



# LA LINGUISTICA FORENSE IN PROSPETTIVA MULTIDISCIPLINARE




FORENSIC LINGUISTICS IN A MULTIDISCIPLINARY PERSPECTIVE



a cura di  
CHIARA MELUZZI E SONIA CENCESCHI



Milano 2023



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

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

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

*Curatore/Editor*

Cinzia Avesani (CNR-ISTC)

*Curatori Associati/Associate Editors*

Franco Cutugno (Università di Napoli), Barbara Gili Fivela (Università di Lecce), Daniel Recasens (Università di Barcellona), Antonio Romano (Università di Torino), Mario Vayra (Università di Bologna).

*Comitato Scientifico/Scientific Committee*

Giuliano Bocci (Università di Siena), Silvia Calamai (Università di Siena), Mariapaola D'Imperio (Rutgers University), Giovanna Marotta (Università di Pisa), Stephan Schmid (Università di Zurigo), Carlo Semenza (Università di Padova), Alessandro Vietti (Libera Università di Bolzano), Claudio Zmarich (CNR-ISTC).

© 2023 AISV - Associazione Italiana Scienze della Voce

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



Edizione realizzata da  
Officinaventuno  
Via F.lli Bazzaro, 18  
20128 Milano - Italy  
email: [info@officinaventuno.com](mailto:info@officinaventuno.com)  
sito: [www.officinaventuno.com](http://www.officinaventuno.com)

ISBN edizione cartacea: 978-88-97657-60-6  
ISBN edizione digitale: 978-88-97657-61-3  
ISSN: 2612-226X

## Sommario

CHIARA MELUZZI, SONIA CENCESCHI  
La linguistica forense e la sfida della multidisciplinarietà:  
introduzione al volume 5

LUCIANO ROMITO  
La Linguistica Forense: Prefazione 9

### PARTE I

#### *Approcci alla ricerca in linguistica forense*

CLAUDIA ROBERTA COMBEI  
The Multimodal Expression of Denial: A Case Study on Femicide Suspects 27

LUCIA BUSO  
*CorIELLS*: a specialised bilingual corpus of English and Italian  
legal-lay language 45

FRANCESCO SIGONA, SONIA D'APOLITO, COSIMO IAIA, BARBARA GILI FIVELA,  
MIRKO GRIMALDI  
Forensic Automatic Speaker Recognition with dialectal speakers:  
a pilot study on the Taranto and Brindisi varieties 65

UMAR MUHAMMAD-GOMBE, PETER FRENCH, ELEANOR CHODROFF  
A Comparative Analysis of Nigerian Linguistically-Naïve Native Speakers  
and Nigerian Linguist Native Speakers Categorising Four Accents  
of Nigerian English 81

SONIA CENCESCHI, CHIARA MELUZZI  
Transcription and voice comparison of noisy interceptions:  
remarks from an audio forensics report 99

### PARTE 2

#### *La linguistica forense nella pratica legale*

FRANCA ORLETTI, IACOPO BENEVIERI  
La trascrizione forense tra cattive prassi e fallacie probatorie.  
Analisi di una trascrizione di intercettazione ambientale 115

RAFFAELE MANNA, ANTONIO PASCUCCHI, JOHANNA MONTI La misurazione stilistica della Falsificazione. I comunicati delle Brigate Rosse	125
DONATELLA CURTOTTI, GABRIELLA DI PAOLO, WANDA NOCERINO La traccia vocale nelle indagini penali in Italia	139
Autori	151

CHIARA MELUZZI, SONIA CENCESCHI

## La linguistica forense e la sfida della multidisciplinarietà: introduzione al volume

La linguistica forense rappresenta oggi un filone di indagine in espansione nella ricerca linguistica e, in particolare, fonetica. Tuttavia, a causa della sua natura sperimentale e del particolare campo di applicazione, questa disciplina può (anzi deve) essere considerata quale parte di un settore applicativo multidisciplinare che necessita di un costruttivo dialogo con altre discipline e professionalità, quali le scienze giuridiche, l'ingegneria (es. analisi ed elaborazione del suono e del linguaggio), la stilistica e la psicologia, giusto per menzionarne alcune.

Si può inoltre rilevare come vi siano enormi discrepanze nei diversi paesi circa lo statuto epistemologico della linguistica forense, nonché sul riconoscimento del ruolo del linguista forense all'interno del processo giuridico. Un esempio tra tutti riguarda il coinvolgimento dell'esperto in qualità di consulente tecnico. Nel panorama italiano, importanti sviluppi sono stati raggiunti grazie al lavoro dell'Associazione Italiana Scienze della Voce e del suo gruppo di interesse, l'Osservatorio sulla Linguistica Forense. Un documento del luglio 2019 delinea le competenze del perito fonico forense e la sua natura interdisciplinare, sintetizzando inoltre il quadro critico del suo inserimento all'interno del contesto giuridico del nostro paese.

Per la figura professionale del perito fonico forense non esiste, ad oggi, un ordinamento didattico o un percorso di studi specifico, ragion per cui è necessario coinvolgere esperti di ambiti diversi, sia per la risoluzione dei quesiti peritali sia per lo sviluppo della ricerca di base, la quale è condizione imprescindibile per permettere i futuri avanzamenti in campo pratico. Un dialogo fitto e costruttivo, corredato dalla costruzione di un linguaggio comune a diverse discipline, è pertanto necessario per uno sviluppo omogeneo del settore. La stessa natura dei nostri curricula in qualità di curatrici del volume, legandosi rispettivamente al versante linguistico e a quello ingegneristico, vuole essere un segnale in tal senso.

Nel corso degli ultimi anni, non sono mancate iniziative già in questa direzione, soprattutto in seno alla Associazione Italiana Scienze della Voce. Il volume 8 della collana Studi AISV, a seguito del 17° convegno annuale dell'associazione, è appunto dedicato all'analisi dell'individualità del parlante e alle possibili implicazioni tecnologiche e forensi. Parallelamente, in ambiti applicativi strettamente legati all'audio forense, degne di nota sono le numerose attività di divulgazione e dialogo con le istituzioni promosse dall'Osservatorio Nazionale per l'Informatica Forense (ONIF) e dall'associazione culturale e scientifica Forensics Group, oltre che del già menzionato Osservatorio sulla Linguistica Forense.

Questo volume si inserisce, dunque, all'interno del panorama descritto coinvolgendo anche esperti di discipline affini e complementari alla linguistica forense con il dichiarato scopo di favorire un dialogo interdisciplinare ed iniziare la costruzione metaforica di un ponte tra le professionalità che lavorano "sul campo" e quelle più focalizzate sulla ricerca di base. Il volume è indirizzato a giovani linguisti che vogliono avvicinarsi ai temi trattati dalla linguistica forense e dalle altre discipline di settore, ma soprattutto ai professionisti e ai ricercatori che desiderino comprendere come il mondo della ricerca e quello del contesto applicativo reale siano interdipendenti e bisognosi di un trasferimento reciproco di conoscenze e competenze.

Questo volume nasce dai risultati scientifici del workshop "La linguistica forense: dalla ricerca scientifica alla pratica legale" organizzato nel corso del LIV congresso internazionale della Società di Linguistica Italiana (SLI). Alcuni lavori sono stati aggiunti, portando all'organizzazione e all'aspetto attuali del volume. Come curatrici, abbiamo scelto di mantenere una divisione in due parti, diverse (seppur complementari) dal punto di vista sia linguistico che di contenuto.

La prima parte include cinque lavori redatti in lingua inglese che presentano vari approcci alla ricerca linguistica in ambito forense: tutti sono basati su ricerche sperimentali e applicano tecnologie diverse a specifici casi di studio.

Il lavoro di Claudia Roberta Combei (*The Multimodal Expression of Denial: A Case Study on Femicide Suspects*) analizza la struttura linguistica dei discorsi dei sospettati di femminicidio. Il materiale deriva sia da interviste, sia da udienze in tribunale tratte da un corpus di inglese americano, ed è analizzato dal punto di vista linguistico e gestuale. Viene posta particolare enfasi sulla realizzazione morfo-sintattica della negazione, ma anche su ripetizioni, anafora e vaghezza semantico-legale. I risultati ben illustrano la stretta correlazione tra elementi linguistici e multimodali, rimarcando la necessità di una attenta analisi congiunta di questi due fattori in ambito forense.

Lucia Busso (*CorIELLS: A specialised bilingual corpus of English and Italian legal-lay language*) presenta invece una importante risorsa plurilingue per l'indagine linguistica focalizzata sul settore della comunicazione legale: il contributo si colloca pertanto nell'ambito della ricerca linguistica forense definita *language and the law*. L'analisi offre alcuni spunti interessanti nella comparazione tra la lingua inglese e quella italiana, indagandone la complessità dimensionale tramite la *Principal Component Analysis* (PCA). Il corpus CorIELLS è inoltre una risorsa a disposizione gratuita degli studiosi.

Sempre sul versante tecnologico, Francesco Sigona, Sonia D'Apolito, Cosimo Iaia, Barbara Gili Fivela e Mirko Grimaldi presentano uno strumento per il riconoscimento automatico del parlante anche in caso di parlato dialettale (*Forensic Automatic Speaker Recognition with dialectal speakers: a pilot study on the Taranto and Brindisi varieties*). Il contributo tratta quindi nel dettaglio il problema, teorico ma soprattutto pratico, di comparare la voce di un sospettato con un estratto audio di cui non si conosce l'identità del parlante nel caso particolare di variazione dialettale e micro-dialettale.

Il lavoro di Umar Muhammad-Gombe, Peter French ed Eleanor Chodroff (*A Comparative Analysis of Nigerian Linguistically-Naïve Native Speakers and Nigerian Linguist Native Speakers Categorising Four Accents of Nigerian English*) presenta uno studio di riconoscimento della lingua madre dei parlanti basato su una registrazione in lingua inglese, ma con presenza di accento nigeriano. Il lavoro dimostra come non vi siano sostanziali differenze percettive tra linguisti esperti e parlanti nativi, auspicando un'inclusione maggiore di questi ultimi (previa educazione linguistica di base) nei casi riguardanti i richiedenti asilo politico.

Chiude questa prima parte il contributo a firma delle curatrici del volume, Sonia Cenceschi e Chiara Meluzzi (*Transcription and voice comparison of noisy interceptions: remarks from an audio forensic report*). Partendo da un caso reale riguardante l'attribuzione dell'identità di un parlatore in due intercettazioni ambientali, le autrici presentano le diverse tecniche impiegate nel corso della perizia. Le conclusioni hanno portato a un chiaro risultato di falsa attribuzione dell'identità, offrendo la possibilità di discutere e approfondire le problematiche operative e metodologiche che affliggono il trattamento di dati reali, nonché fornire un esempio concreto di collaborazione tra linguisti e ingegneri. Il lavoro discute inoltre la necessità di dialogo tra mondo scientifico e giuridico, al fine di favorire sempre più il trattamento scientifico della prova ed incrementare, in ultimo, la sicurezza del cittadino.

Quest'ultimo contributo della prima sezione traghetta idealmente il lettore alla seconda parte del volume, la quale presenta tre lavori in lingua italiana, dedicati alla linguistica forense nella pratica legale. Il primo contributo, a firma di Franca Orletti e Iacopo Benevieri (*La trascrizione forense tra cattive prassi e fallacie probatorie. Analisi di una trascrizione di intercettazione ambientale*), mette in relazione l'aspetto della ricerca scientifica con quello del mondo forense, partendo dalle rispettive aree di studio e lavoro dei due autori. In particolare, il contributo si concentra su un problema molto attuale e sentito nell'ambito della linguistica forense, ossia la trascrizione di una intercettazione ambientale, discutendo criticamente, a partire da un reale caso di studio, sulla necessità di percorsi formativi per trascrittori in ambito giuridico e investigativo.

Il lavoro di Raffaele Manna, Antonio Pascucci e Johanna Monti (*La misurazione stilistica della falsificazione. I comunicati delle Brigate Rosse*) si concentra sull'analisi linguistica e stilistica, tramite tecniche computazionali, di comunicati delle Brigate Rosse. Il contributo pone l'accento sulle caratteristiche stilistiche e lessicali che accomunano questi scritti, differenziandoli al contempo dal cosiddetto falso comunicato apparso il 18 aprile 1978 in relazione al caso Moro.

Il contributo di Daniela Curtotti, Gabriella di Paolo e Wanda Nocerino (*La traccia vocale nelle indagini penali in Italia*) presenta una panoramica dal punto di vista giuridico concernente lo sfruttamento delle registrazioni vocali all'interno della pratica legale in Italia. Nelle loro conclusioni, gli autori ritornano su un argomento fondamentale del volume, ossia la necessità di maggiore formazione interdisciplinare dei soggetti che operano nel settore audio forense, auspicando una crescente

collaborazione tra mondo accademico, in particolare con le associazioni scientifiche di categoria, ed esperti che già lavorano attivamente sul campo.

Queste stesse posizioni erano già state in parte enunciate in apertura di volume, nella prefazione prodotta da Luciano Romito. Grazie alla sua pluriennale e pionieristica esperienza nel settore, l'autore illustra lo stato della ricerca linguistica forense in Italia e traccia le possibili direzioni di sviluppo della disciplina. Come curatrici, siamo molto grate a Luciano Romito per la interessante ed articolata prefazione al nostro volume, nonché per il sostegno dimostrato nel corso della preparazione dello stesso.

Un sentito ringraziamento va ai membri del comitato scientifico del workshop SLI, da cui ha preso poi forma questa miscellanea, nonché ai revisori dei singoli contributi raccolti in questo volume: la vostra pazienza e precisione critica hanno contribuito notevolmente alla buona riuscita di questa impresa, nonché al miglioramento delle versioni iniziali di molti lavori. In ordine rigorosamente alfabetico, ringraziamo quindi:

Silvia Calamai (Università di Siena)  
Marina Castagneto (Università degli Studi del Molise)  
Chiara Celata (Università di Urbino)  
Lorenzo De Mattei (Università di Pisa – Aptus.ai)  
Wendy Elvira García (Universidad Autónoma de Barcelona)  
José Maria Lahoz-Bengoechea (Universidad Complutense de Madrid)  
Nicholas Nese (Università di Pavia)  
Francesca Nicora (Kobe University, Japan)  
Filippo Pecorari (University of Basel)  
Elisa Pellegrino (University of Zurich)  
Paolo Roseano (Universidad Nacional de Educación a Distancia, Madrid)  
Alessandro Trivilini (SUPSI, Lugano)

Ringraziamo inoltre l'associazione Forensics Group, in particolare nelle persone degli avvocati Giuseppe Cammaroto e Domenico Famà, per il patrocinio alla stampa di questo volume.

Nel licenziare il lavoro ci auguriamo che esso possa contribuire allo sviluppo della collaborazione tra professionisti e ricercatori di ambiti diversi, ma complementari, nell'ottica di future costruttive collaborazioni scientifiche ed operative.



LUCIANO ROMITO

## La Linguistica Forense: Prefazione

La linguistica forense è una disciplina recente, che non ha ancora una precisa collocazione nel panorama scientifico europeo e italiano: la si trova, infatti, incardinata in area giuridica/criminologica, tecnologica o umanistica. Ad esempio, l'associazione europea IAFPA (International Association for Forensic Phonetics and Acoustics), che si occupa di voce in ambito forense, riflette nel suo nome l'attenzione alla fonetica e all'acustica più che alla linguistica, così come la Sociedad Española de Acústica Forense, che concentra i propri interessi soprattutto sulle discipline tecnologiche. Al contrario, in Francia l'AFPC (Association Francophone de la Communication Parlée) ha adottato un approccio orientato verso le discipline umanistiche e psicologiche, mentre in Italia, l'originario Gruppo di Fonetica Forense (oggi Osservatorio sulla Linguistica Forense), gruppo tematico all'interno dell' AISV (Associazione Italiana Scienze della Voce), presenta un orientamento molto più interdisciplinare, poiché mette insieme competenze molto differenti, che spaziano dalla linguistica alla fonetica, fino all'ingegneria e all'intelligenza artificiale.

La difficoltà di collocazione di questa disciplina risiede probabilmente nel fatto che essa non sia annoverata in nessuno dei settori scientifici ERC (in Europa) o delle Aree CUN (in Italia) e questo si riflette anche nella mancanza di una definizione unitaria e condivisa da tutti. La *Linguistics Encyclopedia*<sup>1</sup> riporta che la Linguistica Forense collabora con la giustizia rispondendo a tre domande: cosa dice il testo, cosa significa e chi lo ha scritto, confermando l'interesse iniziale per i testi scritti. Anche in Ollson (1996: 1) possiamo leggere che «letteralmente ogni testo che è in qualche modo implicato in un contesto legale o criminale è un testo forense»<sup>2</sup>. D'altronde, l'interesse per i testi scritti in ambito giuridico è già presente nei lavori di Bryant del 1930, nei quali si tratta la *funzione* delle parole nel linguaggio legale, e anche in Wetter che nel 1960 pubblica un interessante lavoro sullo stile linguistico e sui significati intrinseci delle sentenze scritte dai Giudici, e ancora nel lavoro di Danet del 1980, che verte sul linguaggio della controversia e della discussione in aula di Tribunale, e che è diventato un vero caposaldo per le arringhe. In Italia ricordiamo il mirabile lavoro di P. Bellucci (2002), nel quale si ripercorre il processo come evento linguistico essenziale dell'attività giudiziaria. Al fine di ricostruire la scena e l'accaduto, e interpretare quest'ultimo attraverso la norma giuridica, il giudice deve

<sup>1</sup> The Routledge Linguistics Encyclopedia (2010) states that «forensic linguistics help court to answer three questions about a text – what does it say, what does it mean and who wrote, typed or authored it?».

<sup>2</sup> Original text: «literally any text is somehow implicated in a legal or criminal context then it is a forensic text [...]». [Qui e altrove, viene riportato a testo una traduzione in italiano, mentre in nota si riporta l'originale inglese, n.d.E.]

avere la competenza necessaria per leggere e interpretare discorsi di ogni tipo come le intercettazioni, gli interrogatori e i dibattimenti, che provengono da registri e obiettivi diversi, per poi riuscire a sintetizzare tutto nella sentenza.

La prima volta in cui si accenna alla possibile applicazione delle competenze linguistiche in ambito forense è probabilmente nel volume di Philbrick (1949) *Language and the Law: The semantic of Forensic English*. Il termine *forensic linguistics*, invece, lo leggiamo per la prima volta nel volume *The Evans Statements: A Case for Forensic Linguistics* del linguista svedese Svartvik (1968). Il volume racconta il percorso che ha portato, ingiustamente, alla condanna a morte nel 1950 in Gran Bretagna di Timothy Evans per l'omicidio della moglie Beryl Susanna Thorley e della figlia Geraldine. L'autore effettua un'analisi sulle trascrizioni degli interrogatori di Evans. Lo stile delle trascrizioni non era quello di un alcolizzato quasi completamente analfabeta, ma quello normalmente utilizzato dalla polizia. La trascrizione era sicuramente una re-interpretazione delle affermazioni di Evans. Olsson (2012) riporta che «Jan Svartvik ha esaminato le dichiarazioni e ha concluso che queste contenevano non uno ma diversi stili di linguaggio, la maggior parte dei quali scritti in quello noto come il *registro del poliziotto*»<sup>3</sup>. Dopo tre anni dall'esecuzione capitale di Evans fu arrestato il vero colpevole. Questo caso è importante perché per la prima volta non si analizza un testo autografo, ma una trascrizione/interpretazione di un parlato verbalizzato. Questo argomento è ancora oggi molto dibattuto, soprattutto in Italia, e riguarda la scientificità della trascrizione di intercettazioni, come vedremo in seguito.

Il clamore del caso Evans non favorì l'interesse verso la Linguistica Forense come ci si aspettava. Sempre in Gran Bretagna nel 1953, Derek Bentley fu condannato a morte per l'uccisione di un agente di polizia durante una rapina. Secondo l'accusa Bentley, affetto da un disturbo psichico, fu convinto dall'amico Christopher Craig (minorenne) ad effettuare una rapina. Inoltre sempre secondo l'accusa, Craig gridò a Bentley "Let him have it" interpretata dal giudice come "utilizza/usa la pistola". Mastronardi e Trojani (2014: 20) ritengono che la frase fosse stata mal interpretata e che il significato corretto fosse "consegna l'arma all'agente" e non "usa l'arma contro l'agente". Dopo tanti studi di tipo linguistico e richieste di revisione del processo, nel 1998 Bentley viene dichiarato innocente (anche in questo caso postumo)<sup>4</sup>.

Anche la famosa indagine sull'arresto di UnaBomber<sup>5</sup>, autore di una serie di attentati commessi negli Stati Uniti dal 1978 al 1995, si basò sull'analisi testuale delle rivendicazioni insieme ad un riconoscimento dello stile linguistico da parte del fratello David Kaczynski. A questo proposito Coulthard e Johnson (2007) introducono nel loro lavoro il concetto di *idioletto* per l'identificazione dell'autore

<sup>3</sup> Original text: «Jan Svartvik examined the statements and concluded that they contained not one but several styles of language, most of which were written in what is known as 'policeman's register'».

<sup>4</sup> Molto simile è quanto accaduto in Australia nel caso Pistorius, in cui la cattiva qualità del segnale audio rendeva possibile due diverse interpretazioni: "I think I see a place" (it. Penso di vedere un posto) e "I think I see his face" (it. Penso di vedere la sua faccia / il suo volto).

<sup>5</sup> UnA è un acronimo di Università e Aeronautica o Aviazione obiettivi preferiti dall'attentatore.

del testo: «ha citato in particolare l'uso della frase 'cool-headed logician' come terminologia propria del fratello, dal nostro punto di vista una preferenza *idiolettale*, che aveva notato e ricordato» (Coulthard, Johnson, 2007: 162)<sup>6</sup>. Nel 1998 Theodore Kaczynski confessò tutti i crimini imputatigli e fu condannato all'ergastolo.

Le successive definizioni di Linguistica Forense, come quella presente in *The Cambridge Encyclopedia of the English Language* (Crystal, 2003: 462), lasciano presupporre anche altre attività oltre quelle basate sul testo scritto, ma è sul sito web ufficiale dell'Associazione Internazionale di Linguistica Forense (IAFL, n.d.) che si può leggere, forse, la definizione più articolata: «nella sua accezione più ampia, la *linguistica forense* copre tutte le aree in cui diritto e lingua si intersecano», questa interconnessione la si può suddividere in quattro «macro-campi»: *Language and Law*, *Language in the Legal Process*, *Language as Evidence* e infine *Research/Teaching*. Gibbons e Turell (2008: 1) propongono la seguente sintesi riguardo i molteplici elementi che possono essere considerati:

Le principali aree di studio includono: la lingua scritta della legge, in particolare la lingua della legislazione; il discorso giuridico, in particolare la lingua dei procedimenti giudiziari e degli interrogatori di polizia; le questioni di giustizia sociale che emergono dal linguaggio scritto e parlato della legge; l'evidenza di prove linguistiche, che possono essere suddivise in prove sull'identità/paternità e prove sulla comunicazione; l'insegnamento e l'apprendimento della lingua giuridica parlata e scritta; e traduzione e interpretariato legale<sup>7</sup>.

Dunque è possibile notare come in *Language and Law* ci si occupi di «come questo insolito registro possa essere insegnato e appreso e cosa sia possibile fare per renderlo più accessibile alle persone su cui potrebbe avere una qualche influenza, cioè a tutti» (Gibbons, Turell, 2008: 1)<sup>8</sup>; in *Language in the Legal Process* ci si concentra sulla comunicazione scritta e orale prodotta all'interno della Corte o durante gli interrogatori, spesso associata alla traduzione o all'interpretazione. L'applicazione principale ricade nei casi in cui un testimone (o un accusato) proviene da un ambiente socio-culturale meno privilegiato o si esprime in una lingua differente da quella adoperata dai soggetti coinvolti durante i vari procedimenti. In sostanza si tratta di circostanze per le quali il codice linguistico utilizzato dalla polizia, dagli avvocati e dai giudici non è totalmente intelligibile per la persona considerata, elemento che può dare luogo a contesti di coercizione, di incomprendimento o fraintendimento ai suoi danni. Si pensi ad esempio ai traduttori simultanei che spesso vengono scelti

<sup>6</sup> Original text: «he cited in particular the use of the frase 'cool-headed logician' as being his brother's terminology, or in our terms an idiolectal preference, which he had noticed and remembered».

<sup>7</sup> Original text: «Major areas of study include: the written language of the law, particularly the language of legislation; spoken legal discourse, particularly the language of court proceedings and police questioning; the social justice issues that emerge from the written and spoken language of the law; the provision of linguistic evidence, which can be divided into evidence on identity/authorship, and evidence on communication; the teaching and learning of spoken and written legal language; and legal translation and interpreting».

<sup>8</sup> Original text: «how this unusual register can be taught and learned, and what can be done to make it more accessible to the people it affects – that is, everybody».

anche durante le udienze; *Language as evidence* è il macro-campo di maggiore interesse per i nostri studi: come il linguaggio può costituire una prova nei casi giudiziari. Le funzioni maggiormente attinenti sono l'individuazione dell'autore di un testo, il rilevamento di plagio e il riconoscimento del parlatore. Non viene inserita in questa sezione la trascrizione di una registrazione o tutte quelle richieste che normalmente vengono fatte dal giudice italiano, come: l'identificazione o il riconoscimento di un suono o di un rumore; la determinazione dell'originalità o della genuinità di un nastro, di un supporto o di una registrazione in generale; il filtraggio del segnale vocale, la riduzione del rumore presente o l'enfaticizzazione di una voce e in generale l'aumento dell'intelligibilità; l'identificazione di una voce modificata; la caratterizzazione del parlante; l'identificazione mediante metodi oggettivi del parlante; l'analisi del significato in relazione a registrazioni degradate e poco intelligibili. La discordanza tra le richieste effettuate dal giudice e la definizione presente in IAFL sta proprio nella mancata collocazione della Linguistica Forense in una precisa area scientifica. Quanto appena riportato non richiede esclusivamente competenze o analisi di tipo linguistico, ma in maniera interdisciplinare competenze di teoria della comunicazione, di informatica, di statistica, di psicologia, di acustica, di fonetica sperimentale<sup>9</sup>.

Infine la quarta macro-area, *Research/Teaching*, si occupa della ricerca e della formazione, ambiti ancora una volta molto svantaggiati se non si ha un settore scientifico disciplinare di riferimento nel quale intersecare percorsi di formazione o richieste di finanziamento per progetti di ricerca nel settore.

In tutte le definizioni riportate fin qui non viene mai menzionato il passaggio dal codice parlato a quello scritto. E questo probabilmente perché la trascrizione non viene considerata un'analisi del linguaggio, ma un'operazione secondaria, perché è il segnale sonoro (o la bobina) la vera e unica prova. Questa considerazione è più che evidente nelle diverse sentenze delle Corti italiane, in cui si legge:

In tema di intercettazioni di conversazioni telefoniche o ambientali, la prova è costituita dalle bobine e dai verbali, sicché il giudice può utilizzare il contenuto delle intercettazioni indipendentemente dalla trascrizione, che costituisce la mera trasposizione grafica del loro contenuto, procedendo direttamente al loro ascolto. Ne discende che, ai fini dell'utilizzazione delle registrazioni delle conversazioni monitorate, non è necessario disporre perizia, potendo il giudice conoscere la prova mediante il diretto ascolto: rientra, quindi, nei poteri discrezionali del giudice del dibattimento di scegliere le modalità operative dell'istruttoria e di valutare se sia necessario disporre la perizia ovvero se sia sufficiente l'ascolto delle registrazioni delle comunicazioni intercettate in dibattimento oppure in camera di consiglio, potendo le parti ascoltare dette registrazioni e farne eseguire la trasposizione su nastro magnetico, così da sottoporre al giudice le proprie osservazioni. Peraltro, costituisce regola fondante del processo (cfr. artt. 511 e seguenti e 526 del c.p.p., nonché art.

<sup>9</sup> Si vedano le linee guida di AISV [https://olf.aisv.it/wp-content/uploads/2019/08/OLF\\_Le-competenze-del-perito-fonico.pdf](https://olf.aisv.it/wp-content/uploads/2019/08/OLF_Le-competenze-del-perito-fonico.pdf).

111, comma 4, della Costituzione) quella secondo cui le prove devono formarsi nel contraddittorio (Cassazione penale, sez. Vi, 28/03/2018, n. 24744).

Ciò che deduciamo dalla lettura di questa porzione di sentenza è che l'operazione di trascrizione non è un'analisi (e quindi un'attività scientifica) e che inoltre il Giudice, nella sua figura di *Peritus Peritorum* (Romito, 2016), possiede quelle «specifiche competenze tecniche, scientifiche o artistiche» citate nell'art.220 del c.p.p. che gli permettono di sostituirsi al perito o al consulente tecnico ascoltando direttamente anche segnali degradati<sup>10</sup>. Tante altre sentenze prodotte in diversi gradi di giudizio confermano l'opinione comune che per effettuare una trascrizione di un segnale degradato sia sufficiente avere un buon udito e un paio di cuffie. Si riportano in proposito alcuni esempi da sentenze:

la trascrizione deve consistere, [...] nella mera riproduzione in segni grafici corrispondenti alle parole registrate, [...] (Cassazione penale, sez. i, 24/04/1982, n. 805).

La trascrizione delle registrazioni, non soltanto non costituisce mezzo di prova, ma non può neppure identificarsi come una tipica attività di documentazione, fornita di una propria autonomia conoscitiva, rappresentando esclusivamente un'operazione di secondo grado volta a trasporre con segni grafici il contenuto delle registrazioni. Donde l'ontologica insussistenza, in relazione alle trascrizioni, di un problema di utilizzazione, potendo semmai denunciarsi la mancata corrispondenza fra il contenuto delle registrazioni e quello risultante dalle trascrizioni effettuate. D'altro canto, sarebbe del tutto ultroneo il richiamo alle norme relative alla perizia; non soltanto per il carattere di mera operazione dell'attività di trascrizione, comunque da distinguere dalla struttura gnoseologica dei mezzi di prova, dei quali può, semmai, costituire una mera rappresentazione, ma per la fungibilità, che è propria dell'attività meramente riproduttiva, non in grado di poter essere qualificata alla stregua di un documento e, conseguentemente, di un mezzo di prova (Cassazione penale, sez. Vi, 30/10/1992, in mass. Pen. cass. 1993, fasc. 6,12 s.m.).

In tema di intercettazioni telefoniche, non è inutilizzabile la trascrizione per il mancato preventivo esame dibattimentale della persona che vi ha provveduto su incarico del giudice. (La Corte ha chiarito che il richiamo, contenuto nell'art. 268, comma settimo, cod. proc. pen. a “forme, modi e garanzie” previste per la perizia, opera limitatamente alla tutela del contraddittorio e dell'intervento della difesa rispetto all'attività di trascrizione, e, inoltre, che la trascrizione delle conversazioni intercettate comporta una mera attività ricognitiva e non comprende quei compiti di valutazione) (Cassazione penale, sez. Vi, 06/11/2008, n. 2732).

In tema di intercettazioni di conversazioni telefoniche o ambientali, la nullità della perizia trascrittiva del contenuto delle conversazioni non fa derivare la inutilizzabilità delle risultanze delle stesse, atteso che la prova è costituita dalle bobine e dai verbali e il giudice può utilizzare il contenuto delle intercettazioni indipendentemente dalla trascrizione, che costituisce la mera trasposizione grafica del loro con-

<sup>10</sup> Ricordiamo infatti che quasi tutte le registrazioni intercettazioni soprattutto se ambientali sono coperte da rumori e spesso poco intelligibili.

tenuto, procedendo direttamente al loro ascolto o disponendo una nuova perizia (Cassazione penale, sez. Vi, 15/03/2016, n. 13213).

Altra questione posta è quella relativa alla perizia di trascrizione delle intercettazioni e il reiterato riferimento al mancato contraddittorio dell'esame del perito che ha redatto la trascrizione. Anzitutto, la Corte d'appello sottolinea che il perito è stato sentito in dibattimento sulle operazioni effettuate, che chiaramente non sono di carattere "valutativo", bensì "descrittive" e ciò esclude che la trascrizione possa essere assimilata a una perizia. [...] in tema di intercettazioni di conversazioni o comunicazioni telefoniche, la prova è costituita dalle "bobine", sicché è irrilevante, ai fini dell'utilizzabilità, la mancata effettuazione della trascrizione delle registrazioni (Sez. II, 19 giugno 1992, dep. 1 novembre 1992, n. 11124). Ne discende che la trascrizione delle registrazioni telefoniche si esaurisce in una serie di operazioni di carattere meramente materiale, non implicando l'acquisizione di alcun contributo tecnico scientifico e l'attività trascrittiva è attinente ad un mezzo di ricerca della prova e non rappresenta un mezzo di assunzione anticipata della prova stessa; pertanto, il rinvio dell'art. 268, comma settimo c.p.p. all'osservanza delle forme, dei modi e delle garanzie, previsti per le perizie, è solo funzionale ad assicurare che la trascrizione delle registrazioni avvenga nel modo più corretto possibile. [...] come correttamente rilevato dal giudice d'appello che, precisando gli stessi principi, ribadisce che la c.d. "perizia trascrittiva" è solo mezzo mediante il quale l'attività di intercettazione "è resa ostensibile e verificabile dalle parti" e, in tal momento, la difesa avrebbe avuto l'opportunità di far verificare da un proprio "perito" i contenuti e contestare specificatamente la materiale trascrizione delle conversazioni, tenuto conto della disponibilità delle registrazioni (Cassazione penale, sez. VI, 3/11/2015, n. 44415).

La perizia di trascrizione delle intercettazioni sono operazioni non di carattere "valutativo", bensì "descrittive" e ciò esclude che la trascrizione possa essere assimilata a una perizia e il riferimento ai brogliacci non realizza una violazione di legge. [...] Ne discende che la trascrizione delle registrazioni telefoniche si esaurisce in una serie di operazioni di carattere meramente materiale, non implicando l'acquisizione di alcun contributo tecnico scientifico (Cassazione penale, sez. VI, 22/1/2016, n. 3027).

L'incompatibilità ad assumere l'ufficio di perito per chi è stato nominato consulente tecnico in un procedimento connesso, prevista dall'art. 222, comma 1, lett. e), c.p.p., non opera con riguardo all'attività di trascrizione delle intercettazioni, disciplinata dall'art. 268, comma 7, c.p.p., atteso che il rinvio contenuto in tale norma alle forme, ai modi ed alle garanzie previste per l'espletamento delle perizie non comporta l'equiparazione del trascrittore al perito, dovendo il primo – a differenza del secondo, chiamato ad esprimere un "giudizio tecnico" – porre in essere soltanto una "operazione tecnica", non implicante alcun contributo tecnico-scientifico e connessa esclusivamente a finalità di tipo "ricognitivo" (Cassazione penale, sez. I, 26/03/2009, n. 26700).

In Italia oggi registriamo una sempre crescente richiesta di trascrizione di intercettazioni a scopo forense<sup>11</sup>. Come dimostrano le sentenze riportate, la trascrizione

<sup>11</sup> Si pensi che addirittura il sistema giuridico italiano non fornisce una precisa definizione di *trascrizione di intercettazioni* o di *trascrizione forense*. Essa può essere dedotta attraverso la lettura degli articoli del Codice di Procedura Penale dedicati all'*esecuzione della perizia*, oppure delle sentenze di Appello e

non viene considerata una analisi linguistica o una perizia ma esclusivamente un'operazione materiale e meccanica, nonostante sia nota la difficoltà nel trasferire su carta, in maniera corretta e senza una personale interpretazione, tutte le informazioni presenti nel parlato (si veda per approfondimenti Romito, 2013). Purtroppo il problema relativo alle trascrizioni o meglio alle perizie trascrittive non è di semplice soluzione; nonostante si stia facendo molto in questi anni, la normalizzazione dell'intero processo è ancora lontana. Un passo importante potrebbe essere quello di riconoscere la linguistica forense come disciplina da inserire nei percorsi formativi universitari e integrare la definizione rendendola più consona al panorama italiano. Si propone quindi la seguente definizione:

La linguistica forense è una disciplina che attiene alla Linguistica Generale, in particolare alla Fonetica Sperimentale e alle scienze forensi in genere. Ha una propria autonomia sia metodologica che procedurale e si occupa di ogni testo scritto, registrato o prodotto oralmente che sia in qualche modo coinvolto in un procedimento legale o in un contesto criminale. Pertanto, troverà applicazione nelle analisi linguistiche, dialettologiche e fonetiche; nella trascrizione, traduzione e interpretazione di registrazioni; nel miglioramento, filtraggio e digitalizzazione del segnale sonoro; nell'analisi dei rumori e nel restauro dei supporti.

Riprendendo il percorso storico riguardo l'applicazione delle indagini linguistiche in casi giudiziari, anche in Italia abbiamo avuto casi eclatanti come i sequestri di Lancia, di Getty e di Moro<sup>12</sup> o la strage di Peteano. Questi processi richiamarono l'attenzione di esperti americani (come Oscar Tosi e la tecnica del *Voiceprint*<sup>13</sup>), stimolando la discussione e la nascita di un movimento di studiosi con competenze e background scientifici molto diversi tra loro come Ingegneri delle Telecomunicazioni, Fisici Acustici, Linguisti, Dialettologi e anche alcuni Musicisti *avanguardisti*. Ogni perizia, in quegli anni, fu una vera e propria ricerca scientifica con apporti provenienti da tutte le discipline accademiche. Si svilupparono tecniche e studi di Natural Language Processing (NLP) all'interno delle facoltà umanistiche con apporti scientifici di linguisti, dialettologi e psicologi, mentre contemporaneamente,

---

di Cassazione prodotte dai Giudici. L'art. 268 comma 7 del Codice di Procedura Penale – *Esecuzione delle operazioni*, riporta che «Il giudice dispone la trascrizione integrale delle registrazioni ovvero la stampa in forma intelligibile delle informazioni contenute nei flussi di comunicazioni informatiche o telematiche da acquisire, osservando le forme, i modi e le garanzie previsti per l'espletamento delle perizie [art 2201]. Le trascrizioni o le stampe sono inserite nel fascicolo per il dibattimento».

