Celata et al. (2019) in due città ad ovest della regione (Pisa e Livorno) e Nodari e Calamai (2021) nell'isola d'Elba. Questo lavoro si concentrerà, invece, sull'area rurale fiorentina-pratese (Giannelli 1976), che dal punto di vista geolinguistico, si trova appena al di sotto del confine con l'isoglossa attraverso cui si è diffuso il fenomeno di degeminazione e, dal punto di vista storico, presenta poche documentazioni linguistiche sulla varietà locale, avendo subito per diversi secoli una forte influenza del fiorentino (Fiorelli 1980).

Il lavoro prevede l'analisi del fenomeno della degeminazione della consonante vibrante /rr/ in posizione intervocalica a livello lessicale nelle interviste contenute nell'archivio orale raccolto dalla ricercatrice Angela Spinelli nei primi anni 80 del Novecento in tre piccoli paesi dell'Alta val di Bisenzio, zona rurale di Prato (Cantagallo, Vaiano e Montemurlo). L'archivio è composto da serie di interviste faccia a faccia con circa 40 famiglie, corrispondenti a più di 120 ore di registrazione, svolte con lo scopo di raccogliere e conservare la memoria degli abitanti di quelle zone riguardo ai fatti storici avvenuti durante il periodo successivo alla Seconda guerra mondiale.

A partire dalle registrazioni delle interviste è prevista la creazione di un corpus per mezzo dell'ascolto, trascrizione e individuazione di parti di registrazioni di interesse per la ricerca contenenti parole target, relativamente frequenti dato l'argomento e i profili sociali dei parlanti (es. terra, guerra, ferro ecc.).

Ad eccezione delle informazioni di nome e cognome (per garantire l'anonimizzazione e riservatezza dei dati), per ciascun soggetto intervistato verranno raccolte le variabili sociali che permetteranno di ricostruire la rete sociale degli abitanti dei vari comuni, quali l'età, il genere, l'occupazione, il grado di istruzione, il comune di origine, l'eventuale parentela con gli altri intervistati e informazioni su eventuali esperienze migratorie da altri comuni o verso Prato. A partire, invece, dalle parole target verranno estratte le variabili linguistiche, quali la posizione della parola nella frase, la lunghezza della parola (in termini di sillabe), la

posizione della consonante vibrante rispetto all'accento della parola, il tipo di consonante (geminata o scempia) e, in fine, il timbro della vocale precedente e successiva alla vibrante.

Per l'analisi acustica, al momento in fase di svolgimento, è prevista l'estrazione dei valori/parametri di durata delle parole target, delle consonanti rotiche e delle vocali precedenti per mezzo del software PRAAT. L'analisi mira a verificare se i fenomeni di scempiamento si concentrano su specifiche categorie di parlanti (p.e. maschi che lavorano la terra, già anziani al momento dell'intervista) e se hanno un collegamento con la frequenza lessicale della parola. Grazie alla struttura delle interviste e alle modalità di selezione dei soggetti intervistati da parte della ricercatrice, sarà anche possibile approfondire il ruolo che la rete sociale dei parlanti può aver avuto in rapporto alla presenza di degeminazione nelle comunità rurali pratesi.

L'accesso ai dati d'archivio sarà possibile tramite la piattaforma Archivio Vi.Vo. (Calamai et al. 2021) parte dell'infrastruttura italiana di CLARIN(-IT), sviluppata per garantire la conservazione e il riuso di archivi orali storici per vari scopi (ricerca, disseminazione, valorizzazione ecc.) in sicurezza. La futura consultazione dell'archivio per mezzo della piattaforma permette, infatti, di consultare ed estrarre i dati (per l'analisi fonetica e sociolinguistica) limitandone la perdita, la frammentazione e il danneggiamento. (Al tempo stesso, questo lavoro permette di validare la capacità della piattaforma di ospitare diversi tipi di archivi per diversi tipi di utilizzo.)

La decisione di riutilizzare i dati orali che compongono questo archivio è motivata da due fattori. In primo luogo, la struttura delle interviste per la raccolta delle testimonianze ha favorito l'uso di un linguaggio poco controllato e più spontaneo (Labov 1967, 1972), sempre molto difficile da elicitare in laboratorio. In secondo luogo, essendo composto da registrazioni del passato, l'archivio permette di svolgere un'analisi in diacronia, indagando fenomeni linguistici, quali appunto la degeminazione di /rr/, ad oggi poco presenti e/o quasi del tutto scomparsi nel territorio toscano/pratese.

Bibliografia

Calamai, S., & Bertinetto, P. M. (2014). *Le soffitte della voce. Il progetto Grammo-foni* (pp. 1-213). Vecchiarelli Editore srl. Calamai, S., Pretto, N., Stamuli, M. F., Piccardi, D., Candeo, G., Bianchi, S., e Monachini, M. (2021). "Community-based Survey and Oral Archive Infrastructure in the Archivio Vi.Vo. Project". *Selected papers from the CLARIN Annual Conference 2020*. Linköping Electronic Conference Pro- ceedings 180: 180 55–64.