<sup>12</sup> In Italia uno dei primi celebri casi in cui vennero coinvolti specialisti della lingua fu il Sequestro Moro, avvenuto nel 1978 per mano del gruppo terroristico delle Brigate Rosse. Lo scrittore Leonardo Sciascia e il linguista Tullio de Mauro cercarono di estrarre dai comunicati dei terroristi e dalle lettere dello stesso Moro dettagli che potessero agevolare il ritrovamento del ministro o, in un certo qual modo, prevedere ciò che avrebbero fatto i suoi aguzzini [ma si veda anche Manna, Pascucci & Monti, questo volume, n.d.c.]. Infatti, sottolinea Marchetti (2017: 54), «non emergono dei tentativi di collegare i tratti linguistici dei comunicati con i loro autori, e non viene quindi fatta un'analisi sociolinguistica degli stessi comunicati».

<sup>13</sup> Si veda Hollien (1981). Per un approfondimento sulle applicazioni del VoicePrint si vedano Kersta (1962), Tosi et al. (1972), Pollack, Pickett & Sumbly (1954).

nelle facoltà di Ingegneria e di Fisica, si iniziarono le prime ricerche di *speech processing*. La necessità di utilizzare modelli statistici per la valutazione dei dati prodotti da entrambi i filoni di ricerca ha incentivato un humus scientifico molto produttivo e coinvolgente. Le indagini linguistiche si orientarono soprattutto sulla identificazione del parlante o meglio il riconoscimento della paternità di una voce. Questa tecnica nasce negli Stati Uniti quando si inizia a parlare di DNA linguistico, di idioletto e in correlazione con la tecnica del Fingerprint, Kersta (1962) scrive che è possibile rilevare il Voiceprint di una persona:

L'identificazione dell'impronta vocale è un metodo mediante il quale le persone possono essere identificate da un esame spettrografico della loro voce. Strettamente analogo all'identificazione delle impronte digitali, che utilizza le caratteristiche uniche che si trovano nelle impronte digitali delle persone, l'identificazione delle impronte vocali utilizza le caratteristiche uniche che si trovano nelle loro espressioni<sup>14</sup>.

La stessa idea viene riportata anche in Nash, citato da Hollien (1996): «Poiché ciascuna delle creste delle dita o sul palmo della mano differisce l'una dall'altra, così tutte le altre parti del corpo. Sono unici per te ... compreso il tuo meccanismo vocale»<sup>15</sup>. L'interesse e l'attenzione fu tale che si svilupparono addirittura due diversi protocolli, il primo fu realizzato da VIAAS (Voice Identification and Acoustic Analysis Subcommittee) della *International Association for Identification* e il secondo proprio dalla FBI. In Hollien (1996)<sup>16</sup> si può leggere che almeno fino al 1990 le Corti degli Stati Uniti d'America accettarono, almeno il 50% delle volte, l'utilizzo di tali protocolli nonostante Nolan (1983) già evidenziava i propri dubbi sull'utilizzo dei *voiceprints* per l'identificazione del parlante. Di fatto si dovette aspettare molto per ottenere la sentenza dello Stato della California che recita:

L'analisi spettrografica uditiva della voce umana ai fini dell'identificazione forense non ha trovato accettabilità e affidabilità nella comunità scientifica di riferimento e che, pertanto, non sussistono basi per la sua ammissibilità alle prove in questa udienza ai sensi della legge della California (*State of Texas v David Shawn Pope*, 1986, 204th District Court, Case No F85-98755-MQ)<sup>17</sup>.

Quindi fino ad allora, nelle aule dei Tribunali Americani, la capacità espositiva e di convincimento di singoli periti veniva preferita all'evidenza scientifica. Oggi, grazie anche alla sentenza della Suprema corte USA nel caso *Daubert versus Merrel*

<sup>14</sup> Original Text: «Voiceprint identification is a method by which people can be identified from a spectrographic examination of their voice. Closely analogous to fingerprint identification, which uses the unique features found in people's fingerprints, voiceprint identification uses the unique features found in their utterances».

<sup>15</sup> Original text: «As each one of the ridges of your fingers or on the palm of your hand differ from each other, so do all of the other parts of your body. They are unique to you ... including your voice mechanism».

<sup>16</sup> Si veda inoltre, Hollien (2002), Hollien et al. (1982, 1995, 2000).

<sup>17</sup> Original Text: «That the aural spectrographic analysis of the human voice for the purposes of forensic identification has failed to find acceptability and reliability in the relevant scientific community, and that therefore, there exists no foundation for its admissibility into evidence in this hearing pursuant to the law of California».



Dow Pharmaceuticals Inc 1993, ogni nuova teoria proposta in un processo deve soddisfare i criteri di scientificità che sono:

- qualunque teoria o tecnica utilizzata deve essere testata, pubblicata o sottomessa ad un peer review;
- deve prevedere e dichiarare il potenziale errore;
- deve essere accettata in larga misura dalla comunità scientifica<sup>18</sup>.

In Italia la situazione dei metodi utilizzati<sup>19</sup> e delle competenze coinvolte<sup>20</sup> non è regolamentata e la sentenza Daubert non viene mai applicata<sup>21</sup>. È stata condotta una

<sup>18</sup> Original Text: «In order for scientific evidence to be admissible, the judge must consider whether...

- the theory or technique is empirically testable and replicable.
- the theory or technique has been subjected to peer review and publication.
- the potential error rate is reported.
- the theory and technique are generally accepted by the relevant scientific community».

<sup>19</sup> In Italia ben l'11% delle perizie di identificazione vengono effettuate con la tecnica del Voiceprint e questo nonostante la risoluzione dell'IAFPA del 2007 dica: «Voiceprint Resolution presents the organisation's position in relation to voiceprints: IAFPA dissociates itself from the approach to forensic speech comparison known as the "voiceprint" or "voicegram" method in the sense described in Tosi (1979). This approach to forensic speaker identification involves the holistic, i.e., non-analytic, comparison of speech spectrograms in the absence of interpretation based on understanding of how spectrographic patterns relate to acoustic reflexes of articulatory events and vocal tract configurations. The Association considers this approach to be without scientific foundation, and it should not be used in forensic casework». Si vedano Romito & Galatà (2007, 2008) e Romito (2021).

<sup>20</sup> In Italia non esiste un albo dei periti fonici pertanto spesso si viene nominati periti non per titoli posseduti o per la competenza ma solo per conoscenza diretta, una sorta di rapporto fiduciario con il giudice o il PM. Eppure il controllo sulla competenza del perito potrebbe essere svolto in aula dalle parti (PM e Avvocati). In data 29 marzo 2018, il Dipartimento per gli affari di Giustizia del Ministero della Giustizia ha inviato ai responsabili della prevenzione della corruzione e della trasparenza una nota che potrebbe essere utilizzata in fase di incarico: «Gli incarichi di collaborazione e di consulenza conferiti a soggetti esterni alla compagine della Pubblica Amministrazione sono sottoposti a pubblicità obbligatoria per esigenze di trasparenza. Il contenuto dell'obbligo si estende, in particolare, agli estremi dell'atto di conferimento dell'incarico, ai compensi e al curriculum vitae [...]. Come noto il curriculum è un documento che descrive la carriera e il profilo scientifico ed accademico di un soggetto e va in genere allegato alle domande di concorso e di assunzione».

<sup>21</sup> Normalmente la scientificità dei metodi utilizzati non viene definita dalla comunità scientifica ma dal Giudice stesso nella sua veste di Peritus Peritorum. Mi piace però riportare qui la lungimiranza della Corte Suprema di Cassazione V Sez. Penale nr 36080/15 del 27.3.2015 imp. Sollecito R. e Knox A.: «Il dibattito culturale, pur nel rispetto del principio del libero convincimento del giudice, si propone anche di rivisitare criticamente la nozione, oramai obsoleta e di assai dubbia credibilità, del giudice 'peritus peritorum'. In effetti [...] esprime un modello culturale non più attuale e, anzi, decisamente anacronistico, quanto meno nella misura in cui pretenda di assegnare al giudice reale capacità di governare il flusso di conoscenze scientifiche che le parti riversino nel processo, ove invece una più realistica impostazione lo vuole del tutto ignaro di quei contributi, che sono il frutto di un sapere scientifico che non gli appartiene e non può – ne deve – appartenergli. [...]. La conseguenza dell'ineludibile presa d'atto di tale stato di *legittima* ignoranza del giudice, e dunque della sua incapacità di governare «autonomamente la prova scientifica, non può, però essere l'acritico affidamento, che equivarrebbe – anche per un malinteso senso del libero convincimento e di altrettanto malinteso concetto di «perito dei periti» – a sostanziale rinuncia al proprio ruolo mediante fideistica accettazione del contributo peritale, cui delegare la soluzione del giudizio e, dunque, la responsabilità della decisione. [...]. D'altro canto, è a tutti noto che non esiste una sola scienza, portatrice di verità assolute ed immutabili nel tempo, ma

valutazione dei metodi per il riconoscimento del parlatore all'interno della campagna Evalita nel 2009 (cfr. Romito, Galatà, 2009), con la creazione di un corpus di voci intercettate (ancora oggi a disposizione della comunità italiana: *Primula corpus di voci intercettate*), ma la partecipazione da parte dei periti è stata bassissima.

I metodi maggiormente utilizzati sono: il metodo parametrico semiautomatico, automatico, misto o linguistico basato su prove d'ascolto, analisi dialettologiche e comparazione di dati formantici.

Riguardo il metodo parametrico semiautomatico i software statistici utilizzati sono IDEM normalmente utilizzato dal RACIS dell'Arma dei Carabinieri (cfr. Paoloni, 1979, 1984 e 1993; Falcone, De Sario, 1994; Paoloni, Federico, 1998) che utilizza una statistica decisionale bayesiana con l'ausilio di una comunità italiana di riferimento; SMART (Statistical Method Applied to the Recognition of the Talker), sviluppato all'interno di un progetto Europeo con capofila il Ministero degli interni (OISIN II (Rif. JAI/2002/OIS/035) e attualmente utilizzato dalla Polizia Scientifica e IMPAVIDO<sup>22</sup> che integra le statistiche dei metodi precedenti.

Tutti i test utilizzano misure formantiche e calcolano la distanza di mahalanobis, fornendo tabelle con risposta a due uscite (YES – NO), il valore (sottoforma di numero scientifico) del rapporto di verosimiglianza (o Likelihood o LLR) utile per 'pesare' l'errore di falsa attribuzione o di mancato riconoscimento (FFI e FFR) a-priori e a-posteriori.

La differenza tra i metodi è rilevabile nella statistica utilizzata, ma soprattutto nella possibilità di variare la comunità di riferimento in base alla provenienza delle voci da comparare e quindi sull'analisi linguistica preventivamente effettuata sui campioni di voce da analizzare<sup>23</sup>.

I metodi automatici oggi utilizzati sono fondamentalmente di due tipi: quelli basati su i-vector e analisi MFCC e quelli basati su *Deep Neural Networks* o reti neurali e sono tutti commercializzati. I più conosciuti sono: i modelli commercializzati dalla ditta Nuance, per la precisione il modello Nuance Forensics 9.2 che utilizza un sistema PLDA i-vector e il modello Nuance Forensics 11.1 che combina la tecnologia PLDA i-vector con alcune funzionalità di Deep Neural Networks (Jessen et al., 2019a); i modelli commercializzati da Phonexia, per la

---

tante scienze o pseudoscienze tra quelle *ufficiali* e quelle non validate dalla comunità scientifica, in quanto espressione di metodiche di ricerca universalmente riconosciute».

<sup>22</sup> Il software IMPAVIDO (Integrated Methods for PArametric Voice IDentificatiOn), è un nuovo tool per la comparazione forense della voce. Ideale per l'integrazione in un metodo acustico-foneticamente semiautomatico, implementa una comparazione statistica di *feature* spettrale (frequenza fondamentale e frequenze formanti delle vocali della lingua italiana), con due differenti modelli integrati fra loro: modelli parametrici e modelli non parametrici. Il modello parametrico implementato in IMPAVIDO (denominato eIDEM) è ispirato al modulo SPREAD del software IDEM (IDentification Method), tradizionalmente in uso all'Arma dei Carabinieri. E l'implementazione del modello non parametrico (eSMART) si ispira al software SMART, tradizionalmente in uso alla Polizia di Stato. Il software utilizza una metodologia basata sul rapporto di verosimiglianza (Bayesian Interpretation Framework), in linea con le raccomandazioni ENFSI.

<sup>23</sup> Per dettagli si veda la seguente bibliografia: Brutti et al., 2002; Bove et al., 2002, 2003 e 2004; Bove, 2006; Romito et al., 2009.

precisione il modello denominato SID (Speaker Identification)-XL3 che utilizza un sistema PLDA i-vector con due flussi di funzionalità: MFCC in senso classico e funzionalità DNN-Stacked Bottle-Neck e il modello denominato SID-BETA4 che utilizza MFCC in input e un'architettura di incorporamento dei parlanti basata su DNN (Jessen et al., 2019b); e il modello VOCALISE (Voice Comparison and Analysis of the Likelihood of Speech Evidence) basato su Deep Neural Network (DNN, Kelly et al., 2019).

Il terzo metodo misto richiede una grande competenza linguistica dialettologica e sociolinguistica e può essere identificato come la fase preliminare ad un esame parametrico semiautomatico. Non bisogna infatti dimenticare la complessa realtà linguistica della penisola, caratterizzata da una vastissima moltitudine di dialetti, che spesso influenzano la lingua nazionale. Questa peculiarità può spesso rivelarsi un grande vantaggio nelle perizie di riconoscimento del parlatore, giacché esistono studi sulle più minuziose differenze linguistiche fra comunità sociali, soprattutto nella fase iniziale delle indagini. Agli inizi di un'investigazione, quando non ci sono ancora sospettati, è interessante valutare le informazioni che l'esperto può dedurre da una registrazione al fine di creare il profilo del soggetto presente nella prova. È possibile identificare tre categorie di informazioni rilevabili: grammaticali, sociali e biologiche o fisiologiche. I confini fra queste categorie ovviamente non sono rigidi e alcuni elementi possono trovarsi a cavallo tra più aree. Rientrano nelle informazioni grammaticali la lingua nativa (L1 che può anche essere un dialetto), l'accento straniero o la conoscenza di una seconda lingua (L2 che può anche essere italiano) e l'etnoletto. Tra le informazioni sociali, ma anche linguistiche, avremo il socioletto e il regioletto. Invece, fra il sociale e il biologico si inseriscono il sesso o il genere e l'età. Specificatamente fisiologici sono le patologie del linguaggio e persino la dimensione del parlante (altezza, peso). Di fatto, la profilazione dedotta da produzioni vocali è teorica e non pretende di individuare un soggetto specifico (missione ardua anche con le tecnologie più avanzate), ma categorizzare un gruppo umano che abbia determinate peculiarità linguistiche, con l'obiettivo di fornire un'ipotesi di percorso per le indagini.

Concludendo questa carrellata storica sulla Linguistica Forense, in Italia è possibile distinguere due fasi nella storia dell'applicazione di metodiche linguistico-acustiche a scopi giudiziari. La prima, inquadrata fra il 1975 e il 1980, presenta metodologie discutibili, anche se interessanti per il periodo considerato, caratterizzate da un approccio multidisciplinare; la seconda, che va dal 1980 fino ad oggi, mostra invece una crescita esponenziale dei casi in cui il telefono diventa vero protagonista di crimini di vario genere e ciò ha portato il dilagare della richiesta di perizie, con la conseguente nascita di pseudo-professionalità per venire incontro alla crescente richiesta e un allontanamento da parte dei ricercatori e degli studiosi di Linguistica in genere.

Oggi i Tribunali italiani, come 50 anni fa, non chiedono indagini linguistiche, analisi dialettologiche, profilazioni o interpretazioni, ma quasi esclusivamente trascrizioni di intercettazioni. Tali trascrizioni non vengono considerate delle vere

perizie, ma esclusivamente delle operazioni tecniche e meccaniche senza alcun valore scientifico. E questo pur sapendo che non è facile mettere su carta la complessità di una conversazione<sup>24</sup>.

Abbiamo iniziato questo contributo scrivendo che la Linguistica Forense è una disciplina molto giovane, ciononostante i ricercatori sono pienamente consapevoli dei limiti, dei problemi e soprattutto della strada da intraprendere per una veloce soluzione, al contrario di quanto avviene nelle aule dei tribunali e anche nelle commissioni Giustizia del Governo.

Non è possibile inserire la varietà delle richieste normalmente formulate dal giudice sotto la stessa e unica denominazione: *perizia fonica*. Alcune di queste richieste prevedono, infatti, accertamenti tecnici e competenze specifiche molto specializzate come le identificazioni del parlante, il miglioramento del segnale o l'identificazione dei rumori e altre come le trascrizioni forensi prevedono competenze molto differenti. Anche il concetto di *scientificità* non può essere applicato in maniera univoca. Alla domanda se le trascrizioni sono un risultato scientifico o un'interpretazione soggettiva, la prima cosa che va sottolineata è che le trascrizioni sono innanzitutto qualcosa di poco chiaro e di molto complesso. Potremmo sostituire la definizione *oggettiva* con *attendibile*, ma questo vorrebbe dire rendere le trascrizioni poco chiare, perché la traduzione di un segnale sonoro in una trascrizione grafica deve rendere conto delle caratteristiche del sonoro che non sono assolutamente quelle proprie dello scritto. Quando il Giudice richiede *la mera traduzione in parole del segnale orale*, cade in contraddizione perché nel parlato non esistono le parole. Nel parlato abbiamo un flusso continuo dove i suoni si fondono e possono essere frammenti di parola, borbottii o frasi intere, misti a risatine, sospiri e pause. In fonologia una parola è ciò che sta tra due pause di respiro e questa non coincide mai con la parola ortografica che sta tra due spazi bianchi. Quindi una trascrizione *attendibile* risulterebbe poco chiara e l'identificazione delle parole è essa stessa una prima interpretazione; è una segmentazione di un contenuto che non nasce come parole, quindi un'interpretazione soggettiva: la traduzione in parole del parlato è un primo livello di interpretazione. Inoltre, spesso la trascrizione forense è redatta in lingua italiana, mentre invece il sonoro è in lingua dialettale, quindi il perito

<sup>24</sup> «Il parlato comprende un complesso di codici paralleli e concorrenti. Vi è, infatti, la possibilità di utilizzare codici paralinguistici come il volume della voce, il tono, l'intonazione, il ritmo, il silenzio; il codice cinesico o cinestesico con i movimenti del corpo, le espressioni del viso, degli occhi, delle mani; il codice prossemico con la gestione dello spazio e quindi la posizione del corpo e la distanza tra gli interlocutori; il codice aptico attraverso il contatto fisico come la stretta di mano, il bacio sulle guance come saluto ad amici e parenti, un abbraccio, una pacca sulla spalla ecc. in un lavoro di Mehrabian (1972) viene dimostrato che la percezione di un messaggio vocale può essere suddivisa percentualmente in un 55% di movimenti del corpo – soprattutto espressioni facciali – 38% di aspetto vocale come volume, tono, ritmo ecc. e infine solo per il 7% di aspetto verbale, cioè le parole. La percezione e la corretta interpretazione di un messaggio dipende in minima parte dal significato letterale di ciò che viene detto ed è molto influenzato da tutti i codici relativi alla comunicazione non verbale. Questi codici sono tutti di natura sociale e culturale e indicano il tipo di relazione che intercorre fra gli interlocutori» (Romito, 2013: 181).

trascrittore effettua una doppia interpretazione prima delle informazioni mancanti ed in seguito della possibile/probabile traduzione dal dialetto all'italiano; quando invece la trascrizione viene riportata in dialetto si pone il problema della mancata corrispondenza tra i suoni dialettali e i caratteri della lingua nazionale e inoltre è necessario assegnare correttamente i diversi turni di parola ai singoli parlanti, ecc.

Quando si parla di attendibilità, sarebbe opportuno avere indicazioni di quali siano gli aspetti individuati come prioritari<sup>25</sup>, come privilegiare il contenuto e non i fenomeni di esitazione oppure le pause. Ad esempio, sappiamo che esistono *pause lunghe* e *pause brevi*, ma queste non hanno un riferimento paradigmatico o fonologico, cioè non esiste una precisa definizione in durata, di pausa breve. La durata delle pause è strettamente correlata allo stile locutorio del singolo parlante, si pensi ad esempio a Celentano o a Craxi. Le pause, nella trascrizione, devono essere riportate alla storia conversazionale del parlante; è una personale interpretazione definire una pausa "lunga" se il parlante tende normalmente a fare pause. Inoltre, come detto, una produzione ironica da parte di un parlante può non passare attraverso le cuffie, perché magari nell'interazione i personaggi che parlano si danno una strizzatina degli occhi oppure fanno qualche altro gesto 'non acustico' quindi non registrabile. In questo caso l'attendibilità deve essere misurata con il segnale sonoro o con la vera intenzione dei parlanti? Bisogna fare attenzione anche solo alla definizione di *scientificità* o *oggettività* per una trascrizione. Dobbiamo avere la consapevolezza che la trascrizione dell'intercettazione è uno strumento fondamentale, ma ha anche dei limiti. In genere viene usata per guardare a dei contenuti, per cercare dei contenuti, e cercare dei contenuti a volte può essere un atto interpretativo. Spesso la misura cautelare, ad esempio, viene disposta soltanto sulla base dei brogliacci<sup>26</sup> (cioè dei contenuti) e sappiamo bene che oggi non valgono dal punto di vista probatorio e questo, a parer mio, è un controsenso.

Quindi la prima esigenza da affrontare è quella di codificare o comunque chiarire meglio i termini e i limiti della *perizia fonica*; il secondo problema, al fine di avere un contraddittorio valido, e anche per affidare all'esperto la ricerca dei contenuti di cui sopra liberandolo dal limite delle parole grafiche, è quello di avere delle professionalità di particolare livello in modo tale da offrire alla difesa, ma anche agli uffici di Procura delle persone altamente qualificate come consulenti tecnici.

<sup>25</sup> «Nel caso della *trascrizione forense integrale*, l'esperto-trascrittore dovrebbe conoscere l'obiettivo dell'intercettazione in modo da concentrare la propria attenzione su tutte quelle informazioni, presenti nei diversi canali della comunicazione, utili alla corretta comprensione del messaggio. Non avere alcuna conoscenza dell'obiettivo e non essere autorizzato a consultare il fascicolo non è, a parere dell'autore, indice di imparzialità, obiettività e mancato condizionamento» (Romito, 2021).

<sup>26</sup> I brogliacci sono dei verbali sommari del contenuto di una intercettazione. Anche in questo caso è necessario avere delle persone competenti all'interno della sala di intercettazione per non incombere in quello che Goodwin (1994) definisce *visione professionale*, cioè lasciare che il proprio lavoro influenzi l'interpretazione di un dialogo.

### *Riferimenti bibliografici*

- BELLUCCI, P. (2002). *A onor del vero, Fondamenti di linguistica giudiziaria*. Torino: UTET.
- BRYANT, M.M. (1930). *English in the law courts*. Columbia: Columbia University Press.
- BRUTTI, P., FABI, F. & JONA LASINIO, G. (2002). Una proposta di meta-analisi basata sulla combinazione di classificatori per il problema del riconoscimento del parlatore. In *Statistica*, 3, 455-473.
- BOVE, T., BRUTTI, P., FABI, F., JONA LASINIO G., GIUA, P.E., FORTE, A. & ROSSI C., (2002). Tecnologie Informatiche nella Promozione della Lingua Italiana. In T.I.P.I., 121-124.
- BOVE, T., BRUTTI, P., FABI, F., JONA LASINIO G., GIUA, P.E., FORTE, A. & ROSSI C. (2003). Three approaches to the speaker identification problem for forensic use. In *Atti del Convegno Cladag 22-24 settembre 2003*, invited relation.
- BOVE, T., JONA LASINIO, G. & ROSSI, C. (2004). The speaker recognition problem. In *XLII Riunione Scientifica della Società Italiana di Statistica*, 429-440.
- BOVE, T. (2006). SMART II: Analisi della voce, In *Trattamento Automatico Linguaggio*, conference paper.
- COULTHARD, M. & JOHNSON, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence (1<sup>st</sup> edition)*. London: Routledge.
- CRYSTAL, D. (2003). Final frontiers in applied linguistics?. In *British Studies in Applied Linguistics*, 18, 9-24.
- DANET, B. (1980). Language in the courtroom. In GILES, H., ROBINSON, P. & SMITH, P.M. (eds.) *Language. Social Psychological Perspectives*, Oxford: Pergamon, pp. 367-376.
- FALCONE M., DE SARIO N. (1994). A PC speaker identification system for forensic use: IDEM. In: *Proceedings of the ISCA Workshop on Automatic Speaker Recognition, Identification and Verification*. Martigny, Switzerland, pp. 169-172.
- FEDERICO, A. (1998). The parametric approach in forensic speaker recognition. In *Proceedings of the Cost250 Workshop*, 45-51.
- FEDERICO, A.G., IBBA, G. & PAOLONI, A. (1984). Criteri per la caratterizzazione di una voce ai fini del riconoscimento del parlante. In *Rivista Italiana di Acustica*, 8.
- FEDERICO, A. & PAOLONI A. (1993). Bayesian decision in the speaker recognition by acoustic parameterization of voice samples over telephone lines. In *Proceedings of EURO-SPEECH*, 93, 2307-2310.
- GIBBONS, J. & TURELL, M.T. (eds.) (2008). *Dimensions of forensic linguistics*. Berlin: John Benjamins.
- GOODWIN, C. (1994). Professional vision. In *American Anthropologist*, 96(3), 606-633.
- HOLLIEN H. (1996). Consideration of guidelines for earwitness lineups. In *Forensic Linguistics*, 3(1), 14-23.
- HOLLIEN H. (2002). *Forensic Voice Identification*. London: Academic Press.
- HOLLIEN, H., HUNTLEY, R., KUNZEL, H. & HOLLIEN, P.A. (1995). Criteria for earwitness lineups. In *International Journal of Speech, Language and the Law*, 2(2), 143-153.
- HOLLIEN, H., MAJEWSKI, W. & DOHERTY, E.T. (2000). Perceptual identification of voices under normal, stress and disguise speaking conditions. In *Journal of Phonetics*, 10, 139-148.

- HOLLIEN H. & SCHWARTZ, R. (2000). Aural-perceptual speaker identification: problems with non-contemporary samples. In *Forensic Linguistics*, 7(2), 199-211.
- JESSEN, M., MEIR, G., & SOLEWICZ, Y.A. (2019a). Evaluation of Nuance Forensics 9.2 and 11.1 under conditions reflecting those of a real forensic voice comparison case (forensic\_eval\_01). In *Speech Communication*, 110, 101-107.
- JESSEN, M., BORTLÍK, J., SCHWARZ, P. & SOLEWICZ, Y.A. (2019b). Evaluation of Phonexia automatic speaker recognition software under conditions reflecting those of a real forensic voice comparison case (forensic\_eval\_01). In *Speech Communication*, 111, 22-28.
- KELLY, F., FORTH, O., KENT, S., GERLACH, L. & ALEXANDER, A. (2019). Deep neural network based forensic automatic speaker recognition in VOCALISE using x-vectors. In *AES Conference on Audio Forensics*, 1-7.
- KERSTA, L.G. (1962a). Voiceprint-Identification Infallibility. In *The Journal of the Acoustical Society of America*, 34, 1978-1978.
- KERSTA, L.G. (1962b). Voiceprint identification. In *Nature*, 196, 1253-1257.
- MARCHETTI, E. (2017). "Boia imperialisti, spie di regime e corrotti buffoni". La lingua dei comunicati delle brigate rosse durante il Sequestro Moro. In *L'analisi linguistica e letteraria*, XXV, 51-70.
- MASTRONARDI, V., & TROJANI, A. (2014). *Appunti di Linguistica Forense. Introduzione e Reality Monitoring*. Roma: Peritare-Lulu.
- NOLAN, F. (1991). Forensic phonetics. In *Journal of Linguistics*, 27, 483-493.
- OLLSON, J. (1996). The dictation and alteration of text. In *The International Journal of Speech, Language and the Law*, 4(2), 226-251.
- OLSSON, J. (2012). Some Aspects of the Use of Corpora in Forensic Linguistics. In HYLAND, K., HUAT CHAU, M., HANDFORD, M. (Eds.) *Corpus Applications in Applied Linguistics*. London: Bloomsbury.
- PAOLONI, A. (1979). Descrizione delle principali metodologie utilizzate per il riconoscimento del parlatore. In FERRERO, F.E. (a cura di), *L'identificazione della persona per mezzo della voce*. Roma: ESA.
- PAOLONI, A., FALCONE, M. & FEDERICO, A. (1998). The parametric approach in forensic speaker recognition. In *Proceedings of COST*, 250, 45-51.
- PHILBRICK, F.A. (1949). *Language and the Law: The semantic of Forensic English*. London: MacMillan.
- POLLACK, I., PICKETT, J.M. & SUMBY, W.H. (1954). On the Identification of Speakers by Voice. In *The Journal of the Acoustical Society of America*, 26, 403-410.
- ROMITO, L. (2013). *Manuale di Linguistica Forense*. Roma: Bulzoni.
- ROMITO, L. (2016). La competenza linguistica nelle perizie di trascrizione e di identificazione del parlatore (a margine, alcune riflessioni sul difficile rapporto tra scienza e processo). In *Diritto Penale Contemporaneo*, 6, 1-14.
- ROMITO, L. (2021). La trascrizione in ambito Forense. Un vademecum per il trattamento digitale dei dati linguistici. In BERNINI, G., VALENTINI, A., SATURNO, J. & SPREAFICO, L. (a cura di) *Superare l'evanescenza del parlato*. Bergamo: Bergamo University Press, 201-230.
- ROMITO, L. E GALATÀ, V. (2007). Speaker Recognition: Stato dell'arte in Italia, In *III Convegno Nazionale AISV*. Trento: EDK, 223-242.

ROMITO, L. E GALATÀ, V. (2008). Speaker Recognition in Italy: Evaluation of Methods used in Forensic cases. In *Languages Design*, Special issue I, 229-240.

ROMITO, L. E GALATÀ, V. (2009). Forensic Speaker Identity Verification (F-SIV) in Italy, First Evaluation Campaign. In *Evalita-2009*.

ROMITO, L., BOVE, T., DELFINO, S., ROSSI, C. & JONA LASINIO, G. (2009). Specifiche Linguistiche del Database utilizzato per lo Speaker Recognition in S.M.A.R.T. In *Atti del convegno AISV*. Cosenza: EDK Editore.

SVARTVIK, J. (1968). *The Evans Statements: A Case for Forensic Linguistics*. Goteburg: University of Goteburg Press.

TOSI, O., OYER, H., LASHBROOK, W., PEDREY, C., NICOL, J. & NASH, E. (1972). Experiment on Voice Identification. In *The Journal of the Acoustical Society of America*, 51, 2030.

WETTER, J.G. (1960). *The Styles of Appellate Judicial Opinions*. Leyden: Sythoff.





PARTE I

APPROCCI ALLA RICERCA  
IN LINGUISTICA FORENSE

APPROACHES TO RESEARCH  
IN FORENSIC LINGUISTICS





CLAUDIA ROBERTA COMBEI

## The Multimodal Expression of Denial: A Case Study on Femicide Suspects

This pilot study investigates multimodal denial by analysing the discourses of male suspects of femicide that are the victims' (ex-)intimate partners and friends. We deploy an exploratory approach to map recurrent verbal and nonverbal features of denial in an *ad hoc* corpus of North-American English (over 10 hours and 101,000 tokens) of audio-video material, featuring guilty suspects of femicide in two legal contexts (i.e., police interviews and courtroom hearings). Our findings suggest that denial is constructed linguistically by means of several strategies that complement and reinforce each other. Besides an extensive use of the morpho-syntactic realization of negation, suspects deny the accusations through recurrent multimodal features, such as, headshakes, shoulder shrugs, gaze direction (towards law enforcement officers), and open hands (either still or moving onward). We also observe that in our corpus denial co-occurs with repetitions, anaphora, vagueness, and reduced sentence length. These results call for additional systematic research on multimodality in legal interactions and in other contexts.

*Keywords:* negation, denial, multimodality, co-speech gestures, gendered violence

### 1. Introduction

In 1976, Diana Russell mentioned the term 'femicide' for the first time during her testimony at the International Tribunal on Crimes against Women. But it was only in 1992 that Jane Caputi and Diana Russell defined 'femicide' as «the most extreme form of sexist terrorism motivated by hatred, contempt, pleasure or a sense of ownership of women» (Caputi, Russel, 1992: 12). More than two decades later, on the International Day for the Elimination of Violence against Women and through the Vienna Declaration on Femicide, the term 'femicide' was broadly defined as «the killing of women and girls because of their gender» (UN, 2013: 2). In more recent times, the European Institute for Gender Equality (EIGE) defined 'femicide' as «the killing of a woman by an intimate partner and the death of a woman as a result of a practice that is harmful to women» (EIGE, 2017: 28). In this paper, we will adopt the broad definition of 'femicide' formulated by UN (2013).

Femicide differs from male homicide, as most crimes are committed by intimate partners (including ex-partners), other familiar male figures, and they involve ongoing abuse at home, threats or intimidation, psychological violence, sexual and physical violence, or other situations where women have less power or fewer resources than men. In fact, a recent report published by the United Nations Office on Drugs and Crime (UNODC) reveals that in 2020 there were 81,000 women murders around

the world. In 58% of the femicide cases, the offender was the victim's (ex-)intimate partner or some other male family member, friend, colleague, or acquaintance. This means that every eleven minutes, a woman or a girl is killed by someone she knows well. The UNODC (2020) study also claims that although various schemes aimed at reducing or eliminating femicide have been developed worldwide (e.g., VAWG strategy in UK, DdL 99/2013 and L. 119/2013 in Italy, etc.), a significant progress in protecting the lives of female victims is yet to be made. This can be inferred also by the high number of gender-related crimes reported in the past decade that has remained unchanged in large measure. In addition to that, in the American continent there has been a 9% increase of femicide cases with respect to the average global numbers. The afore-mentioned report speculates that collecting accurate data on femicide is difficult because, oftentimes, law enforcement, social service bodies, and medical institutions do not have the necessary information regarding the causes of the homicide, let alone gender-related motives or information regarding the victim–perpetrator relationship.

While the afore-mentioned findings urge for better cooperation between legislators, law enforcement, social services, non-government organization, and scholar research in order to provide more efficient responses to violence against women, in recent years the pervasiveness of femicide has led scholars in various fields of research to explore the facets of this phenomenon in detail. One of the first contributions on this topic is the volume written by Johnson et al. (2010) which presents the results of an international survey on violence against women. The authors discuss the findings of over 23,000 interviews with women from eleven countries that talk about their experiences with gender-based violence. On the same note, Bandelli's (2017) monograph investigates the femicide phenomenon in Italy, by focusing on the discourses of contestation, on the simplistic representations of relationships between men and women, and more specifically, on the aggressor-victim connection.

Several works in the field of linguistics (especially corpus linguistics and discourse analysis) have focused on the topic of violence against women. For example, Sánchez-Moya (2017) explores the discourse of women survivors of intimate partner violence, by means of a corpus-driven examination of a dataset of 120,000 tokens collected from online forums. The paper displays the role of pronominal distribution in shaping the collective identity, as well as some differences in the emotional tone across the groups of victims taken into consideration. Similarly, Sánchez-Moya's (2021) paper analyses the discursive constructions used by female survivors of intimate partner violence when representing themselves and the perpetrators online. The work relies on the semantic categorisation of verbs, recurrent lexical patterns, and corpus statistics measurements (e.g., keyness) to identify these representations. The findings of Sánchez-Moya's (2021) work show that words such as 'abuser' and 'perp' are frequently used by victims to conceptualise male perpetrators. Moreover, the partner's aggressive behaviour is linguistically mirrored by the use of activity verbs such as 'grab', 'punch', and 'break'. Interestingly, some lexical choices adopted

by the victims suggest that they tend to conceptualise themselves largely through their roles as mothers.

Next, Busso et al.'s (2020) study employs tools and techniques from corpus linguistics and natural language processing to explore how femicide is framed in the Italian media. The results indicate that most crimes are perpetrated by intimate partners and that the media discourse around this topic increases in certain circumstances and moments of the year. There is also a tendency of not holding offenders accountable of their crimes. Moreover, the qualitative analysis of figurative language reveals that metaphors are used to delineate both the victims' and the perpetrators' socio-psychological traits. Finally, the episodes of violence are narrated using multimodality, more specifically by means of iconic speech and gestures.

The afore-mentioned studies and our work share the use of techniques from linguistics to explore the domain of discourse regarding gendered-related violence. Nevertheless, unlike previous research, we focus on discourses of the alleged perpetrators (suspects that were later found guilty and convicted), in an attempt to understand the linguistic expression of denial. In particular, through the lens of multimodal analysis and conversation analysis, we investigate how the crime is denied by exploring the discourses of a specific category of suspects of femicide, namely people familiar with the victims from long or close association.

This pilot study is exploratory and qualitative, to a large extent, and our main aim is to detect and describe recurrent verbal and nonverbal features used by the suspects to deny the accusations of femicide. To this end, we collect, transcribe, and annotate a multimodal corpus of over 10 hours (101,000 tokens) of public domain audio-video material of North-American English featuring guilty suspects of femicide. Considering the descriptive nature of this study and the data at hand, no generalisation claims can be advanced with respect to the differences between the strategies used to deny the allegations of femicide and those used to deny other crimes.

Our work attempts to contribute to the general understanding of the multimodal strategies employed to encode denial. In particular, we investigate how denial is communicated by alleged perpetrators (later found guilty) of femicide during police interviews and courtroom hearings. To this end, besides exploring the purely verbal expression of denial, we look at the bodily conduct of the suspects in order to map recurrent multimodal features (e.g., co-speech gaze, co-speech gesture, etc.).

Various studies have explored the role of multimodality in the expression of negation and have indicated that specific gestures and facial expressions are used to express distinct types of speech acts, including those associated with negation (see Prieto, Espinal, 2020, for an up-to-date review). However, previous works have not focused specifically on the multimodal expression of denial, let alone how it is encoded in the discourses of the suspects of crime (femicide or other). We believe that the data produced by the alleged perpetrators of femicide is particularly interesting because of the relationship they have with the victims. Although generalisations over the role of multimodal denial in outlining the profiles of perpetrators of femicide may

not be advanced from the data analysed here, we believe that this type of exploratory study may pave the way to more systematic research on the topic.

The paper is structured as follows: in section 2 we discuss about the theoretical tenets of this study; section 3 describes the corpus, the methods employed and the limitations of this study; results are presented and discussed in section 4; concluding remarks follow in section 5.

## 2. *The study*

### 2.1 Denial

Negation is a basic and essential phenomenon in language and its expression starts in early childhood (Morris, 2003). As is the case with most universal features of human languages, negation has sparked the interest of scholars from various fields of research, such as philosophy, logic, cognitive science, and linguistics. In fact, the beginning of the scholarship on negation dates back to Parmenides, roughly the early fifth century BCE, and over the following centuries it has been further developed by a multitude of scholars, such as Plato, Gottlob Frege, Otto Jespersen, Bertrand Russell, Ludwig Wittgenstein, and Paul Grice (see Speranza, Horn, 2010, for a historical review of negation).

The seemingly simple phenomenon of negation camouflages an elaborate system of forms of expression. Languages are endowed with all sorts of lexical, morphological, and syntactic devices that in combination with other features (e.g., modality) enable speakers to express negation. Adding a single negation feature (e.g., the negative adverb ‘not’) to a sentence will change greatly the message conveyed (e.g., ‘The wall is white.’ vs. ‘The wall is not white.’). In face-to-face communication, the verbal expression of negation is often complemented by nonverbal devices, such as prosody (e.g., speech rate, pitch, etc.) and the multimodal conduct (e.g., gestures, gaze, posture, etc., see also 2.2) – this latter dimension also occurring alone (e.g., in some cultural and linguistic groups head shaking is associated with negation).

According to Roitman (2017: 1), there are three main meanings of linguistic negation: non-existence (e.g., ‘There is no progress.’), rejection (e.g., ‘I do not want this.’), and denial (e.g., ‘It did not happen.’). The studies of Hummer et al. (1993) and Dimroth (2010: 45) indicate that from a developmental perspective, denial occurs later than other functions of negation, because it requires the simultaneous representation of two mental models, one for the true state of the world and one for its false counterpart.

Ripley (2020) investigates the relationship between negation and denial and he asserts that denying a certain claim means performing an act that introduces some information and this new information is the fact that the claim is ruled out. Generally speaking, denial has been regarded as a speech act that enacts objections

to utterances produced by previous interlocutors (van der Sandt, 1991)<sup>1</sup>. This implies that, generally speaking, denial cannot occur alone in natural contexts. However, denial can correct previous utterances that express negation, generating positive sentences.

Geurts (1998) identifies several mechanisms of denial and argues against the unitarian approaches proposed by Horn (1989) and van der Sandt (1991). According to Geurts (1998: 275), denials may be classified as follows: propositional denials (concerned with the content of a previous utterance), presupposition denials (the presupposition of a previous utterance), implicature denials (the implicature of a previous utterance), and form denials (the form, style, or register of a previous utterance).

In this paper we adopt an unitarian approach to denial; in particular, we use ‘denial’ as an umbrella term to refer to any kind of verbal or nonverbal linguistic expression employed by a speaker (let us call it ‘speaker X’) to object or correct the content, the presuppositions, the implicatures, and the form of a previous sentence uttered by another speaker (let us call it ‘speaker Y’).

## 2.2 Multimodal conduct

Before we operationalize denial as a multimodal speech act, it is worth taking a step back with the purpose of introducing and understanding the concept of multimodality. The term ‘multimodality’ is common to several fields of research in linguistics (e.g., conversation analysis, cognitive linguistics, construction grammar, etc.) and consequently it has been defined in many ways. As Mondada (2016: 338) claims, in conversation analysis, ‘multimodality’ is used to “refer to the various resources mobilized by participants for organizing their action – such as gesture, gaze, facial expressions, body postures, body movements, and also prosody, lexis and grammar”.

Following the definition above, we also adopt the concept of ‘multimodal conduct’ from Matoesian & Gilbert (2018) who in turn borrow it from Deppermann & Streeck (2018). In fact, the term appears in Deppermann & Streeck’s (2018) edited volume, which is dedicated to the study of embodied interaction. The book brings together several authors that discuss the complexities of multimodal discourse. These contributions reveal how speakers build and employ modal resources (e.g., speech, gesture, gaze, etc.) to perform social actions. More importantly, Deppermann & Streeck’s (2018) volume advocates for the use of multimodal analysis of video material in the study of social interaction. In their book, Matoesian & Gilbert (2018: 7) use ‘multimodal conduct’ interchangeably with ‘multimodal resources’ to refer to “the integration of speech, gesture, gaze, material artifacts, posture, and movement in the courtroom performance”. To

---

<sup>1</sup> Besides denial, other speech acts are associated with negation, namely rejection and metalinguistic negation (see Déprez, Espinal, 2020, for detailed contributions on this topic).

contribute to the understanding of these modes in legal interaction, the authors propose methods and tools for a systematic analysis of multimodality.

One of reasons why multimodal conduct has long been neglected in the research on legal language is that the analyses are complex and time-consuming, in a context where it is already difficult to investigate the content of the discourse alone. Multimodal analyses of spoken legal language imply the transcription and annotation of a whole range of features such as overlaps, pauses, hesitations, body positions and movements, etc. In turn, each of these features should be categorized in various classes (each having several levels). To make an example, let us consider gaze, namely a type of mode that has been annotated very differently in linguistic research, since (to the best of our knowledge) there are not any unanimous conventions available. For instance, some scholars only annotate the gaze direction (e.g., Kendon, 1967) while others use more granular annotation schemes, coding also the object that the eyes fixate (e.g., Somashekarappa et al., 2020).

Despite being a challenging analysis, the fine-grained investigation of multimodal conduct allows us to better comprehend the dynamics of legal interaction (e.g., between suspects and law enforcement representatives). On the same note, Matoesian (2010: 541) claims that verbal and nonverbal features work together “as co-expressive semiotic partners – as multimodal resources – in utterance construction and the production of meaning”. However, this understanding should not remain confined to the academic research, because a systematic identification, annotation, and analysis of features pertaining to multimodal conduct in the courtroom or in other legal settings may be relevant in shedding light on the interactants’ profiles.

As far as our study is concerned, we aim to demonstrate that suspects negotiate their identity and credibility also by means of multimodal resources. We believe that in all likelihood, multimodality has an effect on the pragmatics of speech acts and on their belief status. On this account, the speech act of denial is operationalized in the present work as a composite variable consisting of both a verbal and a nonverbal dimension that complete and enhance each other. This study explores the strategies employed by suspects to express denial during police interviews and courtroom hearings. In more general terms, this is an attempt to uncover and describe recurrent co-speech multimodal resources able to express denial.

Several studies have already investigated the role of multimodality in the expression of negation, but to the best of our knowledge, none has focused on how the suspects of crime encode denial. Prieto & Espinal’s (2020) review shows that the negation-related speech acts are expressed by means of various prosodic and gestural features across natural languages. As far as denial is concerned, Prieto & Espinal (2020) mention high tones (in tonal languages) and pitch accentual prominence (in intonational languages). An important contribution to the study of multimodal negation is Harrison’s (2018) monograph. He argues that “[n]egation is a linguistic universal with clear grammatical and gestural manifestations” (Harrison 2018: 1) and that there are regularities between grammar and gestures in human communication. Along the same lines, a recent study by Boutet et al. (2021) suggests



that children produce a set of shared gestures in combination with their target languages, generating a bimodal expression of negation. Finally, Bressemer & Müller's (2017) paper on the multimodal patterns of negative assessment demonstrates that recurrent gestures display a fixed form-meaning pairing.

In light of the afore-mentioned studies, investigating how suspects of crime (in our case, femicide) use recurrent verbal and nonverbal features to express denial in a range of legal contexts and interactions may improve our understanding of the universal phenomenon of negation. Moreover, this preliminary work may lay the foundations to more systematic linguistic research on the representation and perspectivation of the femicide phenomenon and on the suspect's responsibility.

### 3. Methodology

#### 3.1 Corpus

This work investigates the multimodal expression of denial by analysing the discourses of a specific category of suspects of femicide, namely individuals that are familiar with the victims from long or close association. For this purpose, we build an *ad-hoc* multimodal corpus of North-American English. We consider four suspects of solved cases of murders perpetrated to women in Canada and in the United States of America. Three suspects were married to the victims and one was the victim's friend. They were accused of killing the victims and one of them was charged of murdering both his ex-wives. Even though all suspects were eventually found guilty, they denied the accusations of murder in the audio-video material we considered in this study. For this reason, we refer to them by using the term 'suspects'.

The four videos are collected from Canadian and American open-access channels dedicated to crime news: The Fifth Estate<sup>2</sup>, Red Circle Interrogations and Confessions<sup>3</sup>, Law & Crime Trial Network<sup>4</sup>, Macon Telegraph Archive<sup>5</sup>. In one of these videos, the suspect is portrayed in a courtroom during evidentiary and judiciary hearings. The other three videos show suspects during police interrogations with law enforcement officers. Besides the four suspects, the other active participants in the interactions are police officers, a county sheriff, a district attorney, a judge and two lawyers. Table 1 summarizes the information regarding the corpus statistics.

<sup>2</sup> The channel is available at this webpage (accessed 21.10.2021): <https://www.cbc.ca/news/fifthestate>

<sup>3</sup> The channel is available at this webpage (accessed 21.10.2021): <https://www.youtube.com/c/RedCircleInterrogationsandConfessions>

<sup>4</sup> The channel is available at this webpage (accessed 21.10.2021): <https://lawandcrime.com>

<sup>5</sup> The channel is available at this webpage (accessed 21.10.2021): <https://www.youtube.com/user/Telegraph247>

Table 1 - *Corpus statistics*

<i>Suspect (anonymized acronym)</i>	<i>Relation with the victim(s)</i>	<i>Legal context</i>	<i>Other participants</i>	<i>Duration of the interaction (HH:MM:SS)</i>	<i>Number of tokens</i>
A.B.	wife	courtroom	district- attorney, judge, lawyers	01:54:17	17,682
C.D.	ex-wife	police interrogation	county sheriff	01:09:44	20,164
E.F.	ex-wives	police interrogation	police officer	04:58:24	46,076
G.H.	friend	police interrogation	police officers	02:03:39	17,234
Total				10:06:04	101,156

The overall size of the corpus amounts to over ten hours of multimodal material corresponding to 101,156 tokens, but the duration of each video and their number of tokens vary due to factors that do not depend on our methodological design.

Generally, the audio quality of the material is good enough to allow us to use software for automatic transcription. The four videos are in colour but they display different characteristics in terms of camera shots and angles<sup>6</sup>. The video of A.B.'s courtroom hearing is mostly a medium shot with an over the shoulder angle – enabling us to have a solid view of the suspect – but close-up shots are also used, and in those cases the suspect covers most of the screen. The medium shot is used also for the video of C.D., but this time with the high-angle technique; since the camera points down on the participants from above, the overview of the interrogation room is possible. For the police interrogations of suspects E.F. and G.H. the long shot and the high-angle techniques are used. All these characteristics allow us to explore and annotate in detail the suspects' multimodal conduct.

After data collection, we use Dragon Professional software<sup>7</sup> for English to automatically transcribe some of the audio-video material<sup>8</sup>. The generated transcripts are manually edited and corrected to match the content of the interviews and the participants' turns. Although the videos are part of the public domain and all cases are closed, we anonymize the data that we consider potentially sensitive (i.e., the names of suspects, of victims, and of other interactants; detailed geographical infor-

<sup>6</sup> For a review on camera shots and angles, see Savardi et al. (2021).

<sup>7</sup> An overview of the software and its characteristics is available at this webpage (accessed 20.11.2021): <https://www.nuance.com/dragon.html>

<sup>8</sup> The transcription methodology had to be adjusted along the way because of the COVID-19 pandemic and the unavailability of licensed software at home. Only part of the material was automatically transcribed with Dragon Professional. The remaining part was manually transcribed by the author with the support of S. Rizza and S. Casari – two trainees of Laboratorio Sperimentale (Department of Modern Languages, Literatures and Cultures, University of Bologna) – to whom we are immensely grateful.

mation, such as town and street names; dates; other information that may reveal the identities of the participants).

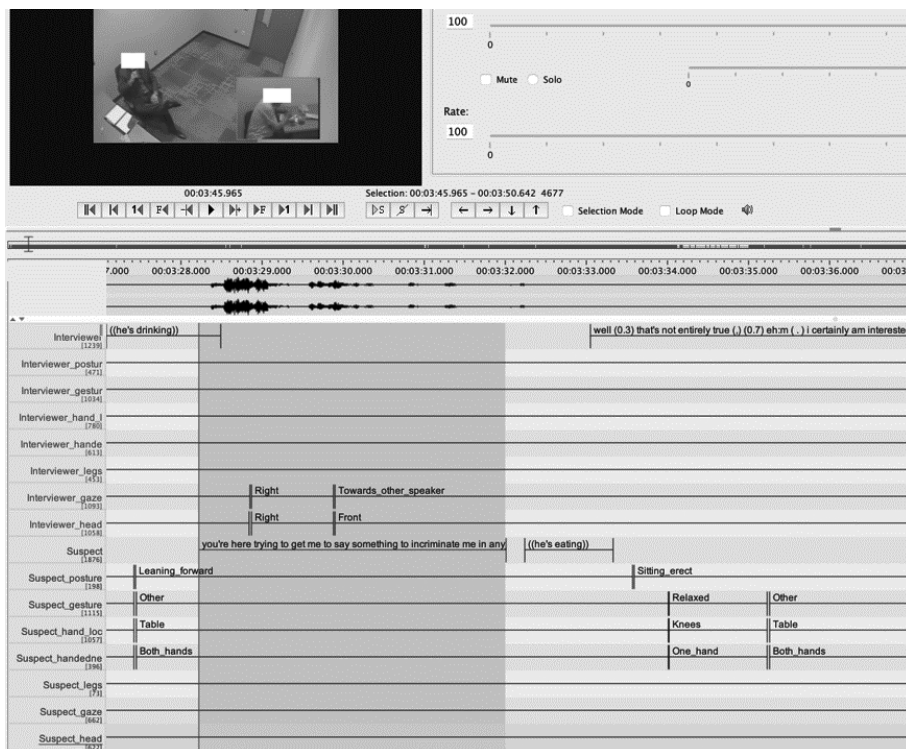
### 3.2 Methods

The edited transcripts of all speakers are imported into ELAN<sup>9</sup> and adapted to match Jefferson's (2004) transcription conventions. ELAN allows us to create tiers for each speaker, to define customised annotation schemes and levels for multimodal conduct, and to generate time stamps.

Even though the entire interactions between participants are transcribed, anonymized, and annotated, for our analyses we only consider the discourses of the suspects.

Therefore, in addition to the speech transcription, we elaborate *ad hoc* annotation schemes for five multimodal resources that we label as follows: gesture, gaze, posture, head, and legs. As displayed in Figure 1 (a screenshot of our transcription and coding in ELAN) each tier corresponds to a specific level of annotation.

Figure 1 - Example of ELAN transcription and annotation of multimodal conduct (own adaptation)



<sup>9</sup> An overview of the software and its characteristics is available at this webpage (accessed 20.11.2021): <https://archive.mpi.nl/tla/elan>

In this study, we use the term ‘gestures’ to refer exclusively to what is traditional known as ‘indexical (or deictic)’ gestures, ‘iconic (or lexical)’ gestures, and ‘beat (or motor)’ gestures produced spontaneously with the hands in face-to-face conversation (Holler et al. 2012; Kendon 2013). Under the macro level of ‘gestures’, we insert three levels of annotation: the hand shape (i.e., angled, cup, deictic, fist, gun, hole, intertwined, jailed, knife, loose, open and still, open moving onward, pursed, relaxed, steepled, peace sign, wall, other), the hand location (i.e., above the head, in the air far from the torso, in the air near the torso, near the ankles, touching the arms, near the back, near the chair, near the head, near the hip, near the knees, on the lap, near the neck, in the pocket, near the shoulders, on the table, touching the torso, touching the waist, other), and the handedness (i.e., one hand, both hands). Our coding scheme for gestures is a simplified free adaptation of the Gesture Studies Coding Manual created by the MIT Speech Communications Group<sup>10</sup>; it allows us to annotate the wide range of gestural manifestations in the corpus.

Next, the term ‘gaze’ is employed here to describe how the speakers use their eyes for regulatory and communicative functions during spoken interactions (Kendon, 1967). Under this category we only code the direction of the gaze (i.e., down, up, left, right, towards the other speaker, other).

Another multimodal resource annotated in this study is posture that we intend as the speaker’s body movements and position when they participate in the conversation (i.e., moving and leaning onward, moving left, moving right, retracting back, erect and sitting, shoulder shrugging, erect and standing, other).

We also code the movements and the position of the head (i.e., back, up front, left, right, still and erect, nodding, shaking, other) and of the legs (i.e., crossed ankles, foot over the thigh, leg over leg, moving legs while sitting, open legs, standing, tight legs, walking, wrapping legs around the chair, other).

Besides the multimodal conduct, we annotate hesitations, pauses, and overlaps. We employ corpus linguistics measurements (e.g., absolute and relative frequencies, concordances, etc.) to account for the ways denial is encoded verbally, namely by means of lexical and morpho-syntactical strategies (e.g., negative markers). This allows us to verify the expression of denial as a multimodal speech act.

The descriptive research proposed here has an important limitation, namely that it does not compare our data with data produced by suspects of crimes other than those involving violence against women. Therefore, we cannot establish whether the characteristics of denial, as they emerge from our data, depend on the type of crime or on other extralinguistic variables (e.g., the emotional state of the suspect, etc.). To the best of our knowledge, there are not any comparable open-access multimodal corpora on legal interactions and, at the time this research was conducted, building another resource on different types of crimes seemed highly impractical.

<sup>10</sup> The Gesture Studies Coding Manual created by the MIT Speech Communications Group is available at this webpage (accessed 01.12.2021): <https://web.mit.edu/pelire/www/manual/>

#### 4. Results and discussion

The first finding of this study reveals a trend on the verbal expression of denial. We observe that suspects make extensive use of lexical, morphological, and syntactic strategies of negation. In Table 2, we present the frequencies of the six most frequent negative markers in our corpus, as compared to their frequencies in the Open American National Corpus – only the spoken subcorpus is considered here (Ide & Suderman 2004)<sup>11</sup>. The frequencies reported for our corpus take into consideration only the suspects' turns and answers to the questions related to the crime. For comparability purposes, we display relative frequencies (per 1,000,000 words).

Table 2 - *The most frequent negative markers*

<i>Lemma</i>	<i>Relative Frequency (1,000,000 words)</i>	
	<i>Our corpus (suspects only)</i>	<i>Open American English Corpus (spoken)</i>
not	16,050	4,408
no	3,902	2,434
never	2,367	930
nothing	888	224
nobody	222	110
none	166	36

As shown above, some of the lemmas (e.g., 'not', 'nothing', 'non') occur up to four times more in the discourses of the suspects in our corpus as compared to the data in the Open American English Corpus. This may not be surprising considering the data at hand, but it strengthens our assumptions with respect to the fact that the suspects of femicide may deny the allegations, in an attempt to abandon their sense of responsibility.

We further investigate the verbal expression of denial in our corpus, by performing a qualitative analysis of a sample of forty turns (ten for each video) that contains at least one instance of morpho-syntactic negation. In most of these interactions, we find examples of reactive negation (77.50%) – typical cases of denial – such as those show in (1) and (2) below. The cases of initiative negation like the one in example (3) are less frequent (22.50%).

- (1) **Police officer:** she was gonna move out and live on her own (.)  
**Suspect:** NO (0.5) we were gonna move together and live together (.)

<sup>11</sup> While the Open American National Corpus is not a multimodal corpus of legal interactions, we believe that it is a valuable electronic resource of American English, as it consists of over 40 million tokens. The corpus contains data from texts of all genres and transcripts of recorded spoken data produced from 1990 onward. The corpus is fully open. A complete description of the corpus is available at this webpage (accessed 14.05.2022): <https://anc.org>

- (2) **District attorney:** a.b. you whispered cheating whore right there (?)  
**Suspect:** no (0.7) no (.)
- (3) **Police officer:** in case you need something else to eat (0.6) a cigarette  
**Suspect:** you know like ehm i don't hate my ex-wife (0.5) for what she's done to me (.)

Our qualitative exploration uncovers another interesting finding. In most cases, the reactive negations are supplemented by the repetition strategy. Two of such examples are (2) above and (4) below – the latter is particularly interesting because all negative sentences are repeated twice or three times.

- (4) **Police officer:** but you know that's gonna be in the paper (.)  
**Suspect:** no i'm not not i'm not (0.5) i can't change that of course (1.0) and there's nothing i can do to change this (0.5) ((he is sobbing)) no i can't (0.9) i can't tell you anything other than what i've told you (.)

Another feature that emerges from the analysis of the sample of turns is the suspects' preference for the anaphora, even when the referents are not easily retrievable in the context of the narratives. Victims are rarely called by their names; instead, suspects use the entire set of anaphoric third person pronouns (e.g., 'she', her), possessives (e.g., 'her', 'hers'), and reflexives (e.g., herself). An interesting example is the one shown in (5) below, as the suspect does not refer to his ex-wife by her name during the entire hour of interrogation. He deploys anaphoric expressions such as "that woman" to distance himself from the murder accusations.

- (5) **Suspect:** i didn't do anything to that woman (0.9) i didn't hurt her (.)

In addition, the corpus exploration reveals other (less frequent) strategies employed by the suspects to deny the allegations of murder. The turns of suspects A.B. and G.H. have a reduced average sentence length (i.e., around 3 words per sentence). These two suspects are also those whose speech acts (including denial) are the vaguest. An example is provided in (6) where it is not clear to which question the suspect answers.

- (6) **Lawyer:** did you confront them about the affair (?) did you (?) had you concluded there wasn't an affair (?)  
**Suspect:** no i did not (.)

The description and the discussion of our results continue with the part dedicated to the multimodal resources. Table 3 displays the number of items coded for each suspect.

Table 3 - *Multimodality resources*

Suspect (anonymized acronym)	Number of items
A.B.	1,674
C.D.	1,756
E.F.	4,122
G.H.	351
Total	7,903

The number of items representing instances of multimodality amounts to 7,903, but their distribution in the corpus varies greatly. In particular, if we consider the duration of each video and the related number of tokens, it is clear that, proportionally, suspect G.H. employs the lowest number of multimodal features, while, on the contrary, suspect C.D. the highest number of these features.

In order to account for the multimodal expression of denial, we identify all the turns that contain markers of negation and we manually control whether suspects use simultaneously co-speech multimodal resources in these cases. All these instances are classified and counted. Table 4 displays the most frequent multimodal features (and the related number of items for each suspect) that co-occur with denial in our corpus.

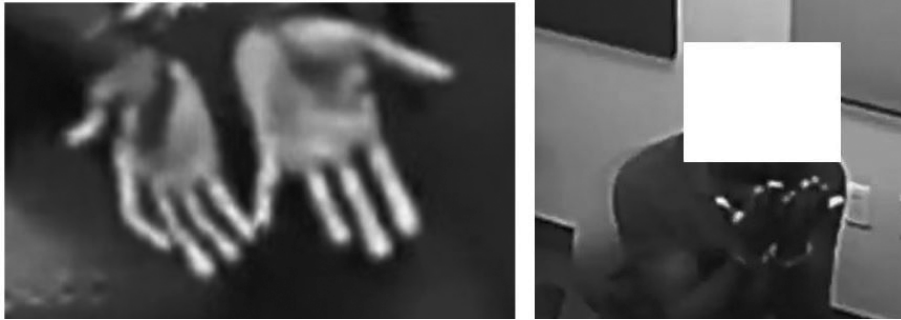
Table 4 - *Multimodal denial*

<i>Suspect (anonymized acronym)</i>	<i>Gesture: Open hands still / moving onward</i>	<i>Gesture: Steepled hands</i>	<i>Head: Shaking</i>	<i>Posture: Shoulder shrugs</i>	<i>Posture: Leaning or moving onward</i>	<i>Gaze: Towards the other participant</i>
A.B.	6	3	37	7	15	28
C.D.	74	21	71	12	14	82
E.F.	48	2	13	5	31	176
G.H.	2	0	1	1	4	29

This analysis unveils several interesting trends. First of all, the verbal dimension of denial is complemented by a set of recurrent co-speech multimodal features: headshakes, open hands (either still or moving onward), steepled hands, shoulder shrugging, leaning or moving onward, and the gaze directed towards the other participant in the legal interaction.

Overall, the results are consistent with previous research on the multimodal expression of negation. According to Bross (2020: 18), the negative headshake is a “statistical universal” acquired during early infancy and, just like other gestures, it may have roots in “object manipulation or action, especially within parent-child interaction”.

As far as the gestures are concerned, the most frequent is the ‘open hands’ gesture (see Figure 2). The hands are either still or moving and their location is generally in the air near the torso, near the lap, on the lap, and on the table. Previous research has reported that the ‘open hands’ gesture (part of the Open Hand Prone gestures family) is usually linked to spoken expression of negations (Kendon 2004). The other co-speech gesture associated with denial in our corpus is the ‘steepled hands’. In this case, the hands are located near the head (more specifically, near the face).

Figure 2 - *Open hands gestures (own adaptation)*

A co-speech multimodal feature that frequently complements denial in our corpus regards the direction of the gaze. In particular, all speakers tend to fixate their interlocutors during denial. One such example is provided in Figure 3 and it corresponds to the question-and-answer exchange in (7). The suspect is sitting still on the chair for almost two hours, his head is still and erect, his hands are relaxed on the table, and his gaze is always directed towards the detective that is asking questions. In all probability, the suspects in our corpus adopt the strategy of looking squarely into the eyes of their interlocutors in an attempt to appear honest and credible.

Figure 3 - *Gaze towards the other speaker (own adaptation)*



- (7) **Police officer:** when was the last time you shot those guns (?)  
**Suspect:** i haven't (.)  
**Police officer:** you've never shot a gun (?)  
**Suspect:** no (.)  
**Police officer:** have you ever shot any gun in your whole life (?)  
**Suspect:** no (.)  
**Police officer:** NEVER (?)  
**Suspect:** no (.)  
**Police officer:** so you own three guns and you've never shot (?)  
**Suspect:** yes (.)  
**Police officer:** wow (.)

Two cues pertaining to the posture set of multimodal features co-occur with denial: shoulder shrugs and the torso leaning or moving onward. Debras (2017: 24) sustains that shoulder shrugs “have to do with the subjective expression of negation”. A qualitative exploration of the shoulder shrugs in our corpus suggest that they are often combined with raised eyebrows – not coded in this study. This is also consistent with previous research, as in the literature, the propensity of the shrug feature to combine with other modes is known as the “compound enactment” (Morris 2004: 165; Debras 2017: 2).

Finally, but not surprisingly, our data do not reveal any interesting trend regarding the movements and the position of the legs.

All in all, our study shows that the speech act of denial may be operationalized as a composite variable consisting of a verbal and a multimodal dimension that co-exist and go along in face-to-face interactions. While we cannot establish whether our findings are relevant, in general, for the speech act of denial, the results discussed above suggest that all suspects analysed in our corpus deny the accusations of by means of multimodal resources. These findings call for more systematic studies in linguistics on multimodality (and on languages other than English) in legal interactions.

### 5. *Conclusions and further research*

Starting from a case study of interviews with guilty suspects of femicide – the victims' (ex-)intimate partners, family members, or friends – this paper explored the multimodal dimension of denial. The data produced by these individuals were considered interesting precisely because of the close relationship they had with the victims. The objective of this descriptive and, to a great degree, qualitative research was to identify recurrent verbal and nonverbal features used by suspects to deny the allegations of crime. The analyses are supported by an *ad hoc* multimodal corpus of over 10 hours (101,000 tokens) consisting of open-access and public domain audio-video material produced by speakers of North-American English.

We showed how suspects expressed denial in two legal contexts (i.e., police interviews and courtroom hearings). The findings suggest that denial is constructed by means of verbal and nonverbal strategies that complete and strengthen each other.

Besides the extensive use of morpho-syntactic realization of negation (negative markers in our corpus were more frequent than those in a reference corpus of American English), suspects deny the accusations through several recurrent co-speech multimodal resources, such as headshakes, shoulder shrugs, gaze towards the interviewer, and open hands (either still or moving forward). We also observed that in our corpus denial co-occurs with repetitions, anaphora, vagueness, and reduced sentence length.

All in all, our results indicate that the speech act of denial is much more than what verbal features could express. Suspects negotiate their credibility through multimodal and verbal resources simultaneously and, to some extent, equally. Our study suffered from limitations common to qualitative methods in linguistics, which, however, provide opportunities for improvement in future research. In particular, the analyses were limited to the data we collected and annotated: due to the lack of comparable corpora, we were not able to adopt a contrastive approach (e.g., taking into consideration denial in discourses produced by suspects of other crimes or by speakers in other communicative contexts). At this stage, we cannot ascertain that the characteristics of denial, as they emerge from our findings, depend on the type of crime or on other extralinguistic variables (e.g., the emotional state of the suspect, legal context, etc.).

For all these reasons, we believe that additional systematic research on multimodality in legal interactions, in other communicative contexts, and in more languages is necessary. Moreover, research on multimodality should not remain confined to the scholarly debate; analysing the multimodal conduct in legal settings may be advantageous in providing insights on the interactants' profiles.

### *Bibliography*

- BANDELLI, D. (2017). *Femicide, gender and violence: discourses and counterdiscourses in Italy*. Cham: Palgrave Macmillan.
- BOUTET, D., BLONDEL, M., BEAUPOIL-HOURDEL, P. & MORGENSTERN, A. (2021). A multimodal and kinesiological approach to the development of negation in signing and non-signing children. In *Languages and Modalities*, 1, 31-47.
- BRESSEM, J. & MÜLLER, C. (2017). The “Negative-Assessment-Construction” – A multimodal pattern based on a recurrent gesture? In *Linguistics Vanguard*, 3(s1), article no. 20160053, 1-9.
- BROSS, F. (2020). Why do we shake our heads? On the origin of the headshake. In *Gesture*, 19(2/3), 269–298.
- BUSSO, L., COMBEI, C.R. & TORDINI, O. (2020). The mediatization of femicide: a corpus-based study on the representation of gendered violence in Italian media. In *L'analisi linguistica e letteraria*, 28(3), 29-48.
- CAPUTI, J. & RUSSELL, D.E.H. (1992). Femicide: Sexist Terrorism against Women. In RADFORD, J. & RUSSELL, D.E.H. (Eds.), *Femicide: The Politics of Woman Killing*. New York: Twayne Publishers, 13-21.

- DEBRAS, C. (2017). The shrug: Forms and meanings of a compound enactment. In *Gesture*, 16(1), 1–34.
- DEPPERMAN, A. & STREECK, J. (Eds.) (2018). *Time in Embodied Interaction*. Amsterdam: John Benjamins.
- DÉPREZ, V. & ESPINAL, M.T. (Eds.) (2020). *The Oxford Handbook of Negation*. Oxford: Oxford University Press.
- DIMROTH, C. (2010). The Acquisition of Negation. In HORN, L.R. (Ed.), *The Expression of Negation*. Berlin: De Gruyter Mouton, 39–71.
- EIGE (2017). Glossary of definitions of rape, femicide and intimate partner violence. [https://eige.europa.eu/sites/default/files/documents/ti\\_pubpdf\\_mh0417297enn\\_pdf-web\\_20170602161141.pdf](https://eige.europa.eu/sites/default/files/documents/ti_pubpdf_mh0417297enn_pdf-web_20170602161141.pdf). Accessed 07.01.2022.
- GEURTS, B. (1998). The Mechanism of Denial. In *Language*, 74(2), 274–307.
- HARRISON, S. (2018). *The Impulse to Gesture Where Language, Minds, and Bodies Intersect*. Cambridge: Cambridge University Press.
- HOLLER, J., TURNER, K. & VARCIANNA, T. (2012). It's on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts. In *Language and Cognitive Processes*, 28(10), 1509–1518.
- HORN, L.R. (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- HUMMER, P., WIMMER, H. & ANTES, G. (1993). On the origins of denial negation. In *Journal of Child Language*, 20, 607–618.
- IDE, N. & SUDERMAN, K. (2004). The American National Corpus First Release. In CALZONARI, N. ET AL. (Eds.), *Proceedings of the 4<sup>th</sup> Language Resources and Evaluation Conference (LREC2004)*, Lisbon, 26–28 May 2004, Paris: The European Language Resources Association (ELRA), 1681–1684.
- JEFFERSON, G. (2004). Glossary of transcript symbols with an introduction. In LERNER, G.H. (Ed.), *Conversation Analysis: Studies from the First Generation*. Amsterdam: John Benjamins, 13–31.
- JOHNSON, H., OLLUS, N. & NEVALA, S. (2010). *Violence against women: an international perspective*. New York: Springer.
- KENDON, A. (1967). Some functions of gaze direction in social interaction. In *Acta Psychologica*, 26, 22–63.
- KENDON A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- KENDON, A. (2013). History of the Study of Gesture. In KEITH, A. (Ed.), *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press, 1–20.
- MATOSIAN, G.M. (2010). Multimodal aspects of victim's narrative in direct examination. In COULTHARD, M. & JOHNSON, A. (Eds.), *The Routledge Handbook of Forensic Linguistics*. London: Routledge, 541–557.
- MATOSIAN, G.M. & GILBERT, K.E. (2018). *Multimodal conduct in the law: language, gesture and materiality in legal interaction*. Cambridge, UK: Cambridge University Press.
- MONDADA, L. (2016). Challenges of multimodality: Language and the body in social interaction. In *Journal of Sociolinguistics*, 20(3), 336–366.

- MORRIS, B.J. (2003). Opposites Attract: The Role of Predicate Dimensionality in Preschool Children's Processing of Negations. In *Journal of Child Language*, 30, 419-440.
- MORRIS D. (2004). *Bodytalk: A worldguide to gesture*. London: Johnatan Cape.
- PRIETO, P. & ESPINAL, M.T. (2020). Negation, Prosody, and Gesture. In DÉPREZ, V. & ESPINAL, M.T. (Eds.), *Oxford Handbook of Negation*. Oxford: Oxford University Press, 1-20.
- RIPLEY, D. (2020). Denial. In DÉPREZ, V. & ESPINAL, M.T. (Eds.), *Oxford Handbook of Negation*. Oxford: Oxford University Press, 1-13.
- ROITMAN, M. (2017). Introduction. In ROITMAN, M. (Ed.) *The Pragmatics of Negation*, Amsterdam: John Benjamins, 1-14.
- SÁNCHEZ-MOYA, A. (2017). Corpus-driven insights into the discourse of women survivors of Intimate Partner Violence. In *Quaderns de Filologia: Estudis Lingüístics*, 22, 215-243.
- SÁNCHEZ-MOYA, A. (2021). From the uncertainty of violence to life after abuse: Discursive transitions among female survivors of Intimate Partner Violence in online contexts. In *Research in Corpus Linguistics*, 9(2), 152-178.
- SAVARDI, M., KOVACS, B A., SIGNORINI, A & BENINI, S. (2021). CineScale: A dataset of cinematic shot scale in movies. In *Data in Brief*, 36, 107002, 1-9.
- SOMASHEKARAPPA, V., HOWES, C. & SAYEED, A. (2020). An Annotation Approach for Social and Referential Gaze in Dialogue. In CALZONARI, N. ET AL. (Eds.), *Proceedings of the 12<sup>th</sup> Language Resources and Evaluation Conference (LREC2020)*, Marseille, 11-16 May 2020, Paris: The European Language Resources Association (ELRA), 759-765.
- SPERANZA, J.L. & HORN, L.R. (2010). A brief history of negation. In *Journal of Applied Logic*, 8(3), 277-301.
- UN (2013). Statement submitted by the Academic Council on the United Nations System, a non-governmental organization in consultative status with the Economic and Social Council. Vienna Declaration on Femicide. [https://www.unodc.org/documents/commissions/CCPCJ/CCPCJ\\_Sessions/CCPCJ\\_22/\\_E-CN15-2013-NGO1/E-CN15-2013-NGO1\\_E.pdf](https://www.unodc.org/documents/commissions/CCPCJ/CCPCJ_Sessions/CCPCJ_22/_E-CN15-2013-NGO1/E-CN15-2013-NGO1_E.pdf). Accessed 07.01.2022.
- UNODC (2020). Killings of women and girls by their intimate partner or other family members: Global estimates 2020. [https://www.unodc.org/documents/data-and-analysis/statistics/crime/UN\\_BriefFem\\_251121.pdf](https://www.unodc.org/documents/data-and-analysis/statistics/crime/UN_BriefFem_251121.pdf). Accessed 07.01.2022.
- VAN-DER SANDT, R. (1991). Denial. In *Chicago Linguistic Society*, 27(2), 331-344.

LUCIA BUSSO

## *CorIELLS*: a specialised bilingual corpus of English and Italian legal-lay language

This contribution presents *CorIELLS* (Corpus of Italian and English Legal-Lay textS). *CorIELLS* is an open-access bilingual corpus of legal-lay language for Italian and English. The corpus rationale, collection, and composition are discussed, together with previous research on the corpus in both Italian and English. Furthermore, a cross-linguistic exploration of lexical and syntactical linguistic complexity is presented, using the *Profiling-UD* tool (Brunato et al., 2020) is further presented. Dimensions of complexity are investigated with Principal Component Analysis (Kassambara, Mundt 2020). Results for the two languages are compared, and similarities and differences in dimensions of complexity are foregrounded.