Canepari, L. (1980). *Italiano standard e pronunce regionali*. Cleup.

Catford, John C. (2001). "On Rs, rhotacism and paleophony". *Journal of the International Phonetic Association* 31(2). 171–185. Celata, C., Meluzzi, C., & Ricci, I. (2016). "The sociophonetics of rhotic variation in Sicilian dialects and Sicilian Italian: corpus, methodology and first results". Loquens, 3(1), e025-e025.

Celata, C., Vietti, A., Spreafico, L. (2019), "An articulatory account of rhotic variation in Tuscan Italian: synchronized UTI and EPG data". In Mark Gibson/Juana Gil, eds., Romance phonetics and phonology, Oxford, University Press: 92–117.

Chabot, A. (2019). "What's wrong with being a rhotic?". Glossa: a journal of general linguistics, 4(1).

Fiorelli, P. (1980) "Il linguaggio dei pratesi." in Storia di Prato, Prato, Edizioni Cassa di Risparmi e Depositi, III

Foulkes, P. & Docherty, G. (2001). Variation and change in British English. In Hans Van de Velde & Roeland van Hout (eds.), 'r- atics: sociolinguistic, phonetic and phonological characteristics of /r/, 27–43. Brussels: ILVP.

Giannelli L. (1976). Toscana. PDI 9. Pisa: Pacini.

Giannelli, Luciano (1983), "Considerazioni sullo stato del rotacismo di l preconsonantico nell'Italia centrale". *Quaderni dell'Istituto di Linguistica dell'Università di Urbino* 1: 135–154.

Hall, T. A. (1997). *The phonology of coronals*. Amsterdam: John Benjamins

Labov, William (1963), "The social motivation of a sound change". Word 19(3): 273–309.

Labov, William/Waletzky, Joshua (1967), "Narrative analysis: Oral versions of personal experience". *Journal of Narrative & Life History* 7(1–4): 3–38.

Labov, William (1972), *Sociolinguistic patterns* (No. 4). Philadelphia, University of Pennsylvania Press.

Labov, William (2013), *The language of life and death: The transformation of experience in oral narrative*. Cambridge, University Press.

Ladefoged, P., & Maddieson, I. (1996). "The sounds of the world's languages". Language, 74(2), 374-376.

Lindau, M. (1985). ''The story of /r/,'' in Phonetic Linguistics, *Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic, New York).

Loporcaro, M. (2009). *Profilo linguistico dei dialetti italiani*. Roma/Bari: Laterza.

Maddieson, I. (1984). "The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis". Journal of Phonetics, 12(1), 9-15.

Mairano, P., De Iacovo, V. (2020). Gemination in northern versus central and southern varieties of Italian: A corpus-

based investigation. Language and Speech, 63(3), 608-634.
Meluzzi, C., & Celata, C. (2020). Style and addressees in the selection of dialectal vs. standard phonetic forms: Sicilian rhotics. In Quaderns d'Italià 25, 137-154
Milroy, L., & Milroy, J. (1992). "Social network and social class: Toward an integrated sociolinguistic model". *Language in society*, *21*(1), 1-26.
Nodari, R., & Meluzzi, C. (2020b). Spie sociolinguistiche e percezione della provenienza geografica nell'italiano di Roma. Studi e Saggi Linguistici, 58 (2), pp. 65 – 98.
Nodari, R., & Calamai, S. (2021). "Degemination in marginal Tuscan speech: temporal analysis in legacy speech data". In D. Recasens, & F. Sánchez-Miret (a cura di), *Sound change in Romance: phonetic and phonological issues* (pp. 67-85). Muenchen: Lincom. Rohlfs G. 1937a. *La struttura linguistica dell'Italia*. Leipzig: Keller
Romano, A. (2013). "A preliminary contribution to the study of phonetic variation of/r/in Italian and Italo-Romance". *Rhotics: New data and perspectives*, 209-225.
Scobbie, James (2006), "(R) as a variable". In Keith Brown, ed., *Encyclopedia of language & linguistics*. 2nd edition. Oxford, Elsevier: 337–344.
Solé, Maria-Josep. (2002). "Aerodynamic characteristics of trills and phonological patterning". Journal of Phonetics 30.655-688. Van de Velde, H. & Van Hout, R. (2001). '*r-atics: sociolinguistic, phonetic and phonological characteristics of /r/*. Bruxelles: ILVP. Vietti, A., Spreafico L., & Romano A. (2010). "Tempi e modi di conservazione delle /r/ italiane nei frigoriferi CLIPS". In Stephan Schmid, Michael Schwarzenbach & Dieter Studer (eds.), *La dimensione temporale del parlato*, 113-128. Torriana: EDK.
Wiese, R. (2001). "The Phonology of /R/". In T. Alan Hall (Ed.), *Distinctive Feature Theory*. (pp. 335- 368). Berlin: Mouton de Gruyter.
Wiese, R. (2010). 'The representation of rhotics', in Mark van Oostendorp, Colin J. Ewen, Elisabeth Hume, and Keren Rice (eds), *The Blackwell Companion to Phonology*. Malden, MA: Wiley-Blackwell.