*Keywords*: corpus linguistics; legal-lay language; language complexity

### 1. Introduction

The present contribution describes a new linguistic resource for the analysis of an under-researched genre: legal-lay language. We define here this textual type as the language of any type of document with legal content aimed at non-specialist audiences (Crystal, Davy, 1969; Tiersma, 1999). Legal-lay language (henceforth: LLL) is increasingly important in our everyday life. Particularly, one constantly finds oneself in need of understanding some sort of legal document, from consumer contracts to privacy notes and terms and conditions of websites.

Since research on LLL as an independent genre is still scarce, this contribution presents the first specialised corpus on LLL, *CorIELLS*, and illustrates some types of research that can be conducted on it. Specifically, the remaining of section 1 (sections 1.1. and 1.2) outline more in depth the background literature and the motivations of the research. In section 2 we describe the corpus collection, its rationale, and the corpus composition. Section 3 outlines previous research that was conducted using the *CorIELLS* (3.1) and presents an innovative analysis of linguistic complexity using dimensionality reduction (3.2). The discussion in section 4 concludes the paper, summarizing the core notions presented and opening new lines of possible research.

#### 1.1 Why legal-lay language?

It is widely acknowledged in the scientific literature that lack of understanding of legal-lay language leads to problems legislative and linguistic alike (Gibbons, 2003;

Curtotti, McCreath, 2013; Benoliel, Becher, 2019). The many issues arising from an incomplete understanding of LLL has been foregrounded by many scholars. For example, Tiersma (1999: 20) claims that “people have the right to know the meaning of the contracts that they sign and for which they will be held legally responsible. When people are entitled to understand a legal document, it should be as free as possible of technical terms and jargons. If technical terms are avoidable, they should at least be explained in ordinary language”. Frade (2007, 2016) also identifies asymmetrical power dynamics between powerful business companies and their customers, with the formers holding a “hidden power”. This notion is cognate to Fairclough’s (2011) concept of ‘technologization of discourse’ – i.e., the manipulation of social practices by the more powerful social force.

Related to this issue, consumers are not keen or motivated in reading the contract provisions or legal notice documents, thus committing themselves to an often legally binding relation they are not aware of.

Despite the centrality of LLL in the contemporary world, studies dedicated specifically to this textual type specifically are not common in the literature (Brunato, Venturi 2014; Lintao, Madrunio, 2015; Van Boom et al., 2016; Conklin et al., 2019). In fact, the scientific literature often assumes that LLL is not a textual type independent from legal language, but only its ‘simplified’ and more accessible version (Bhatia, 1983; Venturi, 2011; Adler, 2012). Particularly, the *Plain Language movement* (Williams, 2004a; Adler, 2012) and the Italian version *Progetto Chiaro* (Williams, 2005) have been advocating for decades the need for a plainer language in drafting legal documents directed at ordinary citizens. This is a difficult balance to obtain, as LLL must crucially preserve its intended legal meaning (Kimble, 2000; Eagleson, 2004) while at the same time be drafted in a language plain enough for non-legal specialists to read and understand.

We argue that this delicate balance between legal content and plainer form constitute *de facto* a novel textual type, autonomous and independent from its ‘parent’ genre, legal jargon. Although the precise connotations of the concept ‘genre’ are eluding and there is no consensus in the literature, one of the most famous definitions of it states that:

genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognised by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style (Swales 1990: 58).

Similarly, Bhatia (2004: 23) notes that

[g]enre essentially refers to language use in a conventionalised communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to stable structural forms by imposing constraints on the use of lexicogrammatical as well as discoursal resources.

Therefore, we argue that the differences in audience and community of practice, the different aims, the conscious effort to simplify both lexicon and grammar, and the different medium (given that LLL is mainly found on the Internet or in consumer contracts) makes it reasonable to consider LLL as an idiosyncratic genre (Busso, 2022a; Busso, 2022b).

Given the scarcity of linguistic studies on LLL as a separate entity from legal language in the literature and the absence of dedicated linguistic resources, the present paper presents the first specialised bilingual corpus of LLL: *CorIELLS* (Corpus of Italian and English Legal-Lay textS). The corpus was originally collected to investigate linguistic complexity in LLL, but it also lends itself to different analyses. Corpus files are freely available from download from the Forensic Linguistics Databank (FoLD) (Petyko et al. 2022, [www.fold.ac.uk](http://www.fold.ac.uk)).

## 1.2 Why a legal-lay language corpus?

As argued by several scholars in the literature, compiling relatively small and specialised corpora – instead of mega corpora of billions of words – allows researchers to investigate fine-grained aspects of language, like a specific register or genre (Flowerdew 2002; Hyland, 2006; Koester, 2010). In fact, while many larger corpora are suitable to research general linguistic phenomena and aim to represent a larger variety of language (e.g., British English), specialised corpora are generally catered to more specific research questions. In specialised corpora,

much closer link between the corpus and the contexts in which the texts in the corpus were produced. Where very large corpora, through their de-contextualisation, give insights into lexico-grammatical patterns in the language as a whole, smaller specialised corpora give insights into patterns of language use in particular settings (Koester, 2010: 67).

Therefore, to investigate features of LLL a new corpus was collected, following Flowerdew (2004) guidelines to build specialised corpora for specific genres and Biber (1993)'s insights on situational (i.e., the range of text types) and linguistic (i.e., the range of linguistic distributions) representativeness. By including a wide variety of text types in the corpus (see section 2 below), the aim is to yield insights not only into the particular corpus, but also into more vast language use in the genre of LLL (Tognini Bonelli, 2001).

## 2. The *CorIELLS* corpus and its use

The corpus was compiled as to be partially parallel between Italian and English. That is, documents included in the final corpus selection are both comparable (i.e., same content in the two languages) and idiosyncratic texts (i.e., same document type but different content). The part of the corpus designed to be parallel only includes document drafted by two different legal teams in the two languages. This prevents – to the best of the author's possibilities – spurious phenomena related to specialised

translation (Matulewska, 2007; Biel, 2009, 2018). Moreover, many scholars have claimed that different versions of the same text can be considered autonomous texts, as they are equally “authentic but not necessarily based on translations from one or the other” (Williams, 2004b: 219). The independence of translated texts is also advocated from within translation studies, as translations should be regarded as “autonomous texts destined to function in the context of the target culture without regard for their relation to the source text” (Garzone, 2008: 48).

The final selection of text types includes four major categories of documents: standard legal notices for bank accounts, summaries of European legislation, terms and conditions of websites, standard consumer contracts for utilities (mobile phone contracts, gas, and electricity). Table 1 outlines corpus composition in the two subcorpora of Italian and English texts.

The corpus was collected semi-automatically with the web-scraping toolkit *Bootcat* (Baroni, Bernardini, 2004).

Final corpus size amounts to over 1.6M words. Despite being relatively small and highly specialized in nature, *CorIELLS* aims at obtaining a representative sample of the genre in question (Biber, 1993; Koester, 2010), while maintaining as a strict parameter of inclusion only freely accessible online texts. In this way, we approximate the types of LLL than any ordinary person could obtain by browsing the web.

Section 2.1 will describe in more detail the selection and collection procedure for each subsection, while section 2.2 will present previous work using the corpus as background to the study here described in section 3.

Table 1 - *Composition of the CorIELLS corpus*

<i>Document type</i>	<i>Document no</i>	<i>English subcorpus</i> (740K words)	<i>Italian subcorpus</i> (880K words)	<i>Comparable</i>
Bank account contracts	15 in total	19%	27%	no
Utilities contracts	15 per language	25%	23%	no
<i>EurLex</i> summaries	247 per language	22%	23%	yes
Terms and conditions	45 per language	34%	27%	yes

## 2.1 Corpus selection

As mentioned, four major categories of documents compose *CorIELLS* (table 1). The composition of the corpus is by no means exhaustive of the different types of texts that can be included under the umbrella term of LLL. The final selection represents a representative sample of document types freely available online. All document types that it was not possible to retrieve freely were automatically excluded. This criterion is underpinned by the assumption that if a document is not available for download/view on the open web, it may very well only shown by professionals explaining the



wordings and the main clauses and provisions (for example, insurance contracts). Table 2 below reports all document sources for the corpus.

The selected documents are:

1. Standard legal notices for bank current accounts. A selection of 15 common bank institutes in both Italy and the UK were chosen (7 for Italian and 8 for English, given the different lengths of the documents). As mentioned, only standard contracts for current accounts freely available online were selected. The documents were first retrieved in their PDF form and text files were obtained using *Bootcat*.
2. Utilities standard contracts. For this section, 15 widely common companies for every-day utilities in Italy and the UK were selected. Standard contract terms for 5 energy supply (gas and electricity), 5 Wi-Fi suppliers, and 5 pay-by-month phones were downloaded as PDF and transformed into text files using *Bootcat*.
3. Terms and conditions (or terms of use) of websites. To have a representative and balanced set of data, the list of the 500 most visited websites in Italy and the UK in 2019 was used. We then selected websites that were present in both lists and only web services with legal notices in both languages were kept. Furthermore, the websites were manually checked to make sure that the two legal notices had been drafted by two legal teams. When terms of use and terms and conditions were in different pages, both were selected. A final selection of 45 websites were included in the corpus. Text files were extracted from the URLs of the terms and conditions pages using the URL collection function on *Bootcat*.
4. European legislation summaries. These texts were retrieved from the EUR-Lex website (<https://eur-lex.europa.eu/homepage.html>). Summaries are defined as “short, easy-to-understand explanations of the main legal acts passed by the EU—intended for a general, non-specialist audience” (*EUR-lex* website). All summaries from 2019 and 2020 (as of December 1st, 2020) were retrieved from the official website (using the Advanced Search function), in both their Italian and English version. These documents are originally drafted in English and later adapted by specialised translators and legal experts in each language of the European Union, as prescribed in EU style guides (Interinstitutional Style Guide, 2015:54-62). In this case as well, URLs were inputted into *Bootcat* to extract their plain text version.

Table 2 - Documents retrieved in corpus selection

<i>Corpus section</i>	<i>Documents</i>
Bank contracts	<p><b>ITA:</b> Illimity, BPM, Finecobank, Banca Mediolanum, Deutsche Bank, Banca Sella, Banca d'Italia.</p> <p><b>ENG:</b> RBS, N26, Starling Bank, Monzo, HSBC, Lloyds, Nationwide, Revolut.</p>

<i>Corpus section</i>	<i>Documents</i>
Utilities contracts	<p><b>ITA:</b>            Energy: Enel Energia, Alleanza Luce&amp;Gas, AGSM, Illumia, Iren;            Phone: Vodafone IT, Tim, Iliad, Wind, Tre;            Wi-fi: Tiscali, Tim ADSL, Vodafone Casa, Wind ADSL, Fastweb.</p> <p><b>ENG:</b>            Energy: Southern Electric, EDF Energy, Gazprom, Go Power, Npower;            Phone: EE, GiffGaff, O2, Tesco Mobile, Vodafone UK;            Wi-fi: NOW TV, BT broadband, SKY UK, Vodafone Broadband</p>
T&C	AirBNB (terms of use + t&c), Alibaba (terms of use + t&c), Amazon, Asos, Booking, DropBox, EasyJet, Ebay, Etsy, Europeancommission, Facebook, Fandom, Firefox, Google, Groupon, H&M, HP, Ikea, Instagram, JustEat, LinkedIn, LiveJasmin, Livesport, Microsoft, Netflix, Pinterest, Primevideo, Ryanair, Samsung (terms of use + t&c), Shein, Skyscanner, Spotify, Tripadvisor, Trustpilot, Tumblr, Twitch, Twitter, Vice, WhatsApp, Wikimedia Foundation, Wikipedia, Yahoo, Youtube.
EurLEX summaries	All summaries from January 2019 to November 2020 at <a href="https://eur-lex.europa.eu/eu-summary/eu-summary-search.html">https://eur-lex.europa.eu/eu-summary/eu-summary-search.html</a>

As mentioned, the corpus files are freely available from download from the Forensic Linguistic Databank (FoLD, <https://fold.aston.ac.uk/>), an innovative online repository hosted by Aston University<sup>1</sup>.

## 2.2 Previous research on legal-lay language

A couple of studies (Busso, 2022a, 2022b) have used CorIELLS to investigate different aspects of LLL in terms of its lexico-grammatical features and its comprehensibility.

Specifically, Busso (2022) presents a quantitative text-based analysis rooted in the core principles of Construction Grammar (Goldberg, 2006, 2019). The study analyses four constructions typical of Italian and English legal jargon and LLL (Mortara-Garavelli, 2001; Williams, 2004; Chovanec, 2013; Mori, 2019). The same constructions are also used for the study in Busso (2022b):

1. Modal verbs (MOD)
  - a. *I paesi UE devono notificare alla Commissione eventuali obblighi e requisiti in materia di comunicazione* (EU countries must notify the Commission of any communication obligations and requirements)
  - b. We must be satisfied of your identity and can refuse instructions if we doubt your identity
2. Nominalisations heading PP attachment chains NOM-pp
  - a. *Informazioni sul sistema di risoluzione delle controversie di cui alla Delibera Consob* (Information on the system of resolution of litigations referred to in the Consob resolution)

<sup>1</sup> FoLD is a “permanent, controlled access online repository for forensic linguistic data” (Petyko et al., 2022).

- b. Mandatory collective management of rights for retransmissions of radio and television programmes by means other than cable.
- 3. Reduced participial relatives PART
  - a. *Nota informativa concernente il trattamento e la protezione dei dati personali.* (Information note treating the processing and protection of personal data)
  - b. The ‘application publisher’ means the entity licensing the application to you as identified in the Store.
- 4. Passive constructions PASS
  - a. *Il reclamo può essere presentato anche dopo la data di entrata in vigore della variazione.* (The complaint can also be submitted after the effective date of the change)
  - b. Payments (...) will be sent on the next working day.

First, a collostructional analysis (Stefanowitsch, 2013) is performed on the selected set of constructions, and statistically significant results are contrasted with the ‘Nuovo Vocabolario di Base’ (De Mauro, Chiari, 2016). Secondly, results from the first part of the study are further compared to neighbouring genres (i.e., legal jargon and written prose) in a contrastive frequency analysis. These genres are represented with the CORIS corpus (Rossini-Favretti, 2002). Specifically, the legal and narrative subcorpora are used to approximate respectively legal jargon and general domain written prose.

Findings from this study suggest that LLL is a ‘blended’ genre containing features from both specialistic and non-specialistic registers. The contrastive analysis moreover shows that some lexico-grammatical features are used significantly differently in LLL than in both legal jargon and general-domain written prose. Specifically, all the analysed constructions behave differently in LLL except for NOM-pp (see examples 1-4 below) – which seem to be the most tightly related to the legal genre.

Busso (2022) performs a similar analysis on the English subcorpus of CorIELLS, using the same structures and the New General Service List (Brezina, Gablasova, 2015) as a core vocabulary. Both collostructional analysis and comparative analysis with legal jargon and written prose were performed. For English, CorIELLS is contrasted to an ad-hoc subcorpus of the EurLEX corpus (Baisa et al., 2016) which includes only legislative texts spanning from the 90s to 2015, and to the imaginative subcorpus of the BNC (BNC, 2007).

Findings are comparable with Italian, in that NOM\_pp is the most specialist construction out of the four investigated. Moreover, results also show that LLL displays linguistic features quantitatively different from the other two genres. Moreover, a representative sample of concordances is analysed in terms of text-based readability metrics and presented to native speakers in a survey. The comparison between results from these two methods (i.e., readability scores and speakers’ judgments) suggests that readability metrics might underestimate the readers’ ability to understand LLL texts.

### 3. *Language complexity in CorIELLS: an exploratory analysis*

As outlined in paragraph 2.2 above, previous research using the CorIELLS corpus has explored the lexico-grammatical level and readability. In the present contribution we expand the existing exploration of this corpus by presenting an analysis of lexico-syntactic complexity of the CorIELLS corpus as compared to legal jargon and written prose, in both Italian and English.

As anyone working with the multifaceted notion of linguistic complexity knows, “complexity” is an ill-defined phenomenon. Many scholars in linguistics have given different definitions and quantitative measures of it (among many others: Benoit, 1990; Kusters, 2003, 2008; Pallotti, 2015). In line with Pallotti (2015), we advocate for a clear-cut working definition of complexity, «treating it as a purely descriptive category, limiting its use to structural complexity and excluding from its definition any theoretical assumption about when, how and why it increases or remains constant» (Pallotti, 2015: 119). Therefore, we here denote “complexity” in the most neutral way possible, simply referring to it as a property of language arising from lexico-syntactic and semantic features of it. We also consider the notion of complexity as cognate – although not synonymic – with “comprehensibility”.

#### 3.1 Data and methodology

In order to compare the two languages in terms of complexity, a subset of the linguistic parameters provided by the online tool *Profiling-UD* (Brunato et al., 2020) are extracted. This text analysis tool allows for the extraction of a vast number of “linguistic profiling” features across different levels of linguistic annotation. Particularly interesting for the purposes of the present paper is the fact that Profiling-UD is specifically devised to support cross-linguistic analyses, since it is based on the Universal Dependencies (UD) representation (Nivre, 2015).

The comparative analysis was performed using the same corpora of legal and general written language used for previous works: the legal and fiction subcorpus of CORIS for Italian, and EurLEX and the BNC imaginative subcorpus for English (see section 3.1 above). The Italian corpus CORIS was chosen as it is the reference corpus for contemporary written Italian, and reflects a type of Italian variety that can be defined as ‘written-written’, following Nencioni (1983)’s criteria. It was accessed from the corpus web interface (<https://corpora.ficlit.unibo.it/TCORIS/>). The fiction subcorpus (CORIS\_narr) includes both novels and short stories. The narratives are further subdivided into adult, children, adventure, science-fiction and women literature. Total word count amounts to about 25M words. The legal subcorpus (CORIS\_law) is one of the most comprehensive datasets of Italian legal language, including strictly legal, bureaucratic, and administrative texts. It amounts to 10M words.

For English, two separate corpora were used: for legal language, an ad-hoc subcorpus of the *EurLEX* (Baisa et al., 2016) corpus was created, to include legislative documents in English ranging from the 1990s to 2015 (EUR), for comparability reasons; the resulting subcorpus amounts to 600M words. As a

proxy for general written language, the “imaginative” subcorpus of the BNC (BNC, 2007) was used, amounting to 20M words. Both corpora were accessed via the *SketchEngine* web interface.

As a first step, a random sample of concordances was extracted from the Italian and English subcorpus of CorIELLS and from the reference subcorpora. The concordances were selected containing the 4 constructions already used for previous analyses (see section 3.1). CQL and CQP searches for the 4 constructions were performed respectively on *SketchEngine* (Kilgarriff et al., 2004) and on the *CORIS* web interface. A total dataset of 120 concordances per language was extracted. The concordances are normalised for length, measuring between 100 and 200 characters, and averaging around 30 tokens per sentence (English, mean tokens per sentence: 28.67, SD: 2.5; Italian, mean tokens per sentence: 28.65, SD: 5.5).

The concordances were run through the online demo version of Profiling-UD<sup>2</sup>. A total of 14 parameters was selected from all levels of analysis provided by the tool: Raw Text Properties, Lexical Variety, Morphosyntactic information, Verbal Predicate Structure, Global and Local Parse Tree Structures, and Use of Subordination. Table 3 outlines the parameters extracted and their function. The explanation is taken directly from the “Linguistic Profile Legend” provided by Profiling-UD.

Table 3 - *Linguistic parameters extracted from Profiling-UD*

<i>Level of analysis</i>	<i>Linguistic parameters</i>	<i>Explanation</i>
Raw text properties	n_tokens	Total number of tokens
	char_per_tok	average number of characters per word (excluded punctuation)
Lexical variety	Ttr_lemma_chunks100	Type/Token Ratio (TTR) calculated with respect to the lemmata in first 100 tokens of a document. It ranges between 1 (high lexical variety) and 0 (low lexical variety).
	Ttr_form_chunks100	Type/Token Ratio (TTR) calculated with respect to the word forms in first 100 tokens of a document. It ranges between 1 (high lexical variety) and 0 (low lexical variety).
Morphosyntactic information	lexical_density	Ratio between content words (nouns, proper nouns, verbs, adjectives, adverbs) over the total number of words in a document

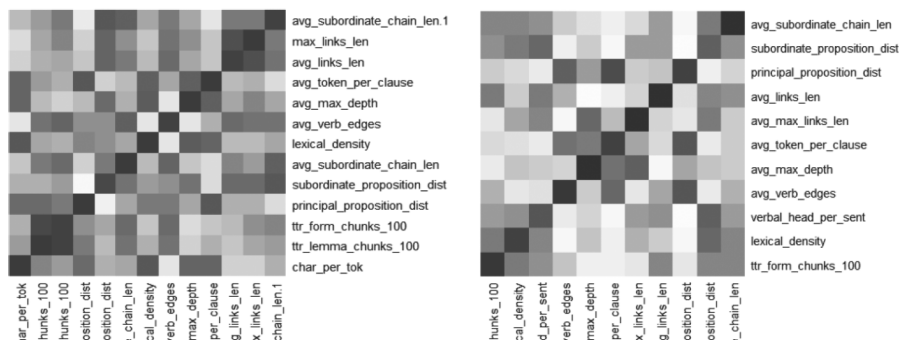
<sup>2</sup> The tool is available online at: <http://linguisticprofiling.italianlp.it>

<i>Level of analysis</i>	<i>Linguistic parameters</i>	<i>Explanation</i>
Verbal Predicate Structure	verbal_head_per_sent	average distribution of verbal heads in the document, out of the total of heads.
	avg_verb_edges	verbal arity, calculated as the average number of instantiated dependency links (covering both arguments and modifiers) sharing the same verbal head, excluding punctuation and auxiliaries bearing the syntactic role of copula according to the UD scheme
Global and Local Parse Tree Structures	avg_max_depth	mean of the maximum tree depths extracted from each sentence of a document. The maximum depth is calculated as the longest path (in terms of occurring dependency links) from the root of the dependency tree to some leaf.
	avg_token_per_clause	average clause length, calculated in terms of the average number of tokens per clause, where a clause is defined as the ratio between the number of tokens in a sentence and the number of either verbal or copular head.
	avg_max_links_len	mean of the longest dependency links extracted from each sentence of a document.
	avg_links_len	average number of words occurring linearly between each syntactic head and its dependent (excluding punctuation dependencies).
Use of Subordination	principal_proposition_dist	distribution of principal clauses
	subordinate_proposition_dist	distribution of subordinate clauses, as defined in the UD scheme: <a href="https://universaldependencies.org/u/overview/complex-syntax.html#subordination">https://universaldependencies.org/u/overview/complex-syntax.html#subordination</a> .
	avg_subordinate_chain_len	average length of subordinate chains, where a subordinate 'chain' is calculated as the number of subordinate clauses embedded on a first subordinate clause.

After extracting these measures for all the concordances, the dataset was analysed using Principal Component Analysis (henceforth: PCA). Although genre and register analysis generally employs Factor Analysis, it has been shown in the literature that PCA yields results comparable to Factor Analysis, especially if the variables are correlated among them and the number of variables is sufficiently large (Field et al. 2012: 760; Levshina, 2015). The PCA analysis was conducted using the statistical environment R and the package *factoextra* (Kassambara, Mundt, 2020). As figure 1 shows, variables appear to be overall moderately correlated to

each other (darker squares indicate higher correlation), justifying the use of PCA as a method of analysis.

Figure 1 - Heatmaps showing correlations among variables in English (red, left) and Italian (blue, right)



### 3.2 Results: PCA of linguistic complexity in Italian LLL

Since the variables extracted from Profiling-UD pertain to both lexical and morphosyntactic level of analyses, we perform two separate PCA analyses: one with more lexically oriented variables (number of tokens, characters per token, lexical density, and TTRs) and one for more syntactically oriented variables (the rest of the 14 variables in table 3).

The scree plots in figure 2 show the percentage of variance accounted for the dimensions for the two analyses. Following the so-called Kaiser criterion, we retain in our analysis only the components with eigenvalues higher than 1 (Levshina, 2015: 355). Therefore, only the first two dimensions are retained in the analysis of both lexical and syntactic complexity.

Manually inspecting the individual contributions of variables to each dimension (see figures 3 and 4 below) the dimensions of lexical complexity are labelled as (1) *Lexical Diversity*, and (2) *Lexical Density*. For syntactic complexity (figure 4), the dimensions are labelled as (1) *Dependency structure*, and (2) *Internal constituents' complexity*.

Figure 2 - Scree plot of the percentage of variance accounted for by the different dimensions of lexical and syntactical complexity

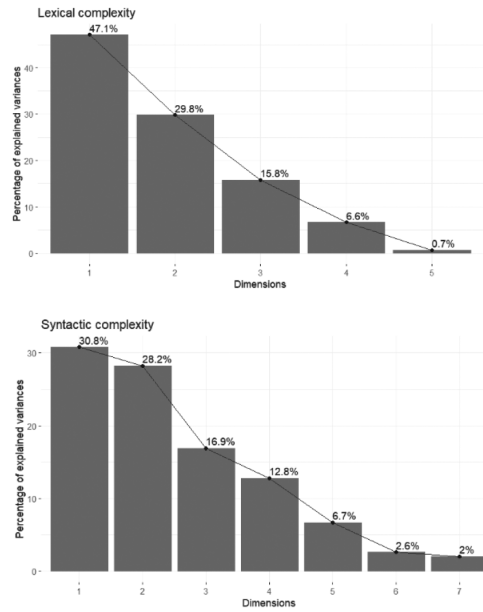


Figure 3 - Individual contributions of variables to the first 2 dimensions of lexical complexity

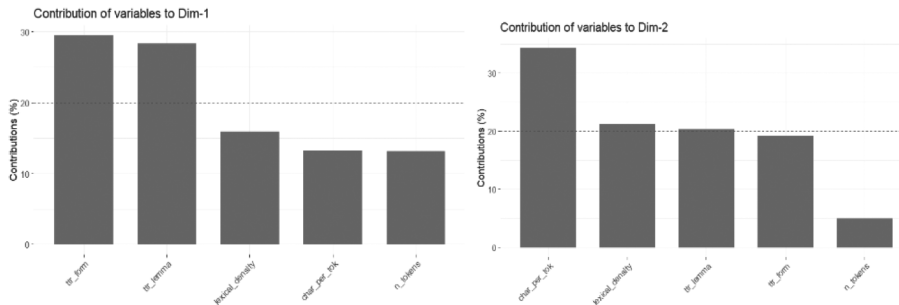
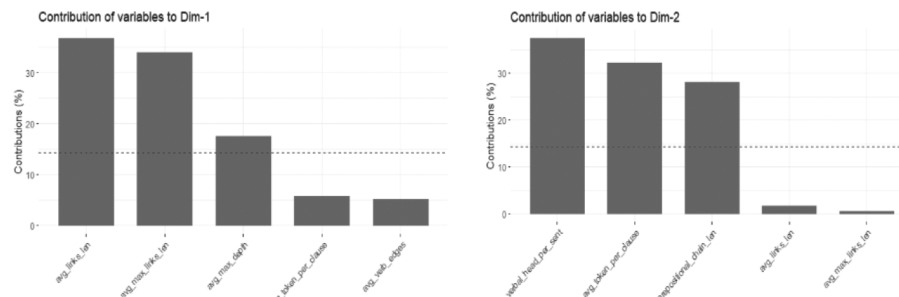


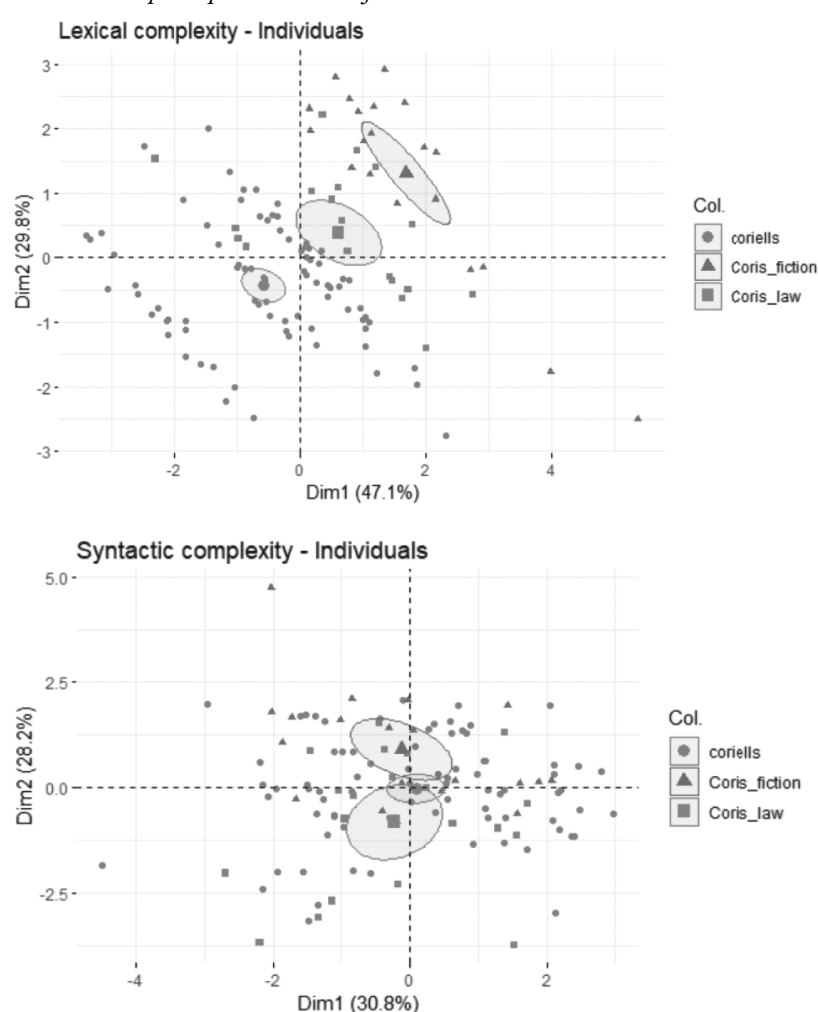
Figure 4 - Individual contributions of variables to the first 2 dimensions of syntactic complexity





After having identified the principal components of variance, the similarities and differences across the three corpora are explored. As it can be seen from figures 5A and 5B below, lexically it appears that LLL is pretty distant from both legal jargon and written prose. Syntactically – however – LLL appears to be a ‘perfect mix’ of the two neighbour genres. In other words, it appears that while for morphosyntactic dimensions of complexity LLL is a mix between specialistic and general-domain written language, the lexical aspect of it might be more idiosyncratic<sup>3</sup>.

Figure 5 - *First two dimensions of lexical and syntactic complexity for the three corpora. Ellipses represent 95% confidence intervals around the centroid*



<sup>3</sup> It must be noted that the sample of variables that can be extracted from Profiling-UD is unbalanced, with fewer lexical information and the majority of variables pertaining to syntactic profile. However, this tool was used as it is the only one – to the author’s best knowledge – that allows for cross-linguistic comparisons for its use of UD.

### 3.3 Results: PCA of linguistic complexity in English LLL

The same process applied for Italian has been applied to English. The scree plots in figure 6 shows that in this case the first 2 dimensions are considered for lexical complexity, and the first 3 for syntactic complexity. Very similarly to Italian, the dimensions of lexical complexity (see figure 7 below) are labelled as (1) *Lexical Diversity*, and (2) *Lexical Density*. For syntactic complexity (figure 8), the dimensions are labelled as (1) *Syntactic Structure*, (2) *Dependency Length*, and (3) *Internal Constituents' Complexity*.

Figure 6 - Scree plots of the percentage of variance accounted for by the different dimensions of lexical and syntactical complexity

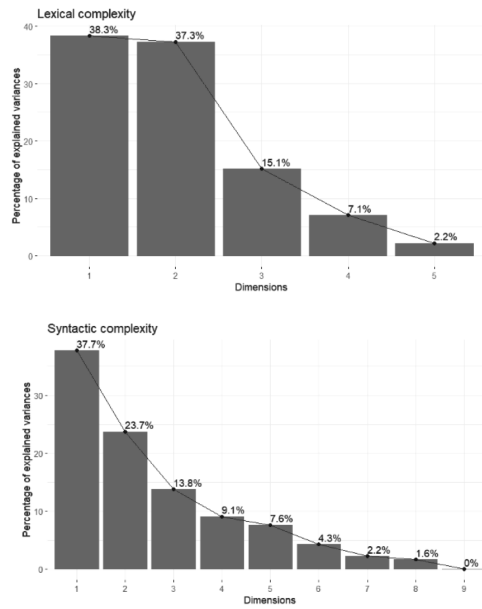


Figure 7 - Individual contributions of variables to the first 2 dimensions of lexical complexity

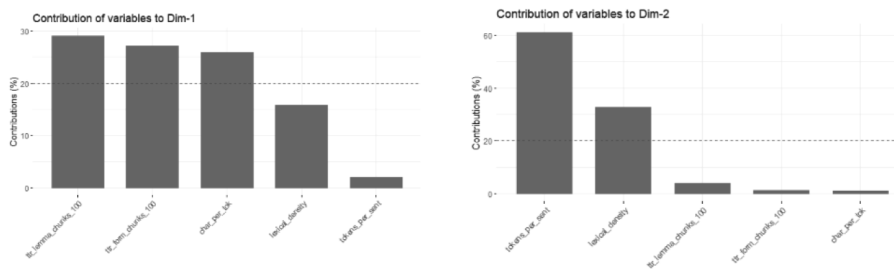
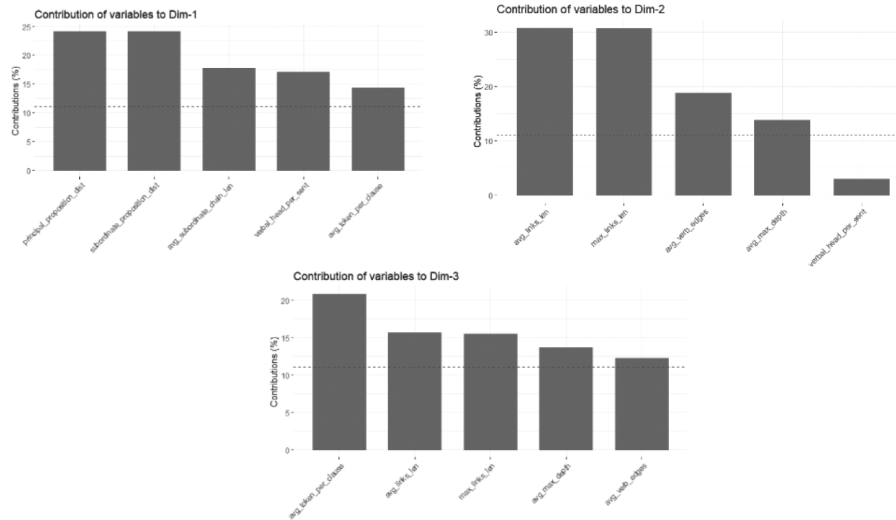


Figure 8 - Individual contributions of variables to the first 3 dimensions of syntactic complexity



From the distribution of the data across the dimensions of complexity, the situation is quite different from Italian. Lexically (figure 9), English LLL is extremely similar to legal jargon – differently from what was noted for Italian (see section 3.2.1 above). For the three dimensions of syntactic complexity on the other hand there is much more overlap across the three corpora. Particularly, what the graphical overlaps shows is that at the morpho-syntactical level LLL is still similar to legal jargon but with a great influence of general-domain language.

Figure 9 - First two dimensions of lexical complexity for the three corpora.  
Ellipses represent 95% confidence intervals around the centroid

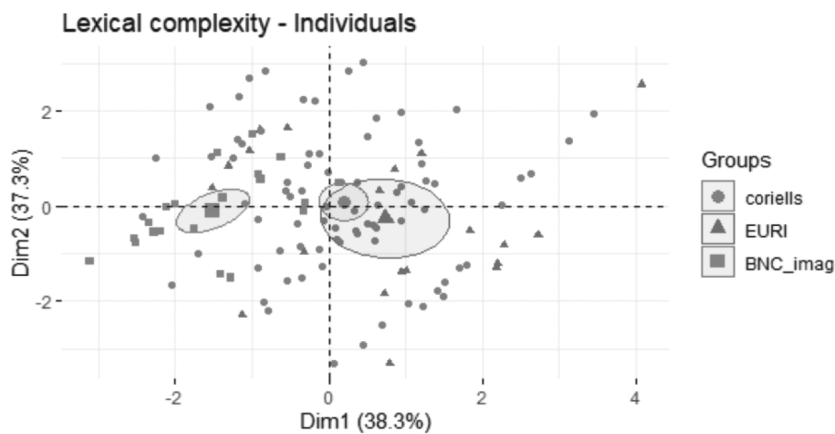
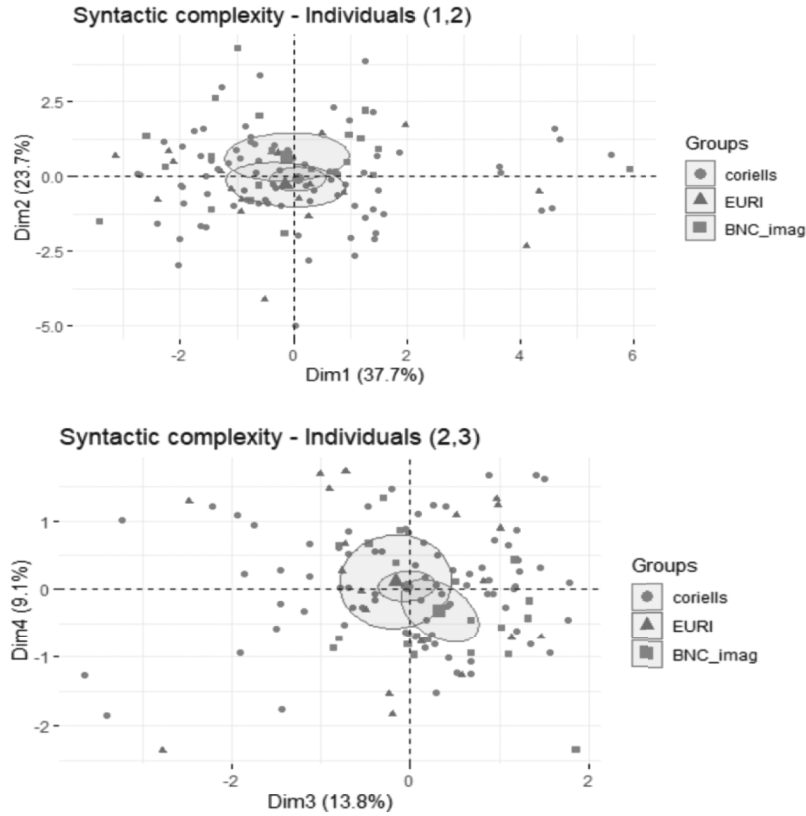


Figure 10 - First two dimensions of lexical complexity for the three corpora.  
*Ellipses represent 95% confidence intervals around the centroid*



#### 4. Discussion

The present paper presented CorIELLS, the first corpus of Italian and English legal-lay language. To show the potential of this corpus and in general of small, specialised corpora, a number of exploratory analyses have been presented. Previous research on CorIELLS mainly aimed at establishing legal-lay language as an idiosyncratic genre, or textual type, based on quantifiable linguistic properties. In this contribution, we expand this line of research by presenting a Principal Component analysis of lexical and syntactical complexity measures.

This type of analysis shows how legal-lay language is positioned across different dimensions of linguistic complexity compared to legal jargon and general-domain prose. Findings, albeit preliminary, seem to provide cross-linguistic support to our hypothesis of legal lay language being an independent genre. For Italian, we find that along the dimensions of lexical complexity legal-lay language is different from both legal jargon and written prose, while syntactically it appears to be a 'perfect mix' of the two neighbour genres, with vast overlaps. English instead is lexically very

similar to legal jargon, while the dimensions of syntactic complexity show much more overlap among the three genres.

To fully understand the differences in the two languages additional research is required, but it appears that English and Italian approach simplification in legal language from two different standpoint: while both English and Italian simplify the syntactic structure (as shown by the overlaps with general written language), at the lexical level English remains quite “conservative”, while Italian legal-lay lexicon – while being still closer to legal jargon than to general written language – shows more idiosyncratic characters. Overall, legal-lay language for both Italian and English is confirmed to have structurally autonomous characters, and to be a ‘blended’ genre made of a mixture of highly specialised features with more general features.

The research outlined here is by no means exhaustive, and possible other lines of research expanding it include a more in-depth look at processing difficulties of legal-lay texts (following for example Gunnarson 1984; Hotta, Fujita 2012; Conklin et al., 2019), especially for vulnerable groups such as L2 speakers or not completely scholarised adults.

### *Bibliography*

- ADLER, M. (2012). *The Plain Language Movement*. Oxford: Oxford University Press.
- BAISA, V., MICHELFEIT, J., MEDVEĎ, M., & JAKUBÍČEK, M. (2016). European union language resources in sketch engine. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 23-28 May 2016, Portorož, Slovenia. European Language Resources Association, 2799-2803.
- BARONI, M. & BERNARDINI, S. (2004). BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, May 26-28, 2004, Lisbon, Portugal. European Language Resources Association, 1-4.
- BENOIT, C. (1990). An intelligibility test using semantically unpredictable sentences: Towards the quantification of linguistic complexity. In *Speech Communication*, 9(4), 293-304.
- BENOLIEL, U. & BECHER, S.I. (2019). The Duty to Read the Unreadable. In *SSRN Electronic Journal*.
- BHATIA, V.K. (1983). Simplification v. Easification – The Case of Legal Texts. In *Applied Linguistics*, 4(1), 42–54.
- BHATIA, V.K. (2004). *Worlds of Written Discourse. A Genre-Based View*. London: Continuum International.
- BIBER, D. (1993). Representativeness in corpus design. In *Literary and linguistic computing*, 8(4), 243-257.
- BIEL, Ł. (2009). Corpus-based studies of legal language for translation purposes: Methodological and practical potential. In *Reconceptualizing LSP. Online proceedings of the XVII European LSP symposium*, 1- 15.
- BIEL, Ł. (2018). Corpora in Institutional Legal Translation: small steps and the big picture. In PRIETO RAMOS, F. (Ed.) *Institutional Translation for International Governance. Enhancing Quality in Multilingual Legal Communication*. London/New York: Bloomsbury, 25-36.

BNC Consortium (2007). *The British National Corpus, XML Edition*, Oxford Text Archive, <http://hdl.handle.net/20.500.12024/2554>.

BREZINA, V. & GABLASOVA, D. (2015). Is there a core general vocabulary? Introducing the new general service list. In *Applied Linguistics*, 36(1), 1-22.

BRUNATO, D. & VENTURI, G. (2014). Le tecnologie linguistico-computazionali nella misura della leggibilità di testi giuridici. In *Informatica e diritto*, XL (XXIII), 111-142.

BRUNATO D., CIMINO A., DELL'ORLETTA F., MONTEMAGNI S., & VENTURI, G. (2020). Profiling-UD: a Tool for Linguistic Profiling of Texts. In *Proceedings of 12th Edition of International Conference on Language Resources and Evaluation (LREC 2020)*, May 11-16, Marseille, France. European Language Resources Association, 7145-7151.

BUSSO, L. (2022a). Lexicon and Grammar in Legal-Lay Language: a Quantitative Corpus Study in Italian. *Studi Italiani di Linguistica Teorica e Applicata*, LI(1), 5-32.

BUSSO, L. (2022b) An investigation of the lexico-grammatical profile of English legal-lay language. *Journal of Language and Law*, 9(1), 146-184.

CHOVANEC, J. (2013). Grammar in the Law. In CHAPPELLE, C. (Ed.) *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell, 1- 8.

CONKLIN, K., HYDE, R., & PARENTE, F. (2019). Assessing plain and intelligible language in the Consumer Rights Act: a role for reading scores? In *Legal Studies*. 39(3), 378-397.

CRYSTAL, D. & DAVY, D. (1969). *Investigating English style*. London. Longman.

CURTOTTI, M. & MCCREATH, E. (2013). A right to access implies a right to know: An open online platform for research on the readability of law. In *Journal of Open Access to Law* 1(1), 1-56.

DE MAURO, T. & CHIARI, I. (2016). *Il Nuovo vocabolario di base della lingua italiana*, available at: <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.

EAGLESON, R. (2004). Plain Language.Gov – Improving communication from the Federal government to the public. Retrieved on January 10, 2022 at <http://www.plainlanguage.gov/whatisPL/definitions/eagleson.cfm>.

FAIRCLOUGH, N. (2011). *Critical discourse analysis – The critical study of language*. 2nd ed. Harlow: Longman.

FIELD, A., MILES, J., & FIELD, Z. (2012). *Discovering Statistics Using R*. Los Angeles: Sage.

FLOWERDEW, L. (2002). Corpus-based Analyses in EAP'. In FLOWERDEW, J. (Ed.) *Academic Discourse*. London: Pearson, 95-114.

FLOWERDEW, L. (2004) The Argument for Using English Specialized Corpora to Understand Academic and Professional Settings. In CONNOR, U. & UPTON, T. (Eds) *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 11-33.

FRADE, C. (2007). Power dynamics and legal English. In *World Englishes*, 26(1), 48-61.

FRADE, C. (2016). The power of legal conditionals in international contracts. In HAFNER, C.A., WAGNER, A., BHATIA, V.K. (eds.) *Transparency, Power, and Control*. London: Routledge, 45-64.

GARZONE, G. (2008). International commercial arbitration rules as translated/re-written texts: an intercultural perspective. In BHATIA, V.K., CANDLIN, C.N. & EVANGELISTI

ALLORI, P. (Eds.) *Language, Culture and the Law. The Formulation of Legal Concepts across Systems and Cultures*, Frankfurt am Main: Peter Lang, 47-73.

GIBBONS, J. (2003). *Forensic linguistics: An introduction to language in the justice system*. Oxford: Blackwell.

GOLDBERG, A.E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

GOLDBERG, A.E. (2019). *Explain Me This*. Princeton: Princeton University Press.

GUNNARSSON, B.-L. (1984). Functional comprehensibility of legislative texts: Experiments with a Swedish act of parliament. In *Text – Interdisciplinary Journal for the Study of Discourse*, 4(1–3), 71-106.

HOTTA, S. & FUJITA, M. (2012). The psycholinguistic basis of distinctiveness in trademark law. In TIERSMA, P.M., & SOLAN, L.M. (Eds.) *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press, 478-486.

HYLAND, K. (2006). *English for Academic Purposes: An advanced resource book*. London: Routledge.

KASSAMBARA, A. & MUNDT, F. (2020). *Factoextra Extract and Visualize the Results of Multivariate Data Analyses*. R Package Version 1.0.7. <https://CRAN.R-project.org/package=factoextra>

KILGARRIFF, A., P. RYCHLÝ, P. SMRZ, & D. TUGWELL. (2004). Itri-04-08 The Sketch Engine. In *Information Technology*, 105-116.

KIMBLE, J. (2000). The Great Myth That Plain Language Is Not Precise. In *Scribes Journal of Legal Writing*, 7, 109-118.

KOESTER, A. (2010). Building small specialised corpora. In O'KEEFFE, A., & MCCARTHY, M. (Eds.). *The Routledge handbook of corpus linguistics*. London: Routledge, 66-79.

KUSTERS, W. (2003). *Linguistic complexity: The influence of social change on verbal inflection*. Leiden: LOT.

KUSTERS, W. (2008). Complexity in linguistic theory, language learning and language change. In MIESTAMO, M., SINNEMÄKI, K. & KARLSSON, F. (Eds.) *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins, 3-22.

Interinstitutional Style Guide (2015), available at <https://publications.europa.eu/code/en/en-000100.htm> (last retrieved: 15 February 2023)

LEVSHINA, N. (2015). *How to do linguistics with R. Data Exploration and Statistical Analysis*. Amsterdam-Philadelphia: John Benjamins.

LINTAO, R.B. & MADRUNIO, M.R. (2015). Analyzing the lexical structures of a Philippine consumer-finance contract. In *Journal of Teaching English for Specific and Academic Purposes*, 2(3), 359-370.

MATULEWSKA, A. (2007). *Lingua legis in translation*. Frankfurt am Mein: Peter Lang.

MORI, L. (2019). Complessità sintattica e leggibilità. Un monitoraggio linguistico per la valutazione dell'accessibilità dei testi legislativi europei e italiani. In *Studi Italiani di Linguistica Teorica e Applicata*, 48, 627-657.

MORTARA GARAVELLI, B. (2001). *Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani*. Torino: Einaudi.

- NENCIONI, G. (1983). Parlato-parlato, parlato-scritto, parlato-recitato. In NENCIONI, G. (ed.), *Di scritto e di parlato. Discorsi linguistici*. Bologna: Zanichelli, 126-179
- NIVRE, J. (2015). Towards a universal grammar for natural language processing. In *Computational Linguistics and Intelligent Text Processing*, 3–16.
- PALLOTTI, G. (2015). A simple view of linguistic complexity. In *Second Language Research*, 31(1), 117-134.
- PETYKO, M., BUSSO, L., GRANT, T., & ATKINS, S. (2022). The Aston Forensic Linguistic Databank (FoLD). *Language and Law/Linguagem e Direito*, 9(1), 9-24. [https://doi.org/0.21747/21833745/lanlaw/9\\_1a1](https://doi.org/0.21747/21833745/lanlaw/9_1a1)
- ROSSINI-FAVRETTI, R., TAMBURINI, F. & DESANTIS, C. (2002). CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In WILSON, A., RAYSON, P. & MCENERY, T. (Eds.) *A rainbow of corpora: Corpus linguistics and the Languages of the world*. Munich: Lincom-Europa, 27-38.
- PETYKO, M., ATKINS, S., BUSSO, L. & GRANT, T. (2022). The Forensic Linguistic Databank. *Journal of Language and Law*, 9(1), 9-24.
- STEFANOWITSCH, A. (2013). Collostructional analysis. In HOFFMANN, T. & TROUSDALE, G. (Eds.) *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press.
- SWALES, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- TIERSMA, P.M. (1999). *Legal language*. Chicago: University of Chicago Press.
- TOGNINI BONELLI, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- VAN BOOM, W.H., DESMET, P., & VAN DAM, M. (2016). “If It’s Easy to Read, It’s Easy to Claim” – The Effect of the Readability of Insurance Contracts on Consumer Expectations and Conflict Behaviour. In *Journal of Consumer Policy*, 39(2), 187–197.
- VENTURI, G. (2011). Semantic annotation of Italian legal texts: a FrameNet-based approach. In *Constructions and Frames* 3(1), 46-79.
- WILLIAMS, C. (2004). Legal English and Plain Language: an introduction, In *ESP Across Cultures*, 1, 111-124.
- WILLIAMS, C. (2004b). Pragmatic and cross-cultural considerations in translating verbal constructions in prescriptive legal texts in English and Italian. In *TEXTUS XVII/1*, 217- 246.
- WILLIAMS, C. (2005). Progetto Chiaro! and the Plain Language Movement in Italy. In *Clarity* 53, 30-32.



FRANCESCO SIGONA, SONIA D'APOLITO, COSIMO IAIA,  
BARBARA GILI FIVELA, MIRKO GRIMALDI

## Forensic Automatic Speaker Recognition with dialectal speakers: a pilot study on the Taranto and Brindisi varieties

In the Forensic Speaker Recognition, the choice of the reference linguistic population plays a key role in evaluating the typicality of the voice recordings to be compared, and therefore the strength of the evidence, within the Likelihood Ratio framework. In the present study we carry out multiple comparison tests among voice recordings of the dialectal speech of Taranto and Brindisi varieties (Puglia, Italy), through an Automatic Speaker Recognition system, using two reference populations (databases): one of “standard” Italian, by professional speakers, and one of dialectal speakers of the same variety as the voices to be compared. The aim is to observe whether the accuracy of the recognition system improves if the dialect reference population is used instead of the Italian spoken one.

*Keywords:* forensic voice comparison, likelihood ratio, reference population, spoken dialect, accuracy.

### 1. Introduction

#### 1.1 The forensic voice comparison

In the Forensic Voice Comparison (FVC), one or more audio recordings of the voice of a known speaker (hence, known samples) are compared to an audio recording of the voice of a speaker of questioned identity (hence, questioned samples)<sup>1</sup>: the goal is to understand to what extent the two samples of voices can probably be attributed or not to the same person.

To this purpose, a Bayesian approach has been widely established, and recently recommended by European Network of Forensic Science Institutes (Drygajlo et al., 2016). Accordingly, the task of the forensic scientist is to provide the court with a strength-of-evidence statement in answer to the question: “How much more likely are the observed differences between the known and questioned samples to occur under the hypothesis that the questioned sample has the same origin as the known sample than under the hypothesis that it has a different origin?” (Morrison, 2009). The answer to this question is quantitatively expressed as a Likelihood Ratio (LR): the LR represents the relationship between similarity and typicality of the compared voice samples and quantifies how similar the characteristics of the

---

<sup>1</sup> Also, Forensic Speaker Recognition (FSR) is frequently used.

recorded signal of the two voice samples are to each other, relative to the diffusion of the same characteristics in a reference linguistic population. The reference linguistic population, therefore, must be defined with particular attention and must be as homogeneous as possible (gender, age, language of the speakers, etc.) with respect to the characteristics of the recorded samples to be compared (Rose, 2002, 2005; Jessen, 2008). Failure to implement an adequate reference population is one of the main factors that makes the estimate of the LR, as well as the strength of the evidence, inaccurate (Robertson, Vignaux, 1995; Rose, 2006).

A crucial issue for FVC is represented by dialectal variation (and also micro-variation)<sup>2</sup>, not yet fully addressed within this field of research. At the moment, the databases of speech populations used in semi- and full-automatic systems refer to idealized speakers: that is, speakers broadly identified in respect of a national language (English speakers, Italian speakers, German speakers, etc.). Conversely, we know very well that national languages are abstract (administrative) entities and that speakers normally use regional varieties (more or less markedly) in their real life: this is particularly true for the Italian linguistic area. Furthermore, the varieties used by speakers are, in many cases, characterized by different phonological systems and different suprasegmental patterns. Also, phonological systems may drastically vary within a limited linguistic space, showing puzzling micro-variation; and each phonological system may show systematic phonological processes that change phonemes at the phonetic surface. So, what we today compare in forensic practice are, at best, common (general) features eventually shared by linguistic systems.

We do not know what biases this fact introduces within the forensic approaches (for instance, the Bayesian approach). Finally, this issue is inherently linked to socio-phonetics aspects: in fact, speaker's features may vary in respect of sociolinguistics variables (age, sex, literacy, contexts of use, etc.). The semi-automatic approach, indeed, allows an expert to select general properties of the vowels analyzed, avoiding that macro- and micro-variation have a drastic impact on the LR. On the contrary, this is impossible when an automatic approach is used.

For what concerns the Italian domain, previous contributions have highlighted the necessity to appropriately develop speech databases for FVC taking care of dialectal variation (cf. Romito et al., 2009; Romito, Galatà, 2008). An example is given by the *Primula corpus*, which contains over 900 recordings of 4 Calabrian speakers. It is characterized by three types of recording channels: high fidelity, environmental and telephone recordings. The recordings have been captured under different conditions that determine their quality: silent room, tapping in and out of a car, calls effected in the car, in the street, and in the classroom (Romito, Galatà, 2008). Unfortunately, this corpus is not, at the moment, available for research scopes.

---

<sup>2</sup> With the term 'micro-variation' we refer to the fact that dialects may often manifest subtle and irregular variations in respect of a general phonological or morpho-syntactic phenomena.

## 1.2 The aim of the present work

Along this line of research, we collected a dialectal database from speakers of two different varieties spoken in Southern Apulia (Italy). In this pilot study, we aim to investigate to what extent the use of a dialectal population, compared with a standard Italian speech population (produced by professional speakers), can influence the performance of state-of-the-art Forensic Automatic Speaker Recognition (FASR) systems. We assume that the accuracy of the comparison can improve when dialectal variation is taken under consideration.

This pilot study focuses only on male speakers for the following reasons. Firstly, it is well known that the production of speech sounds from male and female speakers is strongly dependent on biological differences, such as inner dimension of the mouth, throat, and vocal folds (Simpson, 2009; Hillenbrand & Clark, 2009). These important differences therefore justify a differentiated experimentation between the two genders, using two distinct reference population samples. On the other hand, collecting and analyzing audio samples (and especially dialect audio samples) is time-consuming. So, having a limited amount of time, we decided to focus our efforts on getting a large enough sample of one gender, rather than risk getting two too few samples for both genders. In the near future, however, it will always be possible to repeat the study with a sample of the other gender, integrating the results also in a comparative perspective between the two genders. The choice to begin with the male gender, instead of female, was essentially random.

Furthermore, the study has been carried out with audio samples at the classic telephonic audio quality, as this kind of signal happens very frequently in the forensic field (as in the case of wiretapping).

## 2. Methods

### 2.1 Datasets

#### 2.1.1 The Italian dataset

A dataset made of 150 Italian male professional speakers has been collected by recordings of freely available audio samples (<https://www.audible.it>). The characteristics of the collected samples are summarized in Table 1.

Table 1 - *Characteristics of the recorded audio samples for the Italian spoken dataset*

Average duration	3 minutes
Number of samples	150
Number of samples per speaker	1
Speaker gender	Male 100% (Female 0%)
Audio encoding (format, sampling rate, bit depth)	Mono, wav, 44100 Hz, 16 bit
Background sounds	no
Signal-to-Noise Ratio (SNR)	> 25 dB

In short, the voices of this dataset were those of professional speakers reading audiobooks (in the same case the reader impersonates two characters conversing on specific arguments). The average age of the speakers could not be determined.

Each audio recording was band-pass filtered between 300 and 3400 Hz and downsampled to 8000 Hz, using the Praat software (Boersma, Weenink, 2020), when necessary, after the sample collection.

### 2.1.2 The dialect datasets

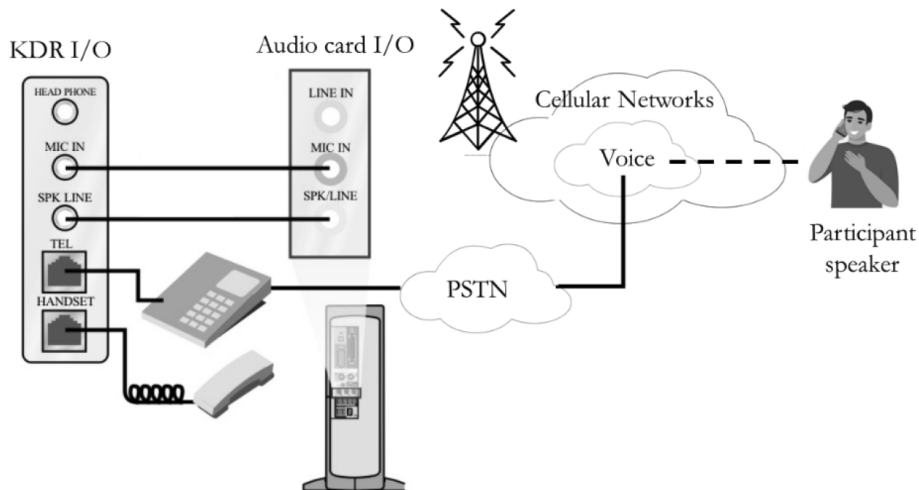
Speakers from Taranto and Brindisi areas were interviewed in order to collect dialectal data. Specifically, during a semi-guided telephonic interview, they were asked to translate, as spontaneously as possible, into dialect some Italian target phrases proposed by the interviewer. The speakers were called to their own mobile phone from a landline connection using a dedicated hardware (KDR, by Sistel s.r.l.) connected to a Microsoft® Windows®-based Personal Computer running the software Audacity<sup>3</sup> which was used to capture and record the audio signal. The whole setup is depicted in Figure 1.

And so, we collected dialectal data as follows:

- 22 male speakers (mean age 44) for the Taranto area: 14 speakers from Taranto city, 4 from Grottaglie, 2 from Monteiasi, 1 from Fragagnano, and 1 from Carosino;
- 22 speakers for the Brindisi area (mean age 40): all speakers come from Francavilla Fontana (see Figure 2).

All the audio samples were resampled to 8000 Hz, and have SNR > 20 dB.

Figure 1 - *Diagram of the setup for the dialect varieties speech samples collection*



<sup>3</sup> Audacity® software is copyright © 1999-2021 Audacity Team. Web site: <https://audacityteam.org/>. It is free software distributed under the terms of the GNU General Public License. The name Audacity® is a registered trademark.

### 2.1.3 Dialectal properties of the Taranto and Brindisi varieties

The Taranto varieties shows clear features of the Apulian dialects: i.e. diphthongization of stressed vowels within open syllables ([a'tʃejtə] *vinegar*; [ˈpaɪlə] *hair*), weakening of the final unstressed vowels ([ˈtandə] *many*; [kapeɟːə] *hair*), reduction of the metaphonic diphthongs [je], [we] to [i], [u] ([ˈpjedə] > [ˈpidə], *foot*, [ˈmwertə] > [ˈmurtə]), etc. (Mancarella, 1998).

On the other hand, the Francavilla Fontana variety, although characterized by the Apulian features, preserves many Salentino features, as the metaphony affecting the mid-high vowels [e], [o] developed by the Latin vowels ĭ, ē and ŭ, ū. The outcome is the raising of [e], [o] to the high vowels [i], [u]: [la ˈpera] *the pear* Sg F / [lu ˈpiru] *the pear* Sg M; [lu ku ˈlore] *the color* Sg / [li ku ˈluri] *the colors* Pl, etc. Also, the vowels [ɛ], [ɔ] derived by the Latin vowels ĕ, ō may be affected by metaphonic diphthongization: i.e., [lu ˈpeti] *the foot* Sg. / [li pjetɪ] *the feet* Pl; [la ˈnotti] *the night* Sg. / [li ˈnwetti] *the nights* Pl. (Ribezzo, 1912).

Figure 2 - Maps of the investigated dialect areas

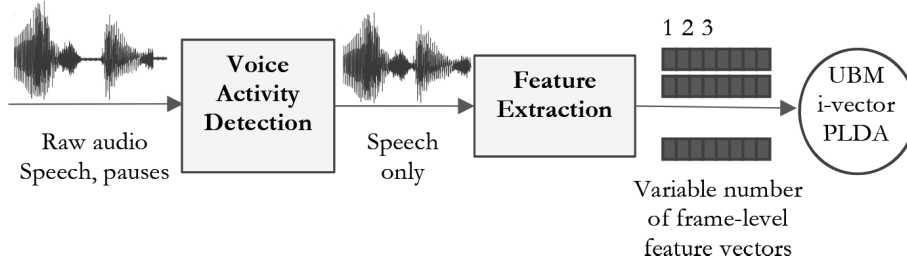


It is interesting to note that for what concerns the Taranto area within the Monteiasi, Carosino, and Fragagnano varieties, Apulian features coexist with Salentino features (Ribezzo, 1912; Mancarella, 1998). Thus, this fact introduces further dialectal variation within the speech population, which is at the core of our aim (cf. Section 1.2).

## 2.2 Front-end audio processing

The comparison algorithm that has been chosen in the current study requires a preliminary conversion of the recorded audio samples into a sequence of vectors of numbers, which give an alternative representation of the information contained in the waveform of the voice signal. These “feature vectors” are the output of a front-end audio processing stage, that for the current study has been simplified into a chain of only two processing blocks: the Voice Activity Detection (VAD) and the real Feature Extractor (see Figure 3).

Figure 3 - Block diagram of the front-end audio processing, which transforms the input audio file into a variable number of feature vectors required by the modelling block (the UBM/i-vector/PLDA)



### 2.2.1 Voice Activity Detection

Voice Activity Detection (VAD), often referred to as Speech Activity Detection (SAD) or simply Speech Detection, is the task of locating speech segments within an audio recording. VAD plays a key role in any speech processing system, including speaker recognition applications where, at least for those approaches based on short-term spectral features, it is required to prevent unnecessary processing of non-speech segments. To this purpose, many methods have been proposed. Classic digital signal processing methods usually classify voiced/unvoiced frames based on scalar features such as short-term energy, zero-crossing rate (Benyassine et al., 1997), periodicity (Tucker, 1992) or spectral divergence (Ramirez et al., 2004); these methods are quite simple and effective on clean condition, but the classification accuracy tends to suffer on low SNR. Statistical model-based approaches have been explored (Sohn et al., 1999; Shin et al., 2010), assuming that the spectral coefficients follow a particular parametric distribution, where the VAD decision is sought by calculating the likelihood ratio based on the hypothesized models. The statistical methods often outperform the classic methods in the presence of stationary noise, but non-stationary noise conditions remain challenging. Supervised models have also been studied, using machine learning techniques and leveraging prior knowledge in large, annotated audio collections (Ng et al., 2016; Plchot et al., 2016; Wu, Zhang, 2011; Zhang, Wu, 2013; Thomas et al., 2015). Such VAD approaches tend to be sensitive to acoustic mismatch between the training and test. Adaptive supervised VADs, such as Huijbregts et al. (2007) and Kinnunen and Rajan (2013) have also been considered, that represent a compromise between the powerful supervised approaches, such as neural networks, and statistical model-based methods which require no prior training but whose parametric modelling assumptions might be over-simplistic.

In this work we decided to use a VAD technique which was adequate to the quality of the available speech recordings and for which a software implementation was already available. We found both requirements met in VoiceBox<sup>4</sup>, a freely

<sup>4</sup> VOICEBOX: Speech Processing Toolbox for MATLAB. Web site: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>

available toolbox for Matlab®, that comes with an implementation of Sohn et al.'s (1999) approach.

### 2.2.2 The Feature Extractor

The output of the VAD is an audio waveform that is input to the real Feature Extractor algorithm. Even in this case, we decided to choose baseline features, such as the Mel-Frequency Cepstral Coefficients (MFCCs). MFCCs are short-time spectral features, meaning that the speech utterance is first divided into (usually overlapped) fragments (frames), that are small enough to be able to assume that in each one the characteristics of speech are time-invariant, then a MFCCs vector is computed for each frame. MFCCs were introduced in the early 1980s by Davis and Mermelstein for speech recognition, and then adopted in many studies on speaker recognition. MFCCs are computed with the aid of a psychoacoustically motivated filter bank, followed by logarithmic compression and discrete cosine transform (DCT). The step required to compute MFCCs are:

- pre-emphasis: this step refers to a filtering operation that emphasizes the speech signal at higher frequencies and is considered in many speech processing applications.
- framing: this operation has been described above. The frame length is usually fixed, but pitch-synchronous analysis has also been (Nakasone et al., 2004; Zilca et al., 2006; Gong et al., 2008) and is still studied (Chen and Miller, 2020).
- windowing: this operation aims at tapering the signal to zero at the beginning and end of each frame, to deal with the finite-length effect of the Discrete Fourier Transform (DFT). The Hamming window is commonly used but the choice of the window function is not considered critical (Kinnunen, Li, 2010).
- DFT: the very well-known technique. It is computed frame by frame to get the spectral amplitude of the signal inside each frame.
- Mel filter bank processing: the output of the DFT is multiplied by a bank of filters to achieve the so-called mel-spectrum. A mel is a unit of measure based on the human perception of tones. The human auditory system apparently does not perceive pitch linearly, so the mel does not correspond to the value of the physical frequency.
- Logarithmic compression: since voiced sounds can be modelled by a source signal filtered by the resonance cavity of the vocal tract, which is a multiplication in the frequency domain, applying logarithm operation allows to get multiplied factors of the spectrum (in this case, the mel-spectrum) into additive ones. This results in a signal in the cepstral domain with a quefrequency peak corresponding to the pitch of the signal and several formants representing low quefrequency peaks.
- Discrete Cosine Transform (DCT): since the vocal tract is smooth, the energy levels in adjacent bands tend to be correlated. Also, the filters in the filter-banks are overlapped, so the energy from ones next to each other is being spread between two. DCT applied to the transformed mel frequency coefficients produces a set of cepstral coefficients.

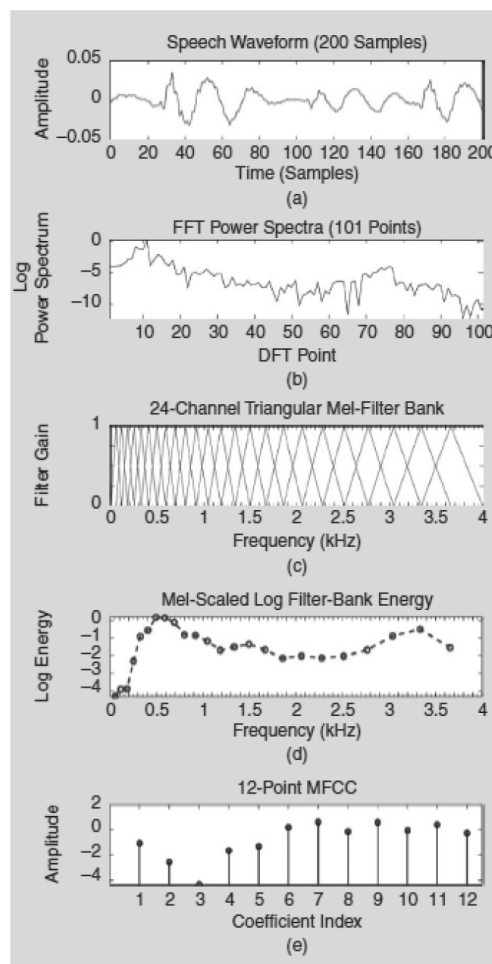
Figure 4 summarizes the MFCCs computation processing, while further details can be found in any speech processing manual.

The specification of the MFCCs extraction used in this study are:

- Frame duration: 16 ms (i.e., 128 samples @ 8000 Hz);
- Frame overlap: 50% (i.e., 64 samples)
- 14 static coefficients, including 0-order and log energy
- 28 dynamics coefficients (delta, delta-delta), corresponding to the first- and second-order derivatives of the static coefficients.
- $F = \text{Total number of MFCCs} = 42$ .

The actual computation has been performed using the VoiceBox toolbox.

Figure 4 - Steps in MFCC feature extraction from a speech frame:  
 (a) 200-sample frame representing 25 milliseconds of speech sampled at a rate of 8 kHz,  
 (b) DFT power spectrum showing first 101 points, (c) 24-channel triangular Mel-filter bank,  
 (d) log filter-bank energy outputs from Mel-filter, and (e) 12 static MFCCs obtained  
 by performing DCT on filter-bank energy coefficients and retaining the first 12 values  
 (adapted from Hasan, Hansen, 2015)





### 2.3 Automatic Speaker Recognition Algorithm

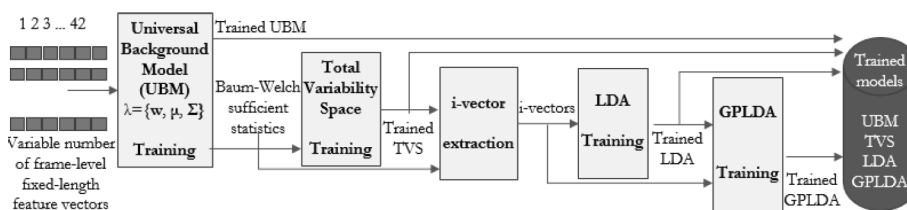
The chosen algorithm for the voice comparison is based on the UBM/i-vector/PLDA paradigm, which is a well-known state-of-art approach. In short:

- the Universal Background Model (UBM; Reynolds, 1997) is a probability density function in the form of a Gaussian Mixture Model (GMM) with  $K$  components, which is trained with all the feature vectors coming from all the speakers in the reference population. The output of the training is the set of UBM parameters (one estimated mean vector and covariance matrix for each Gaussian component in the mixture) as well as Baum-Welch sufficient statistics, that will play a role in the next blocks of the chain.
- i-vectors refers to a particular vector representation of the single utterance on a vector space called “Total Variability Space” (TVS, Dehak et al., 2011), which is particularly convenient for the purpose of voice comparison. The basic idea is that every single utterance of a generic speaker contains information that is a combination of a bias due to the target population, plus speaker-dependent and speaker-independent information. Speaker independent information depends on external factors such as the quality of the recording device, the amount and type of noise superimposed on the voice, etc., and is usually referred to as “channel / session dependent”. The i-vector approach has proven to be a viable and effective way to represent speaker-dependent and independent information, without any bias due to the target population, using far fewer elements than the amount needed to work directly with Gaussian Mixed Models.
- GPLDA (Gaussian Probabilistic Linear Discriminant Analysis; Kenny 2010; Garcia-Romero and Espy-Wilson, 2011) is a back-end stage, that extracts the speaker-dependent information from the i-vectors (possibly after a further dimensionality reduction by Linear Discriminant Analysis), providing at the same time the framework to compare two i-vectors in terms of a Likelihood Ratio-based score.

Figure 5 depicts the processing steps that occur to train all the mentioned models, while Figure 6 depicts the processing steps that occur in the computation of the models of the reference speaker (the known speaker in a real FVC case) and the test speaker (the questioned sample, in a real FCV case), up to the scoring stage.

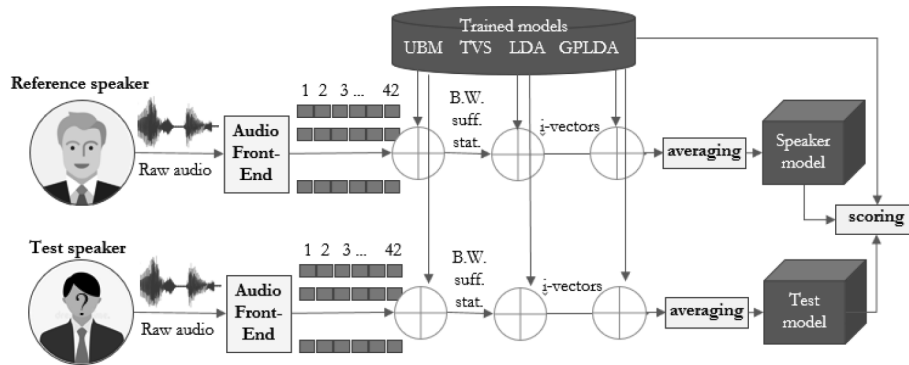
The computation of the models has been done by means of the Microsoft® MSR Identity Toolbox<sup>5</sup> for Matlab®.

Figure 5 - Train of the models of reference population required by the UBM/i-vector/GPLDA paradigm



<sup>5</sup> <https://www.microsoft.com/en-us/download/details.aspx?id=52279>

Figure 6 - Processing steps to compute the reference and the test speaker models, up to the scoring



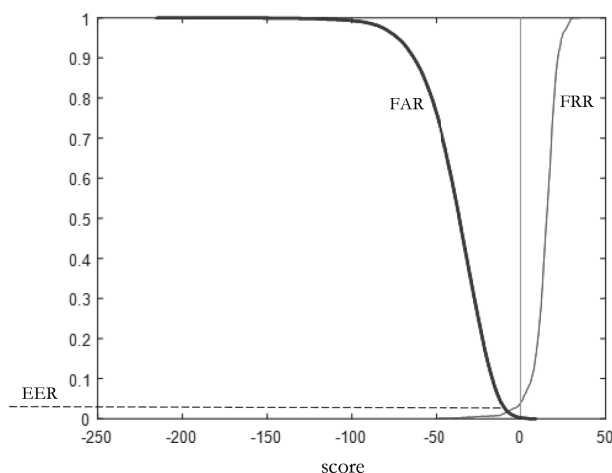
#### 2.4 Performance metric

The main global metrics to assess the accuracy performance of a speaker recognition system are the estimated probabilities that the system response supports contrary-to-fact hypothesis. When the system's response score supports the identification hypothesis of two utterances spoken by two different speakers, a False Acceptance (or False Identification) error occurs; when the system's response score supports the rejection (i.e. no identification) hypothesis of two utterances spoken by the same person, a False Reject error occurs.

By collecting the scores coming from a multitude of comparison tests between couples of utterances spoken by different speakers, it is possible to build a distribution of the scores and the distribution of the False Acceptance error Rate (FAR) as a function of the score, in this different-origin hypothesis. Similarly, by collecting the scores coming from a multitude of comparison tests between couples of utterances spoken by the same speaker, it is possible to build a distribution of the scores, and the distribution of the False Rejection error Rate (FRR), in this same-origin hypothesis. FAR and FRR computed in this way represent an estimate of the respective a priori error probabilities, while the a posteriori estimates depends on the outcome of the real forensic casework.

A very common global metric used to assess the accuracy of the system is then the Equal Error Rate (EER), which is the point where the FAR equals the FRR. The lower the EER, the higher the accuracy of the FASR system. It can be visualized using the Tippet plot as depicted in Figure 7.

Figure 7 - Example of Tippet plot, showing the FRR (red line) and the FAR (blue line) as functions of the score). The intersection point gives the value of the Equal Error Rate (EER)



## 2.5 Design of matched comparison tests

To assess the FASR performances in terms of EER, multiple test comparisons must be performed. In the case of reference population matched with the dialectal variety of the speaker, we proceeded as follows. Firstly, we divided the MFCCs matrix of each speaker in two equally sized matrices, in order to simulate two different utterances (sessions) for each speaker. Then, assuming that  $N$  is the number of dialect speakers in the dataset, to assess the distribution of the FAR we adopted a one-leave-out cross-validation, in which we compared the two sessions of each speaker, using the remaining  $N-1$  speakers of the dialectal dataset to build the models of the reference populations. By this way, there is no chance that such models were biased due to any information about the test speaker. To assess the distribution of the FRR, we adopted a similar strategy, in which we compared the first session of each speaker with the second session of any other speaker, using the remaining  $N-2$  speakers to build the models of the reference populations. Even in this way, there is no chance that such models were biased due to any information about the test speakers. Finally, the EER is computed as the score where  $FAR=FRR$ . We repeated such a cross validation procedure 100 times, in order to average with respect to some random initializations of the reference population models, occurring during their computation.

## 2.6 Design of mismatched comparison tests

The mismatched comparison tests are about the comparison of dialectal speaker using the Italian reference models. Since the Italian dataset is much larger than each dialectal dataset, we assumed that was not methodologically correct to build Italian reference models using the whole Italian dataset. Therefore, before to perform the comparison test, we have randomly split the whole Italian dataset into  $D$  smaller ones, having the same size ( $N$ ) of the dialectal datasets. For each of the  $D$  sub-dataset,

we computed the reference population models and performed the comparison test between the two sessions of a same dialectal speaker (to assess FAR) a between the first session of each dialectal speaker with the second session of any other dialectal speaker. Moreover, we repeated the whole procedure (including the random composition of the D Italian sub-datasets) 100 times.

### 3. Results

The results achieved in terms of EER in the mismatched and matched conditions are summarized in Table 2 for Taranto speakers, and Table 3 for Brindisi/Francavilla speakers.

Table 2 - *Descriptive statistics of the set of EER values coming from the multiple comparison tests, for speakers belonging to the Taranto area, using an Italian reference population, then the matched dialectal one. Coefficient of variation % is defined as  $100 \times \text{standard deviation} / \text{mean}$*

	<i>Mismatched condition (Italian refer. popul.)</i>	<i>Matched condition (dialectal refer. popul.)</i>
Number of trials	100	100
EER mean	7.9 %	3.0 %
EER standard deviation	1.3 %	0.43 %
EER coefficient of variation %	16.4 %	14.1 %
EER minimum	5.2 %	2.2 %
EER maximum	12.7 %	4.2 %
ERR range	7.5 %	2.0 %

Table 3 - *Descriptive statistics of the set of EER values coming from the multiple comparison tests, for speakers belonging to the Brindisi area, using an Italian reference population, then the matched dialectal one. Coefficient of variation % is defined as  $100 \times \text{standard deviation} / \text{mean}$*

	<i>Mismatched condition (Italian refer. popul.)</i>	<i>Matched condition (dialectal refer. popul.)</i>
Number of trials	100	100
EER mean	10.0 %	6.1 %
EER standard deviation	1.35 %	0.74 %
EER coefficient of variation %	13.5 %	12.0 %
EER minimum	7.8 %	4.5 %
EER maximum	14.0 %	8.3 %
ERR range	6.2 %	3.8 %

#### 4. Discussion

For both the dialectal varieties, we achieved a reduction in every EER-based metric in the matched (dialectal speakers and dialectal reference population) over the mismatched (dialectal speakers and “standard” Italian reference population) condition.

As for the Taranto dialectal area, the mean EER shows that it decreases from 7.9% for the mismatched condition to 3.0% for the matched one. Further, even the standard deviation improves in the matched condition, as it lowers from 1.3% to 0.43%, as well as the coefficient of variation even if quite slightly.

These results are also confirmed considering the Francavilla dialectal data. Indeed, the mean EER improves from the mismatched condition (10%) to the matched condition (6.1%). The standard deviation also improves as it lowers from 1.35% to 0.74% in the matched condition and the coefficient of variation slightly improves too.

The achieved results clearly support our initial research hypothesis that in caseworks where dialectal speakers are involved, FVC systems can provide more accurate results when a reference population of the same dialectal variety is used, instead of a generic one.

However, it should be noted that further studies would be needed to consolidate these results. In fact, the present study, which is only in the preliminary phase, has only two areas in consideration, while it would be appropriate for the study to be repeated for other areas also outside Puglia. Furthermore, more robust results would be obtained by sampling each area more numerously, doubling the number of participants per area, and including an equal number of speakers of the other gender.

#### 5. Conclusions and future directions

In the presented study, we have investigated to what extent the use of a dialectal population, compared with a standard Italian speech population, can influence the performance of state-of-the-art Likelihood Ratio-based FASR systems, also assuming that the accuracy of the comparison can improve when dialectal variation is taken under consideration in the reference population.

Although the results obtained in terms of improvement of the EER values are not yet generalizable, due to the small number of dialectal varieties in question, and the limited number of speakers sampled in each area, they clearly support at least the validity of the research hypotheses and establish an excellent starting point for future experiments.

Indeed, we plan to improve the above-mentioned limitations, by increasing the number of investigated areas, the number of speakers for each area, and including female speakers. Also, we plan to test different paradigms for FASR, like recent developments in the field of Deep Learning.

From a different point of view, the achieved results invite to carefully reflect on the importance and need to collect databases of dialectal language, for research purposes, not only at the level of academic institutions, but also and above all in collaboration with the relevant institutions and law enforcement agencies that daily carry out

investigations on the basis of intercepted conversations. By sharing under appropriate non-disclosure agreements, the certainly huge amount of phonic materials collected over years of investigations, it could give a significant boost to research in the field of the study of dialectal variations and their impact in the field of forensic phonetic.

### *Bibliography*

- BENYASSINE, A., SCHLOMOT, E. & SU, H. (1997). ITU-T recommendation G729 Annex B: a silence compression scheme for use with G729 optimized for v.70 digital simultaneous voice and data applications. In *IEEE Commun. Mag.*, 35, 64–73.
- BOERSMA, P. & WEENINK D. (2020). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.05. <http://www.praat.org/>
- CHEN, J.C., MILLER, D.A. (2020). Pitch-Synchronous Analysis of Human Voice. In *Journal of Voice*, 34(4), 494-502.
- DAVIS, S.B. & MERMELSTEIN, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition. In *IEEE Trans. on Acoustics, Speech and Signal Processing*, 28(4), 357–366.
- DEHAK, N., KENNY, P., DEHAK, R., DUMOUCHEL, P. & OUELLET, P. (2011). Front-end factor analysis for speaker verification. In *IEEE TASLP*, 19, 788-798.
- 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/>
- GARCIA-ROMERO, G. & ESPY-WILSON, C.Y. (2011). Analysis of i-vector length normalization in speaker recognition systems. In *Proc. Interspeech*, 249–252.
- GONG, W.G., YANG, L.P. & CHEN, D. (2008). Pitch synchronous based feature extraction for noise-robust speaker verification. In *Proc. Image and Signal Processing*, 5, 295–298.
- HANSEN, J.H., & HASAN, T. (2015). Speaker recognition by machines and humans: A tutorial review. In *IEEE Signal processing magazine*, 32(6), 74-99.
- HILLENBRAND, J.M., & CLARK, M.J. (2009). The role of  $f_0$  and formant frequencies in distinguishing the voices of men and women. In *Attention, Perception, & Psychophysics*, 71(5), 1150–1166.
- HUIJBREGTS, M., WOOTERS, C., ORDELMAN, R. (2007). Filtering the unknown: speech activity detection in heterogeneous video collections. In *Proc. Interspeech*, 2925–2928.
- JESSEN, M. (2008). Forensic Phonetics. In *Language and Linguistics Compass*, 2/4, 671-711.
- KINNUNEN, T. & LI, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. In *Speech communication*, 52(1), 12-40.
- KINNUNEN, T. & RAJAN, P. (2013). A practical, self-adaptive voice activity detector for speaker verification with noisy telephone and microphone data. In *Proc. ICASSP*, 7229–7233.
- KENNY, P. (2010). Bayesian speaker verification with heavy tailed priors. In *Proc. Odyssey: The Speaker and Language Recognition Workshop*, 69(7), 349-356.

- MANCARELLA, G.B. (1998). *Salento. Monografia regionale della "Carta dei Dialetti Italiani"*. Lecce: Edizioni del Grifo.
- MORRISON, G.S. (2009). Forensic voice comparison and the paradigm shift. In *Science and Justice*, 49, 298-308.
- NAKASONE, H., MIMIKOPOULOS, M., BECK, S. & MATHUR, S. (2004). Pitch synchronized speech processing (PSSP) for speaker recognition. In *Proc. Speaker Odyssey: the Speaker Recognition Workshop (Odyssey 2004)*, 251–256.
- NG, R., NICOLAO, M., SAZ, O., HASAN, M., CHETTRI, B., DOULATY, M., LEE, T. & HAIN, T. (2016). The Sheffield language recognition system in NIST LRE 2015. In *Proc. Odyssey 2016: The Speaker and Language Recognition Workshop*, 181-187.
- PLCHOT, O., MATEJKA, P., FER, R., GLEMBEK, O., NOVOTN, O., PESAN, J., VESEL, K., ONDEL, L., KARAFIAT, M., GREZL, F., KESIRAJU, S., BURGET, L., BRUMMER, N., SWART, A., CUMANI, S., MALLIDI, S.H. & LI, R. (2016). BAT system description for NIST LRE 2015. In: *Proc. Odyssey 2016: The Speaker and Language Recognition Workshop*, 166-173.
- RAMIREZ, J., SEGURA, J., BENITEZ, C., TORRE, A.D.L. & RUBIO, A. (2004). Efficient voice activity detection algorithms using long-term speech information. In *Speech Communication*, 42, 3-4.
- REYNOLDS, D.A. (1997). Comparison of background normalization methods for text-independent speaker verification. In *Proc. Eurospeech*, 963–966.
- RIBEZZO, F. (1912). *Il dialetto apulo-salentino di Francavilla Fontana*. Bologna: Forni.
- ROBERTSON B., & VIGNAUX, G.A., (1995). *Interpreting evidence*. London: Wiley.
- ROMITO, L. & GALATÀ, V. (2008). Speaker Recognition in Italy: evaluation of methods used in forensic cases, In *Actas del IV Congreso de Fonética Experimental (A. Pamies & E. Melguizo, editors)*. Granada: Método Ediciones, 229-240.
- ROMITO L, BOVE T, DELFINO S, JONA LASINIO G, ROSSI C (2009). *Specifiche linguistiche del database utilizzato per lo Speaker Recognition in S.M.A.R.T.*. In ROMITO, L., ROSITA, L. & GALATÀ, V. (Eds.). *La Fonetica Sperimentale: metodo e applicazioni*, Milano: Officinaventuno, pp. 632-640.
- ROSE, P. (2002). *Forensic Speaker Identification*. New York: Taylor and Francis.
- ROSE, P. (2005). Forensic Speaker Recognition at the beginning of the twenty-first century – an overview and a demonstration. In *Australian Journal of Forensic Science*, 37, 49-72.
- ROSE, P. (2006). Technical Forensic Speaker Recognition: Evaluation, types and testing of evidence, In *Computer Speech and Language*, 20, 159–191.
- SHIN, J., CHANG, J.H. & KIM, N. (2010). Voice activity detection based on statistical models and machine learning approaches. In *Computational Speech Language*, 24(3), 515–530.
- SIMPSON, A.P. (2009). Phonetic differences between male and female speech. In *Language and Linguistics Compass*, 3(2), 621–640.
- SOHN, J., KIM, N. & SUNG, W. (1999). A statistical model-based voice activity detection. In *IEEE Signal Process. Lett.*, 6, 1–3.
- THOMAS, S., SAON, G., EGBROECK, M. & NARAYANAN, S., (2015). Improvements to the IBM speech activity detection system for the DARPA RATS program. In *Proc. ICASSP*, 4500-4504.

TUCKER, R., (1992). Voice activity detection using a periodicity measure. In *IEEE Proc. (Commun. Speech Vis.)*, 139(4), 377–380.

WU, J. & ZHANG, X. (2011). Maximum margin clustering based statistical VAD with multiple observation compound feature. In *IEEE Signal Process. Lett.* 18(5), 283–286.

ZHANG, X. & WU, J. (2013). Deep belief networks based voice activity detection. In *IEEE Trans. Audio Speech Lang. Processing*, 21(4), 697-710.

ZILCA, R., KINGSBURY, B., NAVRÁTIL, J. & RAMASWAMY, G. (2006). Pseudo pitch synchronous analysis of speech with applications to speaker recognition. In *IEEE Trans. Audio, Speech and Language Processing*, 14(2), 467–478.



UMAR MUHAMMAD-GOMBE, PETER FRENCH, ELEANOR CHODROFF

## A Comparative Analysis of Nigerian Linguistically-Naïve Native Speakers and Nigerian Linguist Native Speakers Categorising Four Accents of Nigerian English

In the field of LAAP (Language Analysis in the Asylum Process), there has been debate over who should undertake the task of inferring a speaker's country and region of socialisation from their language and dialect. This chapter investigates whether native speaker linguists perform more accurately than native speaker non-linguists in determining the first language of Nigerian speakers of English. Eighty non-linguist and 25 linguist speakers of Hausa, Igbo, Kanuri and Yoruba were recruited. They listened to 30-second recordings of Nigerian-accented English, and assigned each to an L1 (Hausa, Igbo, Kanuri, Yoruba or non-Nigerian). Listeners of both groups were most accurate in classifying accents of their own L1. Linguists did not differ significantly from non-linguists in accuracy. The results provide empirical support for having educated non-linguist native speakers involved in LAAP casework.

*Keywords:* Accent classification, Linguist native speakers, Language analysis, Asylum seekers, National origin

### 1. Introduction

In the field of LAAP<sup>1</sup> (Language Analysis in the Asylum Process), there has been debate over who should undertake the task of inferring a speaker's country and region of socialisation based on their language and dialect: academic or professional linguists with detailed knowledge of the languages/varieties that may be at issue (see LNOG, 2004; Fraser, 2009, 2011; Patrick, 2010, 2012, 2016, 2019), or linguistically-naïve native speakers of those languages/varieties, or a combination of both (see Cambier-Langeveld, 2010b; 2012; Cambier-Langeveld, 2007, cited in Cambier-Langeveld, 2016; Fraser, 2011; Foulkes, French & Wilson, 2019; Wilson, 2009). Opposing positions within this debate have largely been argued on principle alone, without support from empirical studies. Contributing to this debate, this chapter concerns research which was designed to determine whether native speaker linguists perform more accurately than native speaker non-linguists in determining the first language (L1) of Nigerian speakers of English.

Previous work in this area has largely been based on intuition; however, a few studies have attempted an empirical approach. One pioneering instance of this

---

<sup>1</sup> This area was formerly termed 'LADO'— Language Analysis in the Determination of Origin. This term is still used by some authors; LAAP is used throughout this chapter.

comes from Wilson (2009). This work attempted to determine the most reliable method for identifying people's nationality or region of socialisation from Ghanaian English. Wilson (2009) has four groups: (1) native speakers of Ghanaian English, (2) non-native academic and postgraduate linguists (3) undergraduates of linguistics and (4) LADO/LAAP professionals. Both groups (2) and (3) were provided with working material on features of Ghanaian English in advance of the experiments. Wilson observed that native speakers of Ghanaian English were the best performing group despite their lack of linguistic expertise. Though Wilson (2009) had indeed included linguists as a test group in the experiment, these individuals were not native speakers of Ghanaian English. As mentioned earlier, these participants mainly encountered the tested English variety through the working material given to them in advance of the experiment.

A thorough investigation is necessary to determine whether native speaker linguists outperform native speaker non-linguists (as tested in Wilson, 2009) in language analysis. With this paradigm, one would be able to infer the potential influence of linguistic knowledge on performance in a more clear and exact manner as the two groups have the same language background but differ in linguistic expertise. Determining whether such a difference exists and is statistically significant is of utmost value to the debate concerning who should conduct the task of analysing a speaker's language in the asylum process.

Having identified the need for comparing naïve and linguist native speakers of the same accent group, two research questions were used to guide this research: (1) How and by whom in a LAAP (Language Analysis in the Asylum Procedure) context should the analysis of spoken English be analysed? Is it native speaker linguists? Or are native speaker non-linguists equally accurate? (2) If linguistic training is found to be effective, which specific linguistic expertise will be required?

In response to question (1), we first hypothesised that native speakers of any of the following four Nigerian languages – Hausa, Igbo, Kanuri and Yoruba – would outperform the respective non-native speakers of those four Nigerian languages in identifying the speaker's L1 from their accented English. In addition, we hypothesised that native speaker linguists would outperform their native speaker non-linguist counterparts in identifying fellow speakers of their own native accent. In response to question (2), we hypothesised that native speaker linguists who were also phoneticians would outperform non-phonetician linguists.

We investigated the following questions as a baseline assessment of the accent performance task:

1. Are native speakers indeed more accurate at classifying their own L1 from the accented English samples relative to non-native speakers?
2. Are some languages simply more difficult to classify than others?
3. Is the relationship between confidence and accuracy significant?

In addition, we statistically assessed the following research questions:

1. Are linguists better than non-linguists at accent identification among the four Nigerian languages?

2. Are native speakers specifically with a linguistic background in phonetics better at accent identification than other native speakers?

An accent classification experiment was conducted to address these research questions. The experiment included native speakers of Hausa, Igbo, Kanuri and Yoruba who either did not have linguistic training (non-linguists) or did have linguistics training (linguists). The stimuli were recordings of spoken English from native speakers of these four languages along with two foil languages. The following sections review the methods and results, which include a comparison of the relevant participant groups (non-linguist versus linguist and non-phonetician linguists versus phoneticians). This is then followed by a discussion and conclusion.

## 2. *Methods*

### 2.1 Participants

The native speaker non-linguist group comprised 80 linguistically naïve, educated, native speakers of Hausa, Igbo, Kanuri, and Yoruba. The group was predominantly composed of university students and administrative staff. All participants were recruited at universities in the cities of Kano, Nsukka, Maiduguri and Ibadan. The dominant language of each of these cities corresponds to a relevant L1 test language: Hausa, Igbo, Kanuri, and Yoruba, respectively. Each non-linguist L1 group was represented by 20 speakers, and their ages ranged from 18 to 70 years (mean age = 28, SD = 9). The full native speaker linguist group comprised 25 academic linguists with various specialisations in e.g., phonetics, phonology, syntax, semantics and sociolinguistics of their Nigerian L1. The original plan was to recruit only sociophoneticians in the linguistics group, but their scarcity resulted in the need to draw participants from a broader range of linguistics specialisations. Linguists who self-classified as a phonetician or phonologist were grouped together in the present study to represent experts in the physical sounds or sound patterns of language. In subsequent sections, we refer to this group as the broad “phonetician” group. The linguists were recruited at Bayero University in Kano (6 L1 Hausa linguists including 1 phonetician), the University of Nigeria in Nsukka (9 L1 Igbo linguists including 1 phonologist), the University of Maiduguri (5 L1 Kanuri linguists including 1 phonologist), and the University of Ibadan (5 L1 Yoruba linguists including 1 phonetician and 1 phonologist).

### 2.2 Stimuli

The experiment employed eighteen recordings of accented English that were approximately 30 seconds in duration. The recordings comprised bits of “The Rainbow Passage” and spontaneous speech in which the speaker narrated aspects of their life experience. Sixteen recordings of L1 speakers of Hausa, Igbo, Kanuri and Yoruba speaking in English were selected. These represented four speakers

from each of the four language groups: Hausa, Igbo, Kanuri and Yoruba. These recordings were made during a fieldwork visit to Nigeria. Additionally, two foil recordings (Ghanaian and Guinean English speakers) were added for a total of 18 recordings. These latter recordings were made by Ghanaian and Guinean English speakers who were pursuing their Masters degree at the University of York. They were recorded in a studio in the Department of Language and Linguistic Science.

### 2.3 Procedure

Using Qualtrics Survey Software, each participant was asked to listen to the recordings and assign each of the recordings to an L1 accent (Hausa, Igbo, Kanuri, Yoruba or non-Nigerian). Following classification, each participant was asked to report their confidence in their decision on a scale from 0 to 100.

## 3. Results

To address the research questions, we provide a descriptive and inferential analysis of accuracy in identifying both speakers of native and non-native accents for each of the relevant subgroups: non-linguists versus linguists (section 3.1), and non-phonetician linguists versus phoneticians (section 3.2). For the inferential analysis, we implemented two logistic mixed-effects models to investigate variation in accuracy. Both models assessed the baseline questions of accent performance. The first model specifically targeted whether native speaker linguists were better than native speaker non-linguists at accent identification. The second targeted the question of whether native speaker phoneticians specifically were better than other native speaker non-phonetician linguists at accent identification. In addition, individual performances were also investigated in an exploratory analysis.

### 3.1 Non-linguists versus linguists

#### 3.1.1 Descriptive statistics

Overall, non-linguist and linguist listeners were substantially more accurate at identifying the L1 accent of a recording when it matched their own L1 (Figure 1 and Table 1). This was reasonably consistent across L1 backgrounds, with the one exception of the Kanuri linguist group (Tables 2 and 3). For the non-linguist L1 groups, accuracies at identifying their own L1 accent ranged from 71% to 81% (Table 2), and for the linguist L1 groups, accuracies ranged from 75% to almost 92%, with the exception of Kanuri (Table 3). The Kanuri linguist group had a lower accuracy of 35%. Overall, non-linguists and linguists had highly comparable accuracy rates in identifying fellow speakers of their accents (76.6% and 76.0%, respectively; Table 4). The Hausa, Igbo and Yoruba linguists were numerically more accurate than their non-linguist counterparts in identifying accents of their own L1, but this did not hold for the Kanuri listeners.

Accuracies generally dropped for classification of Nigerian accents that did not match the L1 of the listener (Figure 1). Linguists numerically outperformed their non-linguist counterparts in classifying other L1 Nigerian accents that were not their own, though the results were somewhat more mixed for classification of non-Nigerian accents (Table 1). For non-linguist L1 groups, these accuracies ranged from 37.5% to 50.0% for other Nigerian L1 accents and from 12.5% to 27.5% for non-Nigerian L1 accents. For linguist L1 groups, these accuracies ranged from 45.0% to 56.9% for other Nigerian L1 accents from 8.3% to 44.4% for non-Nigerian L1 accents.

Figure 1 - Accuracy of classifying the accent of own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from non-linguists and linguists in each L1 group. Error bars reflect  $\pm$  one standard error of the proportion

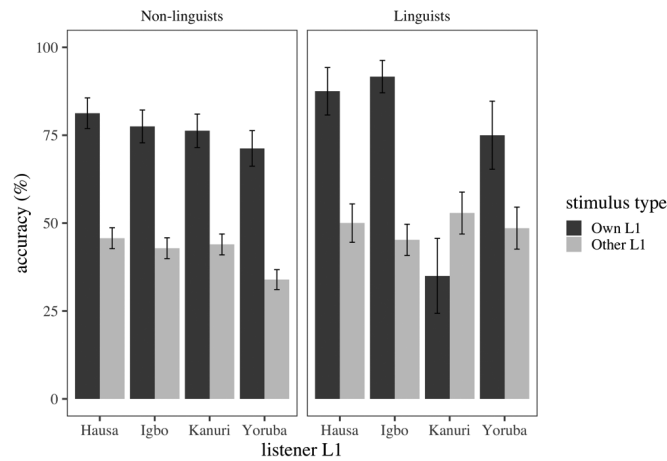


Table 1 - Accuracy of classifying the accent of each stimulus type from non-linguists and linguists

Stimulus Set	Non-linguists	Linguists
Own L1	76.6% (245/320)	76.0% (76/100)
Other Nigerian L1s	45.1% (433/960)	52.0% (156/300)
Non-Nigerian L1s	20.6% (33/160)	28.0% (14/50)
Overall	49.4% (711/1440)	54.7% (246/450)

Table 2 - Accuracy of classifying the accent of each stimulus type from non-linguists in each L1

L1 Group	Non-linguists			Overall
	Own L1	Other Nigerian L1s	Non-Nigerian L1s	
Hausa	81.2% (65/80)	50.0% (120/240)	20.0% (8/40)	53.6% (193/360)
Igbo	77.5% (62/80)	46.2% (111/240)	22.5% (9/40)	50.6% (182/360)
Kanuri	76.2% (61/80)	46.7% (112/240)	27.5% (11/40)	51.1% (184/360)
Yoruba	71.2% (57/80)	37.5% (90/240)	12.5% (5/40)	42.2% (152/360)

Table 3 - Accuracy of classifying the accent of each stimulus type from linguists in each L1 group

L1 Group	Linguists				Overall
	Own L1	Other Nigerian L1s	Non-Nigerian L1s		
Hausa	87.5% (21/24)	56.9% (41/72)	8.3% (1/12)		58.3% (63/108)
Igbo	91.7% (33/36)	45.4% (49/108)	44.4% (8/18)		55.6% (90/162)
Kanuri	35.0% (7/20)	55.0% (33/60)	40.0% (4/10)		48.9% (44/90)
Yoruba	75.0% (15/20)	55.0% (33/60)	10% (1/10)		54.4% (49/90)

Table 4 - Accuracy of classifying the accent of own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from non-linguists and linguists in each L1 group

L1 Group	Own L1		Other L1s	
	Non-linguists	Linguists	Non-linguists	Linguists
Hausa	81.2%	87.5%	45.7%	50%
Igbo	77.5%	91.6%	42.9%	45.2%
Kanuri	76.2%	35.0%	43.9%	52.9%
Yoruba	71.2%	75.0%	37.5%	48.6%

The above findings strongly indicate that listeners were much more accurate when classifying their own L1 accent than when classifying others. In addition, Table 5 presents the confusion matrix among the stimuli accents. L1 accents were not uniformly confusable: a Kanuri accent was most often mistaken for a Hausa accent, whereas a Yoruba accent was most often mistaken for an Igbo accent and vice versa.

Table 5 - Confusion matrix of accent responses against stimulus accents for all participants in the study

Response	Stimulus accent				
	Hausa	Igbo	Kanuri	Yoruba	Non-Nigerian
Hausa	273	4	198	14	23
Igbo	26	252	22	130	63
Kanuri	88	21	151	16	37
Yoruba	11	113	19	234	40
Non-Nigerian	22	30	30	26	47

In the following analysis, we investigated the errors surrounding the classification of accents that do indeed *match* the L1 of the listener. Although the task was a five-way forced choice classification, we can calculate the responses based on whether a native speaker correctly matched the accent in the recording to their own L1 (true positive), whether they incorrectly matched the accent in the recording to their own L1 (false positive), whether they correctly rejected another L1 accent as not the same as their own L1 (true negative), or whether they incorrectly identified another L1 accent as their own L1 (false negative). Table 6 shows the false negative and false

positive rates for each L1 group. The false negative rate is calculated as the number of false negatives divided by the total number of false negatives and true positives: when the accent was indeed the listener's L1, how many times did the listener fail to classify it as the L1? The false positive rate is calculated as the number of false positives divided by the total number of false positives and true negatives: when the accent was indeed not the listener's L1, how many times did the listener classify it as the L1?

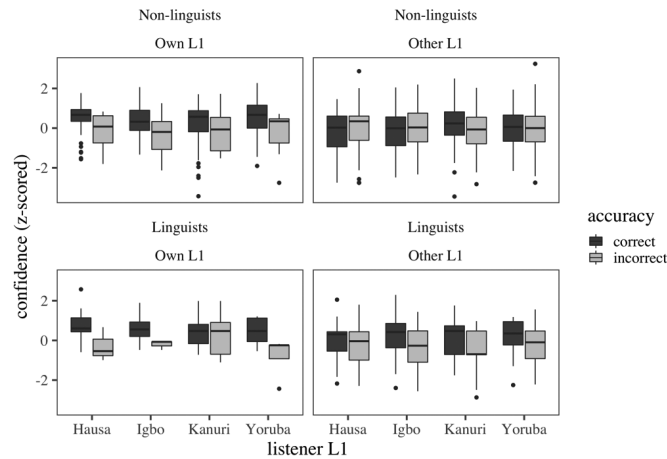
With respect to the false negative rate, the linguist groups had slightly lower false negative rates than their non-linguist counterparts, except for the Kanuri group. This suggests that the Hausa, Igbo, and Yoruba linguists were numerically more precise in identifying accents of their own L1 when presented with them than the non-linguists. For Kanuri, this pattern was reversed. With respect to the false positives, the linguist groups again had slightly lower false positive rates than their non-linguist counterparts except for the Igbo group. Overall, linguists were slightly less likely to accept a non-member of their L1 group as a fellow member based on their spoken English.

Table 6 - False negative and false positive rates for non-linguists and linguists in each L1 group

L1 Group	False negatives		False positives	
	Non-linguists	Linguists	Non-linguists	Linguists
Hausa	18.8%	12.5%	18.9%	15.5%
Igbo	22.5%	8.3%	13.6%	17.5%
Kanuri	23.8%	65.0%	11.1%	4.3%
Yoruba	28.7%	25.0%	16.1%	8.6%

As stated in the methods section above, the listener provided a self-confidence rating on a scale from 0 to 100 after classifying each audio clip. This was to determine how confident the participant was in their classification decision. Figure 2 shows the average z-scored self-confidence ratings by response accuracy across participants in the non-linguist and linguist groups for each of the L1 backgrounds, and when categorising their own L1 and other L1s. For accents that matched the L1 of the participant, confidence was generally higher for correct than incorrect responses. This same general pattern was mostly observed for other L1 accents, particularly for linguists. The two exceptions were the non-linguist Hausa and Igbo non-linguist groups. Further exploration of individual patterns is provided in section 3.3.

Figure 2 - Confidence ratings (z-scored) for each response accuracy when classifying own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from non-linguists and linguists



### 3.1.2 Inferential statistics

To test whether Nigerian linguists significantly outperformed their Nigerian non-linguist counterparts, the full dataset containing both Nigerian linguists and Nigerian non-linguists was used. Accent identification accuracy was analysed as a binary dependent variable (1 = correct, 0 = incorrect) with a logistic mixed-effects model using the lme4 R package (Bates et al., 2015; RStudio Team, 2020). The model included fixed effects of native speaker match, linguist status (“linguist”), stimulus language, confidence rating, the interaction between linguist and confidence rating, as well as a random intercept for participant. Models with more complex random effect structures failed to converge.

Native speaker match had two levels: whether the L1 of the participant (the listener) matched the L1 of the recorded speaker or not. This variable was sum-coded with no-match as the held-out level. Linguist had two levels: whether the participant was a linguist or non-linguist. This factor was sum-coded with non-linguist as the held-out level. The factor for stimulus language had five levels: Hausa, Igbo, Kanuri, Yoruba or non-Nigerian. This factor was sum-coded with non-Nigerian as the held-out level. As mentioned in the methodology section, the experiment included self-rating confidence levels using a sliding scale from 0–100. These confidence levels were converted to z-score confidence values for each participant. The interaction between linguist and confidence was included to check the presence or absence of a significant correlation between Nigerian linguists’ confidence and accuracy in classifying the English accents. The alpha level for determining significance was set to 0.05: predictors with a p-value less than 0.05 were considered significant.

The expectation was for a significant and positive effect of native speaker match, indicating a higher accuracy when the listener’s native language matched the



language of the stimulus. Another expectation was for a significant effect of linguist status, indicating higher overall accuracy for linguists than non-linguists. The effect of stimulus language was included to determine whether some languages were more difficult to identify than others.

As predicted, native speaker match was significant, indicating that listeners exposed to stimuli of their L1 were approximately two times as likely to be accurate than when exposed to stimuli of other L1s ( $\beta_{match} = 0.70, p < 0.001$ ). The result of stimulus language showed that accuracy differed significantly depending on the presented accent, but in different directions. Listeners were significantly more accurate for Hausa, Igbo, and Yoruba classification ( $\beta_{Hausa} = 0.74, p < 0.001$ ;  $\beta_{Igbo} = 0.48, p < 0.001$ ;  $\beta_{Yoruba} = 0.29, p < 0.01$ ), but significantly less accurate for Kanuri classification ( $\beta_{Kanuri} = -0.59, p < 0.001$ ).

The effect of linguist was not significant ( $\beta_{linguist} = 0.14, p = 0.12$ ), indicating that native linguist speakers did not perform more or less accurately than their non-linguist counterparts. A significant positive relationship was also observed between confidence and accuracy for all listeners ( $\beta_{zconf} = 0.25, p < 0.01$ ), indicating that listeners were more confident when accurate. This correlation was even stronger for native linguists than non-linguists ( $\beta_{linguist:zconf} = 0.15, p < 0.05$ ).

## 3.2 Non-phonetician linguists versus phoneticians

### 3.2.1 Descriptive statistics

This section compares the performance of the Nigerian native speaker phoneticians against that of the non-phonetician linguists. We preface these analyses with a reminder that the sample size of the linguist group was smaller than the non-linguist group (25 linguist listeners total), and particularly the number of representative phoneticians (20 non-phonetician linguists; 5 phoneticians: 1 Hausa speaker, 1 Igbo, 1 Kanuri, and 2 Yoruba). Ideally, future research would be able to access a larger sample size for more stable inferential conclusions. We present a high-level overview of observed patterns in the data, and provide a preliminary analysis of whether native speaker phoneticians indeed outperform other native speaker, non-phonetician linguists.

As shown in Figure 3 and Table 7, phoneticians marginally outperformed non-phonetician linguists in overall accuracy, but the performances were otherwise highly comparable. In classifying accents that matched their own L1, non-phonetician linguists had a numerically higher accuracy at 76.2% against the phonetician accuracy of 75.0%. Phoneticians had numerically higher accuracies than non-phonetician linguists in classifying other Nigerian L1s and non-Nigerian L1 accents.

Figure 3 - Accuracy of classifying the accent of own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from non-phonetician linguists and phoneticians in each L1 group. Error bars reflect  $\pm$  one standard error of the proportion

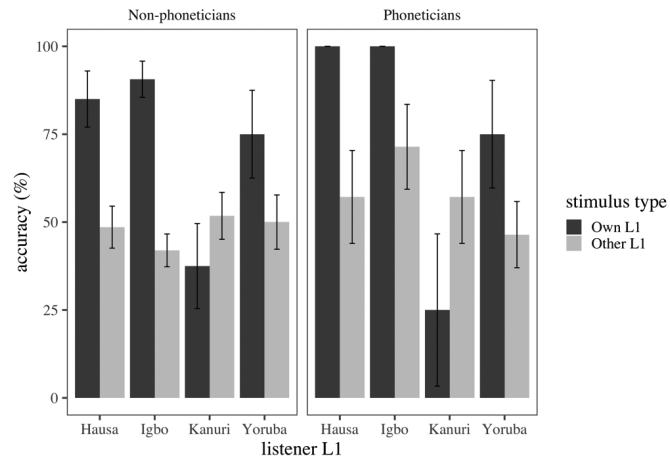
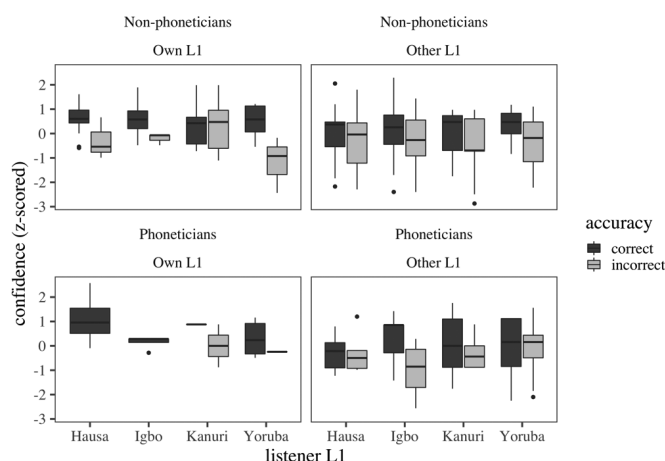


Table 7 - Accuracy of classifying the accent of each stimulus type from non-phonetician linguists and phoneticians in each L1 group

Stimulus Set	Non-phoneticians	Phoneticians
Own L1	76.2% (61/80)	75.0% (15/20)
Other Nigerian L1s	50.0% (120/240)	60.0% (36/60)
Non-Nigerian L1s	27.5% (11/40)	30.0% (3/10)
Overall	53.3% (192/360)	60.0% (54/90)

Figure 4 shows the average z-scored self-confidence ratings by response accuracy across participants in the non-phonetician and phonetician groups for each of the L1 backgrounds, and when categorising their own L1 and other L1s. In the previous section, we found that at a group level, linguists had higher confidence for correct than incorrect responses. In cases when correct and incorrect responses are observed, this pattern also held for each of the non-phonetician and phonetician subgroups. No major observable differences were observed between these two subgroups in the confidence ratings.

Figure 4 - Confidence ratings (z-scored) for each response accuracy when classifying own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from non-phonetician linguists and phoneticians



### 3.2.2 Inferential statistics

To test whether native speaker phoneticians significantly outperformed their non-phonetician linguist counterparts, the same model as described above in section 3.1.2 was run, but using only the linguist data and replacing the two-level factor of “linguist” with the two-level factor of “phonetician”. The two levels of phonetician were: native speaker phonetician and native speaker non-phonetician linguist. The variable was sum-coded with native speaker non-phonetician linguist as the held-out level.

As in the previous model, native speaker match was significant, indicating that the linguist subgroup was also better at identifying their own L1 accent than other accents ( $\beta_{match} = 0.48, p < 0.01$ ). As before, stimulus language also had a significant influence on accuracy: Nigerian linguists regardless of specialty were significantly more accurate with Hausa and Igbo ( $\beta_{Hausa} = 0.84, p < 0.01$ ;  $\beta_{Igbo} = 0., p < 0.01$ ), and significantly less accurate with Kanuri ( $\beta_{Kanuri} = -1.11, p < 0.001$ ); accuracy on Yoruba did not differ from the average accuracy in the task ( $\beta_{Yoruba} = 0.12, p = 0.55$ ).

Though phoneticians were numerically slightly more accurate than non-phonetician linguists, the effect of phonetician was not significant, indicating that Nigerian phoneticians were not significantly better than other linguists ( $\beta_{phonetician} = 0.17, p = 0.32$ ). In addition, the overall relationship between confidence and accuracy was significant ( $\beta_{zconf} = 0.39, p < 0.01$ ); however, no significant difference was observed in the effect of confidence between phonetician and non-phonetician linguists ( $\beta_{phonetician:zconf} = -0.04, p = 0.78$ ).

### 3.3 Exploration of individual participant accuracy

The above results show the overall accuracies of the participants at the group and sub-group levels. This, however, may not precisely indicate consistency across all

individual participants in the experiments. We additionally explored the consistency in the identification task across the individual Nigerian linguists. Figure 5 shows the range of accuracies across the non-linguist listeners in classifying accents of their own L1 against accents of other L1s. Figure 6 shows the equivalent data for the linguist listeners.

A full 74 out of 80 non-linguist participants were numerically better at classifying the L1s of their own accent than they were at their classifying other accents; only 6 out of 80 non-linguist participants showed the opposite pattern. Among linguists, 20 out of 25 participants were numerically better at classifying their own accent than other accents; just 5 out of 25 participants showed the opposite pattern. Overall accuracy of classifying all accents for non-linguists ranged from 16.7% to 83.3% (median = 50.0%, mean = 49.4%). Overall accuracy of classifying all accents for linguists ranged from 27.8% to 77.8% (median = 55.6%, mean = 54.7%). Individual overall performance from non-linguists reached overall higher accuracies than individual linguists; however, the range of non-linguist individual accuracies was much higher and reached overall lower accuracies as well.

Figure 5 - Accuracy of classifying the accent of own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from each non-linguist participant in each L1 group

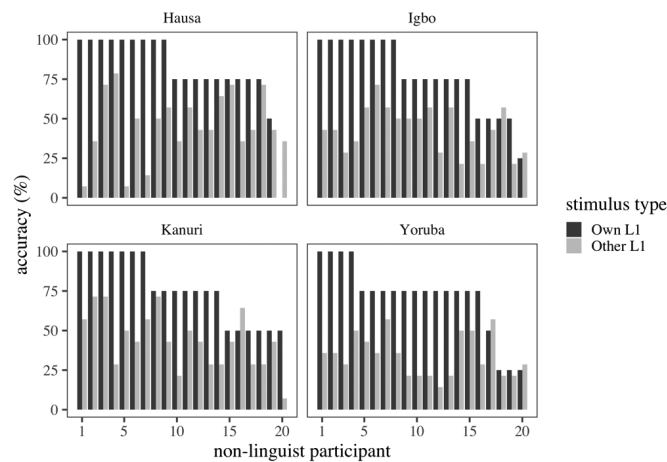


Figure 6 - Accuracy of classifying the accent of own L1-accented stimuli and other L1-accented stimuli (Nigerian and non-Nigerian) from each linguist participant in each L1 group

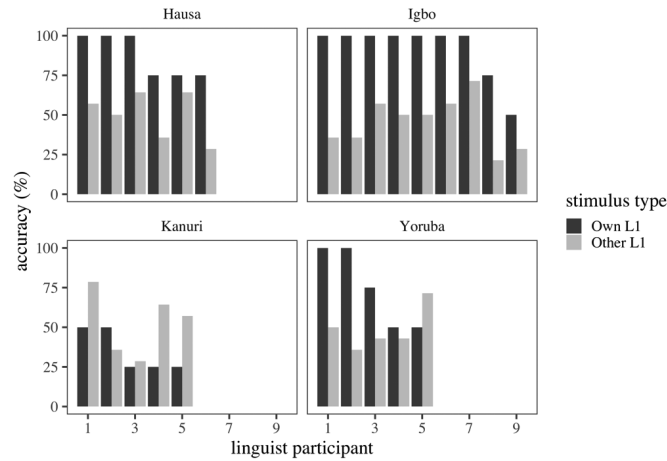


Figure 7 shows the relationship between confidence and accuracy for each individual non-linguist participant. Figure 8 shows the equivalent data, but for linguist participants. The relationship between confidence and accuracy was somewhat variable among non-linguists. In line with the model results in which confidence positively correlated with accuracy, 47 out of 80 non-linguist participants were on average more confident on correct than incorrect responses. However, 33 out of 80 participants were on average more confident on incorrect than correct responses. This relationship between confidence and accuracy was much more consistent among linguists, which reflects the model results. 21 out of 25 linguists were on average more confident on correct than incorrect responses; only 4 linguists showed the opposite ranking.

Figure 7 - Confidence ratings (z-scored) for each response accuracy from non-linguist participants in each L1 group

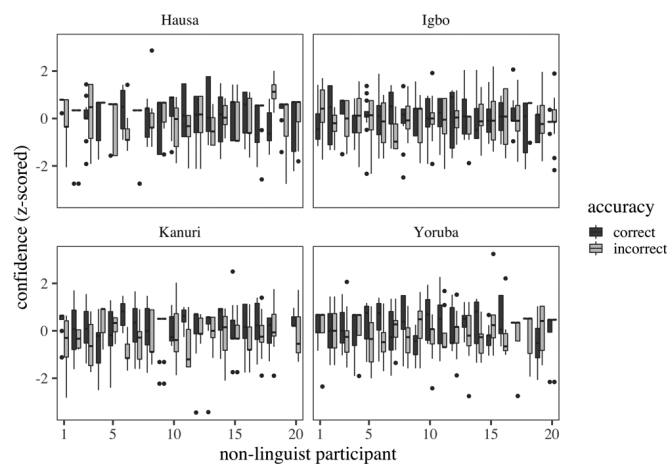
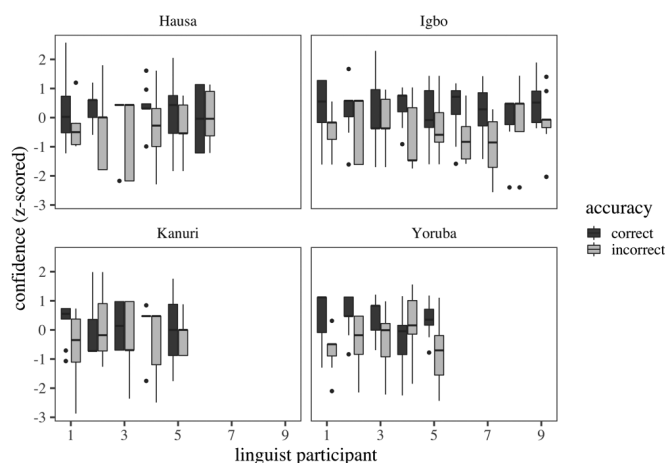


Figure 8 - Confidence ratings (z-scored) for each response accuracy from linguist participants in each L1 group



#### 4. Discussion

Overall, Nigerian native speakers from all four L1 backgrounds, irrespective of linguistic expertise, performed well above chance level. A mixed-effects logistic regression model of accuracy revealed a significant influence of native speaker match, indicating that listeners classifying accents of their own L1 were significantly more likely to be accurate than when classifying accents of other languages. The linguists, however, were only slightly better than the non-linguists on a numerical basis; the difference was not significant and applied to only 3 of the 4 L1 groups.

As anticipated, most native speaker listeners significantly outperformed other listeners when exposed to stimuli of fellow speakers of their native accent. Even Kanuri linguists – who were less accurate compared to other groups – were more accurate than non-native Kanuri listeners in identifying Kanuri-English stimuli. Although most of the linguists outperformed the non-linguists in identifying fellow speakers of their accent, this difference was not statistically significant. Numerically, this finding supports the expectation that linguists would outperform their non-linguist counterparts; some phoneticians even reached 100% accuracy. However, the difference in accuracy between the linguist and non-linguist group failed to reach statistical significance. Thus, this investigation finds general native speaker status as the most reliable variable for significantly boosting accent identification accuracies. Some pieces of past research have reported similar findings to support the value of native speaker expertise, as discussed below.

Hedegard's (2015) study also found that native speakers linguists did not significantly differ from native speakers non-linguists in classifying a Japanese dialect from a spoken audio sample of Japanese. All native speakers were, however, significantly more accurate than non-native linguists with familiarity of Japanese linguistics. Jenkins's (2016) study indicates that native Scottish listeners were the

best in identifying and distinguishing genuine Scottish-accent speakers from non-genuine accent mimics. However, there was no difference between linguists and non-linguist listener groups in making correct judgements. Nolan (2012) argued that native speakers' knowledge of their language differed from the expertise of linguists with special interest in the language in question. He also argued that native speakers' intuition of detecting the speech of a fellow speaker cannot be fully represented by capturing precise behaviour of speech organs in descriptive linguistics using notations of transcription. Hoskin's (2018) study further supports the value of native speaker involvement in LAAP. He argued that non-linguist native speakers of Kurmanji have demonstrated their awareness of heterogeneity, linguistic accommodation and language mixing by identifying further heterogeneity and complexities in spoken Kurmanji, and such features were not included in the available Kurmanji literature.

In the present study, individual accuracies of the two native speaker non-linguist and linguist groups demonstrate that some non-linguists outperformed their linguist counterparts in the classification task, and maximum overall accuracy rates were achieved by individual non-linguist listeners. Considering this finding, we hold the view that success in carrying out language analysis is dependent on the talent and experience of the individual analyst rather than a linguistic qualification alone. Wilson (2009) argued, based on the marginal difference between linguist groups' accuracy and performance, that in-depth linguistic expertise may not be more significant than a short training for a reliable language analysis. This view supports the position already taken by Foulkes, French & Wilson (2019) and Cambier-Langeveld (2010).

### 5. Conclusion

Native speaker linguists were only slightly more accurate than native speaker non-linguists in identifying the L1 accent in spoken English; however, the difference between these two groups was not significant. These findings thus offer empirical support for having educated native speakers involved in LAAP casework, even without linguistic training. Further, the findings of the current study as well as other past studies (such as those mentioned above) have discovered the success rates of native speakers in identifying their own speaker group. It could therefore be argued that the involvement of native speakers in LAAP casework is of utmost significance, and such involvement may only serve as one of the several steps taken in the complex procedure of asylum applications and decisions.

Further research could consider the role of explicit, accent-specific training in language analysis, and whether linguists and non-linguists can employ this training to significantly improve accuracy. Given the above findings, native speakers who receive good training and demonstrate a strong potential when tested may be suitable for the task of language analysis. However, being a native speaker alone does not automatically qualify a person to conduct every forensic speech task. Non-linguist native speakers should receive appropriate training for language analysis and only work in a team under the supervision of a linguist. Our findings do suggest

that native speaker linguists will have less variability in performance than native speaker non-linguists; however, individual native speaker non-linguists can also perform with high accuracy. In addition, it may also be beneficial to have asylum speakers perform an accent classification task as well given the strong role of native language match on accent classification performance. As recommended by Wilson (2016), we also recommend that relevant authorities consider this additional testing method by asking asylum seekers to distinguish speech samples spoken by fellow speakers of their native accent from several other samples.

### References

- BATES, D., MÄCHLER, M., BOLKER, B., & WALKER, S. (2015). Fitting linear mixed-effects models using lme4. In *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- CAMBIER-LANGEVELD, T. & SAMSON (2007). Current methods in forensic speaker identification: Results of a collaborative exercise. In *International Journal of Speech, Language and the Law*, 14(2), 223–243. <https://doi.org/10.1558/ijssl.2007.14.2.223>
- CAMBIER-LANGEVELD, T. (2010b). The role of linguists and native speakers in language analysis for the determination of speaker origin. In *International Journal of Speech, Language and the Law*, 17(1), 67–93. <https://doi.org/10.1558/ijssl.v17i1.67>
- CAMBIER-LANGEVELD, T. (2012). Clarification of the issues in language analysis: A rejoinder to Fraser and Verrips. In *International Journal of Speech, Language and the Law*, 19(1), 95–108. <https://doi.org/10.1558/ijssl.v19i1.95>
- CAMBIER-LANGEVELD, T. (2016). Language analysis in the asylum procedure: A specification of the task in practice. In *International Journal of Speech, Language and the Law*, 23(1), 25–41. <https://doi.org/10.1558/ijssl.v23i1.17539>
- FOULKES, P., FRENCH, P., & WILSON, K. (2019). LADO as Forensic Speaker Profiling. In PATRICK, P.L., SCHMID, M. & ZWAAN, K. (Eds.), *Language Analysis for the Determination of Origin*. Cham, Switzerland: Springer Nature, 91–116. [https://doi.org/10.1007/978-3-319-79003-9\\_6](https://doi.org/10.1007/978-3-319-79003-9_6)
- FRASER, H. (2009). The role of “educated native speakers” in providing language analysis for the determination of the origin of asylum seekers. In *International Journal of Speech, Language and the Law*, 16(1), 113–138. <https://doi.org/10.1558/ijssl.v16i1.113>
- FRASER, H. (2011). The role of linguists and native speakers in language analysis for the determination of speaker origin: A response to Tina Cambier-Langeveld. In *International Journal of Speech, Language and the Law*, 18(1), 121–130. <https://doi.org/10.1558/ijssl.v18i1.121>
- HEDEGARD, H. (2015). *Language analysis for the determination of origin (LADO): The native speaker dimension*. University of York (UK): MSc thesis.
- HOSKIN, J. (2018). Native speaker non-linguists in LADO: An insider perspective. In NICK, I.M. (Eds.), *Forensic Linguistics. Asylum-seekers, Refugees and Immigrants*. Malaga: Vernon Press, 23–40.



- JENKINS, M. (2016). *Identifying an imitated accent: humans vs. computers*. University of York (UK): MSc thesis.
- LNOG (2004). Guidelines for the use of language analysis in relation to questions of national origin in refugee cases. In *International Journal of Speech Language and the Law*, 11(2), 261-266. <https://doi.org/10.1558/ijssl.v11i2.261>
- NOLAN, F. (2012). Degrees of freedom in speech production: An argument for native speakers in LADO. In *International Journal of Speech, Language and the Law*, 19(2), 263–289. <https://doi.org/10.1558/ijssl.v19i2.263>
- PATRICK, P.L. (2010). Language variation and LADO. In ZWAAN, K., VERRIPS, M. & MUSKEYN, P. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedure*. Nijmegen: Wolf Legal Publishers, 73–87.
- PATRICK, P.L. (2012). Language analysis for determination of origin: Objective evidence for refugee status determination. In TIERSMA, P.M. & SOLAN, L.M. (Eds.), *The Oxford Handbook of Language and Law*. Oxford: Oxford University Press, 533–546. <https://doi.org/10.1093/oxfordhb/9780199572120.013.0039>
- PATRICK, P.L. (2016). The impact of sociolinguistics on refugee status determination. In LAWSON, R. & SAYERS, D. (Eds.), *Sociolinguistic Research: Application and Impact*. London: Routledge, 235–256.
- PATRICK, P.L. (2019). Language analysis for the determination of origin (LADO): An introduction. In PATRICK, P.L., SCHMID, M. & ZWAAN, K. (Eds.), *Language Analysis for the Determination of Origin*. Cham, Switzerland: Springer Nature, 1–17. [https://doi.org/10.1007/978-3-319-79003-9\\_6](https://doi.org/10.1007/978-3-319-79003-9_6)[https://doi.org/10.1007/978-3-319-79003-9\\_1](https://doi.org/10.1007/978-3-319-79003-9_1)
- R STUDIO TEAM. (2020). RStudio: Integrated development for R. RStudio, PBC, Boston, MA. URL <http://www.rstudio.com/>.
- WILSON, K. (2009). *Language analysis for the determination of origin: Native speakers vs. trained linguists*. University of York (UK): MSc thesis.
- WILSON, K. (2016). *Language analysis for the determination of origin (LADO): An investigative study*. University of York (UK): MPhil thesis.



SONIA CENCESCHI, CHIARA MELUZZI

## Transcription and voice comparison of noisy interceptions: remarks from an audio forensics report<sup>1</sup>

The paper describes a case study of particular interest and representative of speaker identification and speech attribution problems in real environments. The authors were recruited as technical consultants by the accused (already convicted) person's lawyer to report on a probable misidentification of her client in two environmental recordings, with the goal of obtaining a scientific analysis and subsequently requesting an eventual review of the sentence. The results, carried out by the experts blind to the overall criminal proceeding, led to a conclusion of rare clarity in this field, and to the evident presence of a clear mistake in the speaker's voice attribution. It is believed that the description of the work (albeit anonymized) may be of strong interest to those working for the promulgation of audio forensics and forensic phonetics scientific methodologies in real contexts such as preliminary investigations and legal proceedings.

*Keywords:* Forensic phonetics, Audio forensic, Noisy recordings, Speech perception, Psychoacoustics, Speaker comparison, Speaker recognition.

### 1. *Introduction*

Noisy audio has usually been disregarded by phonetic analysis, especially those dealing with the quantitative comparison of acoustic features extracted from different speech samples. To eliminate interferences, one of the first things researchers learn is to record in a silent environment with no technological devices turned on. Furthermore, audio files have to be recorded in WAVE (.wav) format with a sampling frequency of possibly 44.1 kHz and 16 bit (Di Paolo & Yaeger-Dror 2011, Meluzzi 2022). When working with forensic audio files, the situation is rather different, both in terms of audio quality (i.e., recording format, noise, and so on) and the amount of acoustic data available for a proper and reliable analysis. What happens in forensic cases is that a person is under trial on the basis of many proofs, thus including one or more audio files. These audios are usually intercepted recordings made through different technologies such as micro audio recorders or trojan, and more recently WhatsApp audio and similar kind of social network audio messages (for a full case report see Cenceschi, Meluzzi & Nese, 2020).

---

<sup>1</sup> This work has been jointly conceived and written by the two authors. However, for the Italian evaluation system, Chiara Meluzzi is responsible for section 1, 3, 3.1, 4 and 5; Sonia Cenceschi is responsible for sections 2, 2.1 and 3.2.

This difference between the ‘optimal’ audio data expected in phonetic analysis, and the forensic reality is the main cause of misunderstanding in expectations between the phonetic expert and the so-called ‘third part’, e.g., the judge, attorney or law enforcement officers asking for a forensic report. On the one hand, this implies that the experts have to be very clear in their report: not only regarding which methods they have used, but also which have to be excluded and on what scientific basis. For instance, a statistical analysis can be run only if some pre-requisites of the matrix are fulfilled. To begin with, there must be sufficient data in all the conditions to be tested: if the intercepted file is too short, sometimes less than 10”, and contains very few speech samples from the subject to be tested, it is impossible to provide more than an opinion, albeit well-grounded and justified by descriptive evaluations based on specific knowledge. On the other hand, third parties usually expect a clear Yes-or-No answer, for example, from a phonetic comparison, but this could rarely be asserted without a margin of doubt. Thus, a forensic expert needs to balance the scientific reliability with the practical needs by using a lexicon and examples understandable by naïve users (i.e., not experts). However, without tangible experience in real-world forensic circumstances, these dynamics and balance are impossible to achieve or discover.

In this paper, we will present a practical working scenario of a real-world example on which the authors have worked. For privacy reasons, the name and surname of the person involved in the trial, as well as other information that could identify him have been avoided, by referring to him as ‘Mr. G’. In the course of his trial, Mr. G was accused and convicted of a crime based on two intercepted audio files. Mr. G’s counsel wanted to know whether Mr. G. was correct in claiming his extraneousness to the facts in order to proceed with a judicial review motion. We, hereby, describe the case study in detail (section 2), including the materials, the methodological choices and the important remarks inferred and collected during the work progress. Section 3 contains analyses for both of the attorney’s requests, and the report’s findings are summarized and discussed in section 4 to highlight the issues in working with noisy audio recordings in forensic phonetics, as well as the role of forensic experts in balancing scientific reliability and practical needs. A final section focuses on the conclusion and future possibilities.

## *2. Case study: audio comparison in Mr. G.’s trial*

After the first trial was over, audio forensic experts were called in to help with the case. Since Mr. G. was asking for a judicial review, he asked his lawyer for an audio forensic report on two intercepted environmental recordings, which consisted the unique incriminating evidence. There was no additional technical information available, as it is typical in such circumstances (e.g., placing of the microphones, presence of external sources of noise). The transcription of a recording done by Law Enforcement Officers during the preliminary phases of the investigation, reported the surname of Mr. G. as pronounced by an unknown voice. A second recording

comprised a short excerpt of the speech that the police officers matched with Mr. G.'s speech timbre. On these basis, Mr. G. was prosecuted and then arrested, being later sentenced to 18 months in prison. He spent these months at house arrests, by always claiming his innocence. Only after 10 months, the Mr. G.'s lawyer obtained these audio recording and asked for an expertise report.

In June 2018, the two forensic specialists were called and asked to provide an audio forensic report for two main purposes, that is to determine:

1. whether the first environmental recording contained the surname of Mr. G.;
2. whether the second environmental recording contained the Mr. G.' speech.

The first point is concerned with perceptual, spectral, and phonetic observations, whereas the second is a classic comparison between a Known Voice (KV), that is, the client's voice as recorded by the forensic experts, and an Unknown Voice (UV), as reported by Law Enforcement Officers and identified as the client's voice.

## 2.1 Materials and data

Mr. G.'s lawyer provided the audio recordings to the forensic experts for their report. Table 1 summarizes the data, by dividing the audio belonging to the Unknown Voice (UV) and those pertaining to the voice of the client (Known Voice, KV). The table also reports an indication of the duration of the original audio file, its original saving format, and the typology of recording (i.e., environmental interception, phone call, etc.). No information was available concerning the instruments used in the interception.

Table 1 - *The data available for the report, as provided by Mr. G's lawyer*

<i>Typology</i>	<i>Total duration</i>	<i>Format</i>
<i>Environmental interception With the Unknown Voice (UV)</i>	19'06"	.ogg
<i>Environmental interception containing the surname</i>	19'40"	.ogg
<i>Phone Call, spontaneous speech (Known Voice)</i>	5'37"	.m4a
<i>Whats.App, spontaneous speech (Known Voice)</i>	1'49"	.m4a
<i>Whats.App, spontaneous speech (Known Voice)</i>	42"	.m4a
<i>Whats.App, read speech (Known Voice)</i>	1'	.m4a

The presence of more data for the KV is normal during a phonetic forensic comparison, especially if the KV belongs to the lawyer's client and/or a person under trial. In these cases, the forensic expert could pretend to record more audio samples for the comparison by performing the so-called *saggio fonico*, that is a collected audio sample from speaker comparison (cf. Kersta, 1962). Obviously, what determines the possibility of effectively performing an audio comparison is not the amount of data

at disposal for the KV, but what we have for the UV. In the case presented here, the only audio files belonging to UV to be compared with Mr. G. was an environmental interception recorded in a public domain compressed audio format (.ogg). Despite its length, the audio was very noisy and different voices overlapped with each other. This means that the amount of speech effectively produced by UV without overlapping was very poor both from a qualitative and quantitative point of view.

The \*.ogg files were in stereo format, 16,000 Hz, 16 Bit, and had a useful upper spectral threshold of 8,000 Hz. They were converted to wav format (dummy enhancement) for software processing convenience. Although they were of inferior quality, they had a threshold SNR (signal to noise ratio) sufficient to perform spectral and phonetic analysis with a semi-automatic method (SNR > 60 dB). The materials belonging to KV consist in:

1. a business call from Mr. G;
2. three vocal messages recorded on WhatsApp<sup>2</sup> by Mr. G and sent to the experts. The audio files were recorded with a Huawei P30 PRO, with a sampling frequency of 48 kHz, 16 bit. Two of them contained spontaneous speech, and one WhatsApp audio message for which Mr. G. was asked to read the same linguistic content pronounced by the UV in the environmental recording.

### 3. *Analysis of Mr. G's case*

A forensic report must be clearly organized and convey the results in the most exact but simple manner possible in order to improve the understanding of specialized aspects by so-called “third parts” (i.e., judges, lawyers, etc.). The necessity to maintain a dialogue between the expert and not-expert comes with the requirement to explain in detail the scientific jargon and the different tools normally used in phonetic analysis. It was important, for example, to explain what a spectrogram was, how PRAAT was used, and to provide a brief summary of the results at the beginning. For clarity, results were here organized in the same manner: A first section (3.1) is dedicated to determining the presence or absence of the client's surname in the audio files, and a second section (3.2) is dedicated to demonstrating whether KV and UV could conceivably belong to the same speaker (i.e., Mr. G.).

---

<sup>2</sup> WhatsApp is based on the SILK VoIP (Voice over Internet Protocol) protocol, a codec developed by Skype and now license-free, available in open-source mode. This technology allows you to have a conversation similar to that of a telephone network by using an Internet connection or any other dedicated packet switching telecommunications network that uses the IP protocol for data transport. SIL VoIP is the basis, with CELT, of the hybrid Opus codec, the official format of WhatsApp, which is a lossy audio coding format used to achieve low latency with the best quality. In particular, WhatsApp exports Opus files as m4a (with the AAC codec) because Opus is not supported by many audio player applications.

### 3.1 The presence/absence of the surname

In order to preserve the anonymity of the client, in this section we limit to describe the procedures adopted in the report, without providing the examples included in the report to demonstrate the presence of a different surname than the client's one.

Following pre-processing, the first stage was the extraction of the portion of interest from the original \*.ogg file, which was already been documented as part of the ongoing legal procedure. The portion of interest corresponded to 13 seconds (from 5' to 5'13" of the original file). It was saved as a separate file called "Extracted V1". The file was then converted in .wav format to allow formants' viewing in PRAAT<sup>3</sup>. The new file was transcribed and time-aligned with the help of ELAN software (Lausberg & Sloetjes, 2009) in order to annotate different linguistic and extra-linguistic features on multiple tiers. Only the transcription tier was then exported in a Word file to be attached to the final report. We reported below in Example 1 an anonymized transcription of this file: proper names and surnames have been substituted as XXX and YYY, respectively. The amount of Xs and Ys correspond to the original phones heard during transcription. Voice 1 and Voice 2 indicates the two (male) speakers. The English translation for each line of dialogue is provided below; this was not part of the original report because it was unnecessary.

*Example 1: Anonymized transcription of Extracted V1*

- 1 Voce 1: c'è XXXX, c'è... ..YY(Y) [non chiaro, voci sovrapposte]  
*Voice 1: there's XXXX, there's... ..YY(Y) [unclear, overlapping voices]*
- 2 Voce 2: XXXX?  
*Voice 2: XXXX?*
- 3 Voce 2: chiama(lo) [non chiaro, voci sovrapposte]  
*Voice 2: call him [unclear, overlapping voices]*
- 4 Voce 1: c'è XXXXXXXXXXXX  
*Voice 1: there's XXXXXXXXXXXX*
- 5 Voce 2: YYYYY?  
*Voice 2: YYYYY?*
- 6 Voce 1: sì.  
*Voice 2: Yes.*

According to the excerpt, we were dealing with a very short dialogue between two people, the speakers of which were lately identified based on another audio file provided by the lawyer. The audio quality was very low, thus in *line 1 and 3*, it was specified that the content was unclear and that the two voices were overlapping too much to allow a correct transcription. Although this is not a proper scientific way to indicate overlapping or noise in a transcription (see, for instance, the guideline for transcription in Conversation Analysis; cf., for Italian, Savy 2005), it was necessary

<sup>3</sup> PRAAT now allows also .mp3 files, but at the time of the case it was not possible, and it was thus decided to convert the .mp3 to .wav only for visualization purposes. The authors discourage from extracting acoustic values from originally compressed audio files.

to provide a transcription as clear as possible to the lawyer. Thus, the use of symbols had to be limited or even avoided to enhance the communication with the third part.

For the purpose of the report, we focused on *line 5*, in which the second speaker pronounced the surname of the person that the prosecutor has identified with the client. The transcription already pointed out that this surname was constituted by 2 syllables and 5 phones, with only a small doubt on the third one, which could alternatively be /l/ or /r/. An examination of the spectrogram using PRAAT validated this transcription (Boersma, 2001).

In the report, pictures of the wave form and spectrogram as visualized in PRAAT were included. Each phone was delimited by vertical yellow bars and clearly indicated by an arrow. The explanation of the acoustic nature of each sound was also provided in the text, with a long note containing the major references to the topic.

In the end, it was possible to argue without doubt that the surname pronounced in line 5 of the excerpt (1) clearly show the presence of only two vowels on the spectrogram, thus corresponding to a two-syllables word. The phones constituting the two syllables could also be easily identified, with the only exception of a sound at the end of the first syllable that looked like an [l], but that could also be a rhotic realized as an approximant (cf. Celata et al. 2016 for spectrographic characteristics of Italian rhotic realizations). No fricatives or palatal sounds were detected on the spectrogram.

Since Mr. G.'s surname was four-syllables long, and it contained a fricative and a palatal nasal, it was possible to sustain the claim that in line 5 of the excerpt, the second speaker did not pronounce the client's surname. For this reason, the report claimed that the original transcript was incorrect. The hypothesis was also floated that Mr. G.'s surname was mistakenly added to the original transcript due to a perceptual error. Indeed, police officers frequently do transcriptions and recordings without the correct audio equipment, which can easily lead to blunders such as the one in this case. In this regard, it should be highlighted that the two specialists first examined the spectrograms and then listened to the audio to avoid being swayed by their knowledge of Mr. G.'s surname. Even so, the listening of the audio with closed headphones clearly confirmed the result achieved through the spectrographic analysis.

### 3.2 Speakers' voices comparison

Since the two audio recordings to be compared were both severely damaged in terms of quality and were also relatively brief, it was determined that no semi-automatic speech comparison could be performed. Nevertheless, we aimed at providing a reliable report by comparing the two voice across three different methods:

1. Perceptive test
2. Phonetic and voice quality analysis
3. Sociophonetic profile of the two voices

The perceptive test was conducted on three distinct categories of listeners: a group of 6 trained Italian phoneticians, experts in audio and/or forensic analysis; 6 foreigner phonetic experts, with different L1s and only a superficial knowledge of Italian; a group of 43 naïve listeners (24 women, 19 men), aged between 23 and 64 years, and



native speakers of North-Western Italy. The reason for testing this three different groups was twofold: on the one hand, we needed the opinion of trained experts who had previously worked with noised audio files for forensic purposes, but we wanted to see if the noised audio file's content could alter the experts' impression (e.g., Fraser 2003). On the other hand, we sought to compare the experts' results to the perception of naive speakers, to see if knowledge of the task and its repercussions affected the judgment on the similarities between the two voices.

Each group had to listen to two stimuli previously prepared by the researchers for maximum two times. After the listening, participants had to indicate if they believe the two voices in the stimulus belonged to the same person, the degree of confidence of their judgment and to eventually indicate what has shaped their judgment. Both stimuli contained two audio excerpts of around 10" each, with an interval of 3" of silence between them. Stimulus A contained the UV as extracted from the materials provided by the lawyer and a sample of KV as extracted from a phone call made by the client to his lawyer. Stimulus B contained the same two files but in reversed order (KV followed by UV). Each respondents started the test alternatively with stimulus A or B, in a random order. All groups of listeners agreed that KV and UV belonged to two separate people. Only 9 naive listeners out of 43 total respondents claimed that KV and UV belonged to the same individual, even though they mentioned in the open remarks that the accent and 'speech mode' were different. Comments from naïve listeners' provided useful information on the phonetic and phonological characteristics that have shaped their perception: possible geographic origin, as carried by regional accent, and speech rate were the most common features, together with the pronunciation of the vowel /e/.

A detailed phonetic analysis was provided by extracting the first formants' values of all vowels produced in the recordings containing both KV and UV. Table 2 shows the amount of vowels considered for each voice. Although usually only stressed vowels are considered in phonetic analysis, in this case (as well as in other forensic cases) the scarcity of the audio material forced the experts to use all available vowels, regardless of the stress position. Only diphthongs and vowels included in ruined audio segments were excluded from the analysis. Vowel /ɛ/ was also excluded because of the very rare occurrences in the data.

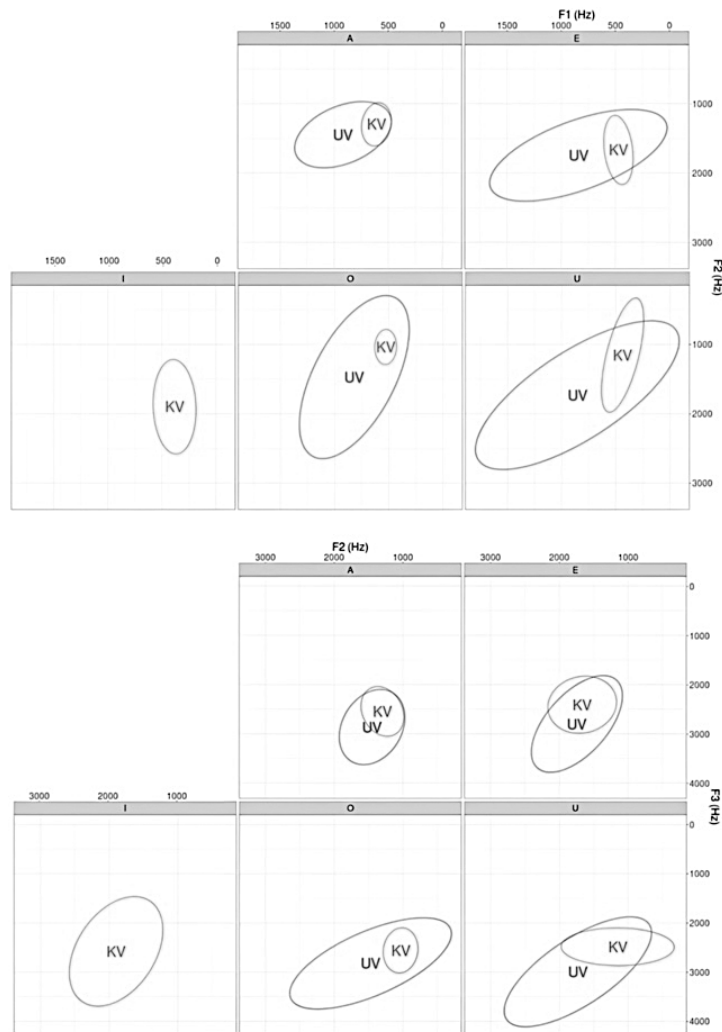
Table 2 - *Vowels available for phonetic comparison of KV and UV*

	<i>Unknown Voice (UV)</i>	<i>Known Voice (KV)</i>
/a/	67	60
/e/	84	64
/o/	17	43
/u/	33	15
/i/	0	32

The data were unbalanced, as it usually happens in forensic research, and many items had to be discarded because of the low quality of the audio file. We focused the analysis

on F1, F2 and F3 values, since F0 was too variable, and it could also be affected by the different registers and speech styles of the two recordings. Then, we plotted the values through Visible Vowels (Heeringa and Van de Velde, 2017) to enhance the clarity of the results for the third part, since graphical representation helps the understanding of reports. We plotted the vowel space for F1xF2 (Fig. 1 above) and for F2xF3 (Fig. 1 below) whose importance was suggested by the Dispersion-Focalization Theory (DFT) of vowel systems (Schwartz et al. 1997; 2012).

Fig. 1 - Comparison between Unknown Voice (UV) and Known Voice (KV) as performed through Visible Vowels. Above: the vowel space for F1xF2. Below: the vowel space for F2xF3. Vowel labels: "A" /a/, "E" /e/, "O" /o/, "I" /i/, "U" /u/.



As evident from Fig. 1, although no instances of /i/ were available for UV, it was still possible to appreciate how the vowels were differently distributed, in particular for what concerned the F1-F2 space. In particular, the distribution of KV's formants is extremely well delimited with respect to that of UV (and more concerning F1). The same is true in the F2-F3 region, where the distribution of KV's formants is also extremely well circumscribed in comparison to that of UV for vowels, particularly F3. Both of these graphs provide evidence that contradicts the possibility of UV and KV belonging to the same person.

On a qualitative level, KV was perceived with a more nasal realization of vowels, especially /a/ and /e/. Although this is not strongly reflected in F2/F3 variability, the Harmonics-to-Noise Ratio (HNR) showed a clear difference between UV and KV, as reported in Table 3.

Table 3 - *Harmonics-to-Noise Ratio in Known and Unknown voice*

	<i>Known Voice (KV)</i>	<i>Unknown Voice (KV)</i>
/a/	7.017 dB	4.897 dB
/e/	7.336 dB	4.977 dB
/o/	7.524 dB	6.039 dB

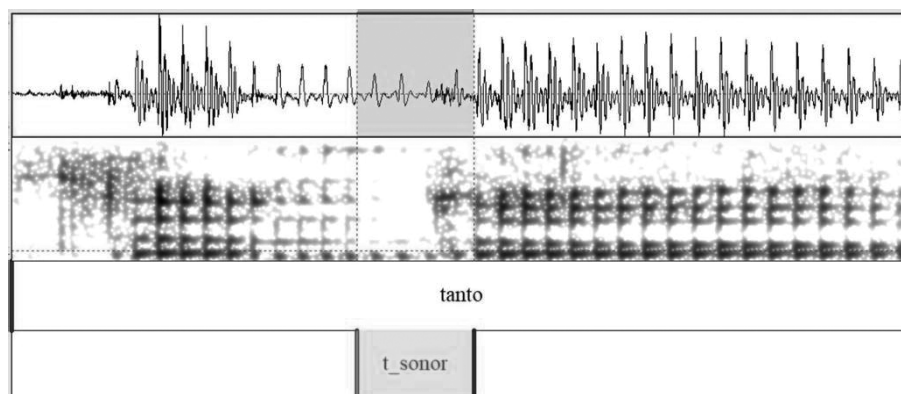
Finally, we presented in the report a sociophonetic profiling of both voices. To avoid biasing the subsequent results, we did this analysis while running the perceptive test and before the pseudo-quantitative formant analysis. Sociophonetic profiling considers the variability of language according to different geographical and social factors, other than the topic and setting of the recordings. Profiling should be accomplished by a forensic audio expert who is also very knowledgeable about the language of the case from a broad linguistic standpoint, taking into account not only phonetics but also dialectology and sociolinguistic variability. This profile operation is typically performed prior to any further perceptual or quantitative testing on the materials, so reducing the expert's potential biases (see also Meluzzi et al., 2020). For this case, giving the aforementioned limitations of the audio materials, the sociophonetic profiling was carried out for both UV and KV on the basis of prosodic cues and the realization of middle vowels /e/-/ɛ/ in stressed position.

As said before, the scarce occurrences of /e/ did not allow for a quantitative comparison, thus we opted for only a perceptive comparison. The two voices were evaluated independently and the findings were compared. UV pronounced [e] with the stressed vowels in the words *proiettili* 'bullets' and *sette* 'seven' of the first audio file, and in the words *merda* 'shit' and *questo* 'this' in the second audio file. From the same audios, it was possible to notice a 'dark' realization of the central vowel /a/ when in stressed position like in the words *puttana* 'bitch' and *cazzo* 'fuck'. This was consistent with the further formant analysis. The prosody corresponds to a Northern Italian variety (Hack, 2012; Cardinaletti & Repetti, 2008), with stereotypical Milanese characteristics (Montreuil, 1991; Cerruti, 2011; Kramer, 2009). Furthermore, based on the message of the intercepted audio, it was possible

to detect that, despite the low register, the speech is articulated in subordinated clues, with no mispronunciations or disfluencies, and no indication of foreigner accent. To sum up, UV could be profiled as an adult male between 35 and 40 years old, with a middle to high level of education; more importantly, the speaker's origins could be placed in Lombardy, especially in the Milanese area.

The profile of the KV was more complicated. The expert should have no prior knowledge of the subject being profiled, but even in this scenario, some preliminary information could skew the study. As a result, the profile must be based solely on objective qualities of the voice as revealed by the audio at hand. Since KV has repeated the same sentences of the intercepted audio containing the UV, it was possible to compare the exact same words across the two recordings. The results showed that KV realized the stressed anterior middle-vowel in closed syllables as [e] (e.g., in the word *sette* 'seven'), whereas the vowel /a/ was realized as more fronted than UV. Furthermore, in a voice note, KV voiced the alveolar consonant following the nasal in the word *tanto* 'a lot,' producing [tan.do], as evidenced by the spectrogram (cf. Fig. 2).

Fig. 2 - The spectrogram of KV's vocal note showing the sonorization of the alveolar consonant in the word *tanto* 'a lot'



These phonetic features point towards a Southern Italian regional pattern. These characteristics were also confirmed by the intonation patterns shown by KV in the spontaneous speech recorded in the phone call. However, since these features were not extremely marked in KV's speech, the hypothesis was made that KV came from Southern Italy, and, hypothetically, from Sicily, and that he maintained relationships in this area, thus justifying the presence of some phonetic and phonological pattern of southern regional varieties of Italian. KV was assumed to be an adult male between the ages of 30 and 40 based on more impressionistic qualities of his voice quality. It was worth noting that the sociophonetic profile found a match in the comments lately made by both Italian experts and naïve speakers on the two voices during the perceptive experiment. This indicates the perceptual importance of even minor phonetic characteristics in conveying sociolinguistic meaning. Furthermore,

the lawyer later confirmed, after the reception of the report, that her client (i.e., KV) was 33 years old and lived in the Milanese area, but he was originally from the area of Agrigento (Sicily), where he regularly spent his summers.

#### 4. *Results and discussion*

Results can be summarized as follows: Mr. G.'s surname was not pronounced in the first environmental recording, and UV and KV are not the same speaker.

It is clear that such a categorical result is extremely rare. It should be noted that the information provided to us after the report was completed supported our view. Indeed the lawyer informed us that the attribution of the vocal timbre to Mr. G. had been conducted on the basis of the surname heard pronounced by another guy in the initial tape. Without calling into question the good faith and investigative preparation of those who carried out the work, we would like to emphasize how this can thus be defined as a classic case of altered perception: the speaker was most likely inferred from a clue previously extrapolated without scientific validation, e.g. "since Mr. G. has been previously cited, he must be the person speaking on the other recording". Furthermore, our understanding of the technical and operational conditions under which many agents operate substantiates the work. Many consecutive hours with headphones, without being able to equalize the audio or intervene on the signal to improve its intelligibility further affects the speaker's understanding and attribution to reinforce the concept that knowledge of the facts is not alone a guarantee of infallibility in this field (Fraser, 2003; French, Fraser, 2018), especially when data are so few and poor from the qualitative point of view.

William Labov once said that doing linguistic research is often a matter of making the best possible use of poor data (Labov, 1972). This aphorism was written with historical sociolinguistics in mind, but it can be applied to any situation in which a lack of data and the objective impossibility of collecting new ones forces the researcher to adapt his paradigms and standards in order to always provide reliable and informative research. In audio forensics, the data at disposal for a phonetic comparison are usually very 'poor' both in quantitative and qualitative terms. As seen in this work case, it is usual that speech data belonging to the UV are very scarce, whereas it is possible to collect many more audio samples from the KV. Furthermore, the quality of the audio, especially in case of UV, is usually very scarce too, in terms of background noise and spectral information. It is thus beneficial to think in a non-mechanical manner, drawing from all available clues to generate conclusions that can be valuable in the forensic profession. Indeed, deductions and logical consequences might lead to perfect conditions for addressing a forensic inquiry, as in this example.

### 5. *Conclusions and further perspectives*

The reason it is necessary to make public work like this is obviously mainly an ethical principle. Justice and trust in it are undermined by innumerable factors widely discussed in the audio forensics and applied forensic linguistics literature (e.g., Maher, 2009; Fraser, 2021; Cenceschi, Meluzzi & Trivilini, 2021). Furthermore, experience in real contexts suggests that highlighting problems and viable solutions with concrete examples of real case works is the best way to fuel a healthy collaboration between the scientific and the judicial worlds, even when the difficulties venture beyond the technical themes, into the complicated sphere of bureaucracy. A second, but no less significant, motivation for this publication is the necessity to warn researchers interested in forensic competence that, in judicial contexts, it is often necessary to virtually forget one's scientific language in order to speak effectively in the cause of truth and justice. It is also necessary, for example, to be able to give up the possibility of completing an analysis, and to build deductions 'out of the box' on the basis of a small number of the available elements, without forgetting the scientific nature of the work.

From a scientific point of view, the forensic data is, as mentioned, unpredictable and often of low quality and duration (Meluzzi, Cenceschi & Trivilini, 2020). This case is the demonstration that these characteristics do not completely wipe out the possibilities for intervention and the work can always lead to useful conclusions. It all relies on who, how, and even when the analysis are performed. The expert can intervene at different moments of the judicial process, and often, due to the structure of the judicial process and the bureaucracy, there is no guarantee that the expertise will be, regardless of its goodness, exploited or considered. However, the expert has neither duties nor powers in this sense, his role is to carry out the analysis and express the objective results in the best feasible way. Again, making scientific terminology comprehensible is critical in order to avoid stalling the justice procedure, especially when examining documents already examined by "ad hoc experts" (French & Fraser, 2018).

### *Bibliography*

- BOERSMA, P. (2001). Praat, a system for doing phonetics by computer. In *Glott International*, 5(9/10), 341-345.
- CARDINALETTI, A. & REPETTI, L. (2008). The phonology and syntax of preverbal and postverbal subject clitics in northern Italian dialects. In *Linguistic inquiry*, 39(4), 523-563.
- CELATA C., MELUZZI C. & RICCI I. (2016) The sociophonetics of rhotic variation in Sicilian dialects and Sicilian Italian: corpus, methodology and first results. In *Loquens*, 3(1), e025. DOI: <http://dx.doi.org/10.3989/loquens.2016.025>
- CENCESCHI, S., MELUZZI, C. & NESE N. (2020). Speaker's identification across recording modalities: a preliminary phonetic experiment. In ROMITO, L. (ed.) *Language change under contact conditions*, Studi AISV 7, Milano: Officinaventuno, 409-428. DOI: 10.17469/O2107AISV000019

- CENCESCHI, S., MELUZZI, C. & TRIVILINI, A. (2021). Audio compression and speaker's discrimination: perspectives for forensic phonetics in the Italian setting. In *Indagatio Didactica*, 13(5), 143-154.
- CERRUTI, M. (2011). Regional varieties of Italian in the linguistic repertoire. In *International Journal of the Sociology of Language*, 210, 9–28.
- DI PAOLO, M. & YAEGER-DROR, M. (2011). *Sociophonetics. A Student's Guide*. London: Routledge.
- FRASER, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. In *Forensic Linguistics*, 10, 203-226.
- FRASER, H. (2021). Forensic Transcription: Legal and scientific perspectives. In BERNARDASCI C., DIPINO D., GARASSINO D., NEGRINELLI S., PELLEGRINO E., SCHMID S. (eds.) *Speaker individuality in phonetics and speech sciences: speech technology and forensic applications*, Studi AISV 8, Milano: Officinaventuno, 19-32.
- FRENCH, P. & FRASER, H. (2018). Why “Ad Hoc Experts” should not Provide Transcripts of Indistinct Audio, and a Better Approach. In *Criminal Law Journal*, 42, 298-302.
- HACK, F.M. (2012). *The syntax and prosody of interrogatives: Evidence from varieties spoken in northern Italy*. Oxford University (UK): Doctoral dissertation.
- HEERINGA, W. & VAN DE VELDE, H. (2017). Visible Vowels: A Tool for the Visualization of Vowel Variation. In *Proceedings CLARIN Annual Conference*, Clarin Eric, 4034-4035.
- KERSTA, L.G. (1962). Voiceprint identification. In *The Journal of the Acoustical Society of America*, 34(5), 725-725.
- KRAMER, M. (2009). *The phonology of Italian*. Oxford: Oxford University Press.
- LABOV, W. (1972). *Sociolinguistic patterns*, Pennsylvania: University of Pennsylvania press.
- LAUSBERG, H. & SLOETJES, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. In *Behavior Research Methods, Instruments, & Computers*, 41(3), 841-849. doi:10.3758/BRM.41.3.841.
- MAHER, R.C. (2009). Audio forensic examination. In *IEEE Signal Processing Magazine*, 26(2), 84-94.
- MELUZZI, C. (2022) La ricerca sul campo (e in campo). In MELUZZI C., & NESE N. (eds.) *Metodi e prospettive della ricerca linguistica*. Milano: Ledizioni, 37-52.
- MELUZZI, C., CENCESCHI, S. & TRIVILINI, A. (2020). Data in forensic phonetics from theory to practice. In *Teanga*, 27, 65-78.
- MONTREUIL, J.P. (1991). Length in Milanese. In *New analyses in Romance linguistics*, 37-47.
- SAVY, R. (2005). Specifiche per la trascrizione ortografica annotata dei testi, in ALBANO LEONI, F. (a cura di) *Italiano Parlato. Analisi di un dialogo*. Napoli: Liguori editore.
- SCHWARTZ, J.L., BOË, L.J., VALLÉE, N. & ABRY, C. (1997). The dispersion-focalization theory of vowel systems. In *Journal of phonetics*, 25(3), 255-286.
- SCHWARTZ, J.L., BOË, L.J., BADIN, P. & SAWALLIS, T.R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. In *Journal of Phonetics*, 40(1), 20-36.







PARTE 2

LA LINGUISTICA FORENSE  
NELLA PRATICA LEGALE

FORENSIC LINGUISTICS  
IN THE LEGAL PRACTICE





FRANCA ORLETTI, IACOPO BENEVIERI

La trascrizione forense tra cattive prassi e fallacie probatorie.  
Analisi di una trascrizione di intercettazione ambientale<sup>1</sup>  
Forensic transcription between bad practices and evidentiary fallacies.  
Analysis of an environmental wiretapping transcript

The paper arises from the scientific interaction of two different skills: that of a criminalist and that of a linguist, based on the awareness of the inseparable link between language and law. The data analysed consists of the transcription of an environmental interception of a conversation between two suspects, a woman and a man, not fully identified, both born and resident in the province of Rome. The aim of the work was to investigate how the lack of knowledge on the part of the judicial police officer of the main guidelines on forensic transcriptions, drawn up by the scientific community of reference, can lead to significant discrepancies between the intercepted speech and its transposition into written form.

*Keywords:* forensic transcription, scientific transcription, covert recordings, evidence, fair trial.

### 1. Introduzione

Il presente lavoro ha come obiettivo quello di evidenziare la rilevanza dei dettagli linguistici, in senso lato, di una interazione comunicativa tra soggetti intercettati, ai fini probatori e processuali e nasce dalla collaborazione, da un lavoro di sinergia tra una sociolinguista attenta alla linguistica forense e un avvocato penalista anch'egli attento a questa disciplina, entrambi consapevoli dell'inscindibile legame tra lingua e diritto. Questo legame è stato sottolineato da un'ampia tradizione di studi in ambito di linguistica forense. Si ricordi, per esempio, la disciplina della fonetica forense avviata dagli importanti lavori di John Bassett Trumper e, sempre in ambito italiano, l'opera di Patrizia Bellucci, *A onor del vero*, nonché l'approfondita introduzione di De Mauro a tale lavoro (v. Bellucci, 2005; De Mauro, 2005).

Nell'affrontare l'analisi del dato oggetto di questo studio, confronteremo le caratteristiche della cosiddetta trascrizione scientifica, quella che nasce come strumento di studio e di ricerca sul parlato, con quelle proprie della trascrizione forense.

---

<sup>1</sup> Il presente lavoro è stato ideato da entrambi gli Autori e l'analisi è frutto di un'elaborazione congiunta. Ai fini accademici sono da attribuire a Orletti i parr. 1, 4, 6, e 8 e a Benevieri i parr. 2; 3; 4; 5. Il paragrafo 7 va attribuito a entrambi.

## 2. *Descrizione del dato oggetto dell'analisi*

Oggetto dell'analisi è la trascrizione di una intercettazione ambientale, che si può in un certo senso considerare canonica della più diffusa prassi trascrittoria giudiziaria perché presenta caratteri che ricorrono frequentemente nelle trascrizioni di intercettazioni effettuate dalla polizia giudiziaria. È caratterizzata da *omissis*: larghe fasi dell'interazione risultano omesse nella trascrizione. È semplificata: cioè non tiene conto delle peculiarità del parlato, come la prosodia, i silenzi, le pause, le sovrapposizioni dei turni, le false partenze. Presenta infine interpretazioni del parlato da parte del trascrittore<sup>2</sup>.

Il dato è costituito da una conversazione captata mediante intercettazione ambientale. Nello specifico, si tratta di una conversazione tra due soggetti, un uomo indagato e una donna di cui non si conosce l'identità. L'intercettazione è avvenuta nella provincia di Roma e la trascrizione è stata effettuata dalla polizia giudiziaria.

## 3. *Le intercettazioni: inquadramento giuridico e giurisprudenziale*

Le intercettazioni rappresentano uno strumento processuale che consente l'acquisizione di elementi di prova, costituiti da conversazioni o comunicazioni telefoniche, ambientali e flussi informatici.

Sotto un profilo tecnico con il termine "intercettazione" si intende il processo attraverso cui, grazie a dispositivi elettronici, conversazioni o comunicazioni, che si svolgono a distanza mediante telefono o altro mezzo ovvero hanno luogo in un determinato contesto, vengono captate a insaputa degli interlocutori da parte di un terzo che non è né parte dello scambio comunicativo, né destinatario delle comunicazioni (Paoloni, Zavattaro, 2007: 81). Si tratta pertanto di una captazione occulta di una comunicazione o conversazione tra due o più soggetti, ovvero di un flusso di dati, attuata a opera di un terzo con l'impiego di mezzi meccanici, elettronici o informatici (Cassazione, S.U., 28.5.2003, Torcasio e altro).

Tradizionalmente le intercettazioni, in ragione delle caratteristiche qualitative della registrazione e delle modalità con le quali vengono attuate, possono essere distinte in intercettazioni di telecomunicazioni, quelle cioè effettuate per mezzo della rete fissa o rete mobile, previste dall'art. 266 comma 1 c.p.p.; intercettazioni tra presenti, cd. "ambientali", quelle effettuate in un ambiente qualsiasi sottoposto a controllo attraverso rete telefonica fissa, rete mobile o con l'ausilio di mezzi diversi da quelli telefonici (registrazioni a distanza con microfoni direzionali, microtrasmettitori occultati materialmente con trasmissione attraverso ponti radio, microregistratori nascosti ecc.), disciplinate dall'art. 266 comma 2 c.p.p.; intercettazioni informatiche e telematiche, attuate mediante il captatore informatico (c.d. "trojan horse"), contemplate dall'art. 266bis c.p.p.

<sup>2</sup> Per un'analisi delle caratteristiche linguistiche delle trascrizioni delle intercettazioni si rinvia a Orletti, 2016, 2017; Duranti, 2006.

Le intercettazioni costituiscono uno strumento di ricerca della prova molto importante, cui si ricorre in modo consistente nel corso delle indagini preliminari qualora sia consentito dal titolo di reato.

Per offrire un dato rappresentativo della rilevanza delle intercettazioni nei procedimenti penali in Italia, basti considerare come nel 2020, ultimo anno in relazione al quale sono stati diffusi dati ufficiali, siano state eseguite circa 106.513 intercettazioni, delle quali il 78% di tipo telefonico (pari a 83.454), il 15% di tipo ambientale (pari a 15.427) e il 7% di altro tipo (Ministero della Giustizia, 2020).

Occuparci di intercettazioni significa pertanto occuparci di un mezzo di ricerca della prova, che, per la sua rilevanza, merita di essere studiato sia nella prospettiva giuridica che in quella linguistica.

#### *4. La trascrizione scientifica e la trascrizione forense*

Con gli studi sul rapporto scritto/parlato, condotti sia nell'ambito dell'analisi del discorso, che in ambito sociolinguistico, che in quello dell'analisi conversazionale, è stato avviato un indirizzo di ricerca specifico dedicato alla trascrizione.

La trascrizione nata dalla ricerca sul parlato (cd. trascrizione scientifica) presenta caratteristiche peculiari che la distinguono dalla trascrizione forense.

La riflessione sulle caratteristiche della trascrizione scientifica nasce dall'importante lavoro di Elinor Ochs (1979), che evidenzia come il trascrivere sia una teoria, non consista cioè in una banale traduzione da un suono a un segno scritto, ma al contrario sia un'attività caratterizzata da una importante componente teorica e costituisca un primo livello di analisi del parlato, implicando selezione e interpretazione da parte del trascrittore.

Inoltre nella trascrizione scientifica, proprio in ragione della selettività del testo trascritto, quest'ultimo non costituisce il dato primario, che è rappresentato invece dall'evento fonico a cui il trascrittore deve continuamente tornare a confrontarsi nella propria attività di analisi. In tale analisi il trascrittore, infatti, deve sempre tenere conto della complessità del parlato, caratterizzato dalla multimodalità: in particolare è tenuto a prendere in esame l'importante ruolo della prosodia, i fenomeni paraverbali, i cosiddetti fenomeni temporali (le sovrapposizioni), i cambiamenti di progetto, le false partenze. La complessità di tali fenomeni deve essere necessariamente riportata nella trascrizione (Orletti, Testa, 1991).

Alla luce degli studi condotti nell'ambito dell'analisi conversazionale, inoltre, la trascrizione scientifica considera il parlato come un fenomeno strutturato e razionale, caratterizzato da una concatenazione sequenziale di turni secondo cui il turno che viene prima si lega al turno successivo<sup>3</sup>.

<sup>3</sup> Sacks, Schegloff & Jefferson (1974). La trascrizione di porzioni di parlato vista in rapporto alla relazione parlato/scritto è diventata un vero e proprio filone di ricerca, a cui hanno contribuito studiosi all'interno dell'analisi del discorso, come Du Bois (1991) e, per l'analisi della conversazione, Jefferson, 2004. Per una visione d'insieme dei vari sistemi notazionali, oltre al già citato Orletti, Testa (1991), si veda anche O'Connell, Kowal (2009).

La trascrizione forense invece mira unicamente a rendere fruibile il contenuto del dato fonico, chiarendo di volta in volta chi parla e cosa dice. Proprio tale finalità induce spesso il trascrittore ad andare al di là del dato fonico, con la conseguenza che la trascrizione costituisce spesso una interpretazione di tale dato.

### 5. *La trascrizione forense secondo l'ordinamento giuridico*

Secondo il nostro ordinamento giuridico la prova è costituita dal segnale fonico acquisito tramite l'attività di intercettazione, mentre la trascrizione è unicamente una "rappresentazione" in veste grafica di tale dato. Secondo un orientamento ormai consolidato della Suprema Corte di Cassazione, l'attività di trascrizione costituirebbe una mera "*attività materiale e meccanica, non valutativa*", che rende il parlato fruibile nel processo tramite la consultazione della relativa trascrizione<sup>4</sup>. Nella prassi giudiziaria, tuttavia, sia il giudice che le parti processuali molto spesso si basano quasi esclusivamente sul dato trascritto, a causa sia delle difficoltà operative di accedere al dato fonico originario (il cui ascolto deve essere comunque autorizzato dall'autorità giudiziaria a seguito di idonea richiesta giustificata da parte del difensore), sia dell'elevato numero di intercettazioni sovente condotte nell'ambito di un medesimo procedimento, tale da rendere estremamente complesso e faticoso l'ascolto dei file audio registrati.

A tali considerazioni deve aggiungersi come la recente riforma sulle intercettazioni telefoniche (D.L. n. 161/19, convertito in Legge n., 7/20) abbia previsto la possibilità di attribuire pieno valore di prova anche alla trascrizione effettuata dalla polizia giudiziaria, laddove le parti prestino il consenso a tale utilizzo. L'art. 268 comma 7 c.p.p., modificato dalla citata legge, stabilisce infatti che "*il giudice, con il consenso delle parti, può disporre l'utilizzazione delle trascrizioni delle registrazioni [...] effettuate dalla polizia giudiziaria*".

Così brevemente delineato il quadro giuridico e giurisprudenziale, è evidente come la trascrizione delle intercettazioni non solo sia un'attività molto complessa, in quanto implica il passaggio tra due codici semiotici diversi, ma possa costituire altresì la prova sulla quale il giudice forma il proprio convincimento e motiva la sentenza. È indispensabile pertanto che l'attività trascrittiva sia effettuata in modo adeguato, proprio allo scopo di evitare che venga alterata la rappresentazione scritta del parlato intercettato. In tale prospettiva, dunque, ogni questione linguistica è questione probatoria: qualsiasi errore, ambiguità o trascuratezza nella trascrizione del dato fonico può determinare rilevanti conseguenze sull'accertamento del fatto e dunque sulla corretta amministrazione della giustizia.

<sup>4</sup> Cfr., tra le ultime sentenze, Cass., Sez. V Pen., 17.2.2020, n. 12737.

### 6. *Analisi del dato*

Il dato, così come è stato tratto dagli atti processuali, è costituito dalla seguente trascrizione in (1).

(1) *...omissis dalle ore 12.07.54 alle ore 12.08.13...*

1. Uomo: – aho... ti stai a muovere?
  2. Donna: – sì!...a che ora vieni?...mi chiami tu quando vai a fare colazione?
  3. Uomo: – eh!...già te l'ho messa da parte quella cosa, eh!...
  4. Donna: – bene, bene!
  5. Uomo: – così però a' Mo' (Morena nome di Donna)
  6. Donna: – beh, pure...pure di meno!
  7. Uomo: – eh va be però ti ho messo quella a' Mo'!
  8. Donna: – e...se non ti do quello della settimana, altri dieci
  9. Uomo: – come ti pare a' Mo', quella ti ho messo da parte!
  10. Donna: – e siccome <<incomprensibile>> come li tengo, tu lo sai, a me sta là...io...
  11. Uomo: – tu domenica chiudi quello vecchio, a' Mo'!
  12. Donna: – sì, sì!
  13. Uomo: – ciao, ciao!
  14. Donna: – ci vediamo domenica!
- ...omissis dalle ore 12.08.44 alle ore 12.20.53...*

Il primo fenomeno oggetto d'interesse è contenuto al rigo 5, dove troviamo una interpretazione di "a' Mò" proposta dal trascrittore tra parentesi come apocope del nome femminile "Morena". Con tale interpretazione il trascrittore effettua una precisa attribuzione di identità all'interlocutrice, preferendola ad altre possibili attribuzioni (per esempio, Monica) in quanto tale nominativo appartiene a una delle persone sottoposte a indagini e pertanto ricorre negli atti procedurali, facendo così parte della enciclopedia di conoscenze del trascrittore. Va detto che l'operazione interpretativa del trascrittore è errorea dal punto di vista linguistico, in quanto l'apocope corretta del nome Morena sarebbe "Moré", ma la sua volontà di integrare il dato nel quadro processuale lo porta a questa soluzione fallace<sup>5</sup>.

L'attribuzione effettuata dal trascrittore costituisce un atto interpretativo del parlato intercettato, che si risolve da parte del trascrittore nel preferire l'attribuzione del nome di Morena anziché interpretare tale dato come l'allocuzione "Amò", che non attribuisce alcuna identità all'interlocutore o all'interlocutrice, ma semmai segnala un livello di intimità nella interazione. L'allocuzione "Amò" presenta il medesimo fenomeno del troncamento e costituisce espressione coerente con il contesto dell'interazione, trattandosi di espressione ampiamente diffusa nel contesto laziale e, ormai, anche nazionale. Tale ultima opzione tuttavia viene probabilmente scartata dal trascrittore il quale preferisce lasciarsi guidare nella propria interpretazione dall'enciclopedia di conoscenze di cui è in possesso, in qualità di ufficiale di polizia giudiziaria, per aver seguito le indagini.

<sup>5</sup> Ringraziamo la/il revisore anonima/o per il suggerimento.

Proprio tale scelta interpretativa del trascrittore sottolinea l'importanza di quanto osserva Helen Fraser (2003), laddove raccomanda come il trascrittore dovrebbe essere il più possibile estraneo al contesto enciclopedico del caso e, pertanto, sarebbe opportuno che non conoscesse l'attività di indagine nella quale si inserisce l'interazione. Solo in tal modo si potrebbe evitare il rischio che il trascrittore proietti le proprie anticipazioni e previsioni sull'ascolto dell'intercettazione.

Tuttavia la interpretazione di espressioni vaghe o dipendenti dal contesto e dal cotesto è molto diffusa nelle trascrizioni forensi. Questo avviene perché l'interazione parlata, per sua natura, è strutturata in maniera ordinata e razionale, sia per quanto riguarda la sua organizzazione generale (Heritage, Greatbatch, 1991), che per il susseguirsi sequenziale delle interazioni (Sacks, Schegloff & Jefferson, 1974). Essa cioè è dipendente sia dal contesto, quello situazionale cui fanno riferimento le espressioni cd. "deittiche" e quello enciclopedico, sia dal cotesto, vale a dire da ciò che è stato detto prima e ciò che è stato detto dopo. Ogni azione dunque deve essere interpretata alla luce dell'azione precedente e dell'azione che la segue (Sacks, Schegloff & Jefferson, 1974).

Nel parlato esistono molte espressioni deittiche fantasmatiche,<sup>6</sup> come le definisce Maria-Elisabeth Conte (1981) riprendendo Karl Bühler (1934), ed è proprio in queste espressioni che spesso il trascrittore interviene per integrare e chiarire in via interpretativa la vaghezza e l'incompletezza del parlato, ciò allo scopo di rendere fruibile la trascrizione ai fini processuali. Tuttavia l'effetto è quello di una estrapolazione del dato dalla dimensione parlata, cui segue la sua ricollocazione in un nuovo testo, quello trascritto, con conseguente impossibilità di interpretare tali espressioni correttamente (è il processo dell'"entextualization" osservato da Bucholtz, 2000). Tali osservazioni consentono di rilevare come i caratteri della vaghezza, della frammentarietà e dell'apparente incompletezza nella interazione, essendo caratteristiche naturali del parlato spontaneo, non possano di per sé giustificare un giudizio di enigmaticità della conversazione, attribuendo alla stessa contenuti criptici e finalità illecite, come spesso enunciato dalla Suprema Corte.

Proprio la struttura sequenziale dell'interazione consente spesso di disambiguare alcune espressioni cui i parlanti possono far riferimento nel corso dell'interazione. Nel dato in esame, tuttavia, la presenza di *omissis*, cioè di parti dell'interazione che non sono state trascritte dall'ufficiale di polizia giudiziaria, impedisce di acquisire informazioni utili per valutare l'appropriatezza dell'interpretazione offerta dal trascrittore. Si fa riferimento soprattutto alla mancata trascrizione dell'*incipit* della conversazione, corrispondente ai primi 19 secondi, momento interazionale che spesso risulta denso di informazioni circa il livello di intimità e confidenza tra gli interlocutori. Tale omissione ci impedisce pertanto di rilevare se l'allocutivo "A' Mo'" fosse coerente alla luce del tipo di rapporto eventualmente segnalato dall'*incipit* della conversazione.

<sup>6</sup> Orletti (2016: 57) porta numerosi esempi di espressioni fantasmatiche tratte da intercettazioni: "quella roba lì; i cari amici tuoi, il quartierino, ecc.". Si tratta di espressioni indefinite, comprensibili solo a chi conosce il contesto, sia enciclopedico che situazionale dell'interazione.



Gli *omissis* dunque, pur costituendo una diffusa pratica nelle trascrizioni forensi curate dalla polizia giudiziaria, interrompono la concatenazione sequenziale della conversazione intercettata e, pertanto, alterano la rappresentazione dell'interazione parlata, dunque del dato probatorio originario.

### 7. *Aspetti linguistici della trascrizione e riflessi sul processo penale*

Quanto osservato evidenzia come il dato linguistico, emergente dalle trascrizioni di una intercettazione, in realtà si rifletta molto concretamente sulla rappresentazione della prova che viene consegnata alle parti processuali e al giudice.

La trascrizione in esame, infatti, ha determinato l'insorgere nel processo di alcune questioni di tipo probatorio.

La prima ha riguardato l'attribuzione effettuata dal trascrittore di una specifica identità all'interlocutrice della conversazione intercettata, attribuzione che costituisce un vero atto identificativo con rilevanti conseguenze sull'accertamento del coinvolgimento di tale soggetto nei fatti di reato oggetto del processo.

La seconda questione è stata relativa alla collocazione grafica di tale interpretazione e attribuzione di identità, che il trascrittore colloca sulla stessa riga del dato trascritto. Nella fattispecie tale scelta tipografica ha impedito alle parti di rilevare con chiarezza il confine tra la voce del trascrittore e la voce trascritta, con la conseguenza che nel corso del processo lo stesso pubblico ministero, titolare delle indagini, ha avuto difficoltà a reperire l'atto, costituito dalla trascrizione in esame, nel quale l'interlocutrice è stata identificata per la prima volta con Morena. A tal proposito giova ricordare come la trascrizione costituisca un modo di riportare le parole altrui tramite le forme del discorso diretto, rispetto al quale l'intervento del trascrittore deve esser distinto e segnalato adeguatamente con vari accorgimenti, non diversi da quelli utilizzati dal narratore in un testo letterario, come ci ricorda Bice Mortara Garavelli (1985). Adeguati segnali di questo passaggio di voci, da quella del soggetto intercettato a quella del trascrittore, avrebbero potuto essere costituiti dal mutamento del carattere tipografico, dall'impiego delle virgolette, dalla dislocazione dell'interpretazione in nota a piè di pagina. Emerge qui la necessità di una formazione linguistica dei trascrittori, su cui la comunità scientifica dei linguisti forensi si è già espressa più volte<sup>7</sup>.

La terza questione, infine, ha riguardato gli *omissis*, la cui presenza nella trascrizione ha impedito di conoscere alcuni aspetti fondamentali della interazione, che possono rivestire una importante valenza probatoria. In particolare, gli *omissis* all'inizio e alla fine della conversazione hanno impedito di comprendere, per esempio, se gli interlocutori si conoscessero, quale fosse il rapporto tra di loro, se l'incontro fosse stato casuale o programmato, se alcuni fatti riferiti fossero stati introdotti per la prima volta all'inizio della conversazione, oppure se facessero già parte degli universi storicamente condivisi tra i parlanti.

<sup>7</sup> Ci limitiamo qui a ricordare Orletti (2017) e Romito, Frontera (2017).

Numerosi dunque possono essere gli effetti distorsivi che, partendo da aspetti linguistici emergenti da una trascrizione di intercettazione, possono riflettersi sulla stessa rappresentazione della prova nel processo penale, con conseguente rischio di alterazione dell'accertamento del fatto.

### 8. Osservazioni conclusive

Le considerazioni che precedono vorrebbero evidenziare la necessità che l'ordinamento giuridico preveda percorsi formativi adeguati per i trascrittori, fondati sulla ricerca sul parlato, che consentirebbero la costituzione di un albo nazionale dei trascrittori, attualmente assente.

Appare altresì indispensabile l'adozione a livello nazionale di linee guida per i trascrittori, quali quelle già elaborate in ambito accademico nel 2016 con specifico riferimento alle intercettazioni telefoniche e ambientali (Orletti, 2016). Giova in questa sede ricordare alcune delle raccomandazioni ivi contemplate: quella di non tradurre, sia da una varietà dialettale sia da un'altra lingua; quella di non interpretare gli enunciati non intelleggibili e di segnalare, al contrario, la difficoltà di comprensione; quella di riportare le peculiarità del parlato; quella di mantenere integra la struttura sequenziale dell'interazione evitando gli *omissis*; quella di non integrare il testo trascritto con parti originariamente mancanti nell'interazione parlata, nella consapevolezza che proprio la vaghezza, la frammentarietà e l'apparente incompletezza costituiscono caratteristiche fondamentali del parlato spontaneo.

Queste linee guida non solo dovrebbero costituire la base della formazione professionale dei trascrittori, siano essi periti e consulenti, siano essi appartenenti alla polizia giudiziaria, ma dovrebbero altresì costituire patrimonio conoscitivo di tutti gli operatori nell'ambito giudiziario (avvocati e magistrati). La conoscenza dei fondamenti di linguistica forense, infatti, consentirebbe per esempio all'avvocato di rilevare elementi di inattendibilità di una trascrizione di intercettazione, consentendo in tal modo di valutare adeguatamente, per esempio, se prestare o meno il consenso all'utilizzo da parte del giudice di tale trascrizione per formare il proprio convincimento. La previsione di percorsi formativi in tema di linguistica forense permetterebbe a magistrati e avvocati di verificare le modalità con le quali la prova fonica, oggetto dell'intercettazione, sia stata rappresentata nella trascrizione e se tali modalità risultino adeguate. Solo in tal modo sarà garantito un effettivo controllo di legalità nella formazione della prova secondo il modello del "giusto processo" previsto dall'art. 111 c. 2 Cost.

Proprio seguendo tale prospettiva la Camera Penale di Roma, già da qualche anno, ha costituito una Commissione sulla linguistica giudiziaria con lo scopo di creare un ponte tra giuristi e linguisti forensi e di avviare così percorsi formativi per gli avvocati sul tema della linguistica forense.

*Riferimenti bibliografici*

- BELLUCCI, P. (2005). *A onor del vero. Fondamenti di linguistica giudiziaria*. Torino: Utet.
- BÜHLER, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fisher.
- BUCHOLTZ, M. (2000). *The politics of transcription*. In *Journal of Pragmatics*, 32 (10), 1439-1465.
- BÜHLER, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fisher.
- CONTE, M.E. (1981). *Texts deixis und Anapher*. In *Kodikas/Code*, 3(2), 121-132.
- DE MAURO, T. (2005). Introduzione. In BELLUCCI, P. (Ed.), *A onor del vero. Fondamenti di linguistica giudiziaria*. Torino: Utet.
- DU BOIS, J.W. (1991). Transcription design principles for spoken discourse research, In *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 1(1), 71-106.
- DURANTI, A. (2006). *Transcripts, like shadows on a wall*, in *Mind, Culture and Activity*, 13(4), 301-331.
- FRASER, H. (2003). *Issue in transcription: Factors affecting the reliability of transcripts as evidence in legal cases*, in *International Journal of Speech, Language and the Law*, 10(2), 203-226.
- HERITAGE, J. & GREATBATCH, D. (1991). *On the Institutional Character of Institutional Talk. The Case of News Interviews*. In BODEN, D., ZIMMERMANN, D.H. (Eds.), *Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis*. Cambridge: Polity Press, 93-137.
- JEFFERSON, G. (2004). *Glossary of transcript symbols with an introduction*. In LERNER, G.H. (Eds.), *Conversation Analysis: Studies from the first generation*. Amsterdam: John Benjamins Publishing Company, 13-31.
- MINISTERO DELLA GIUSTIZIA (2020), <https://webstat.giustizia.it/Analisi%20e%20ricerche/Rapporto%20su%20Intercettazioni%20fino%20al%202020.pdf>.
- MORTARA GARAVELLI, B. (1985). *La parola d'altri*. Palermo: Sellerio.
- OCHS, E. (1979). Transcription as theory. In OCHS, E., SCHIEFFELIN, B. (Eds.), *Developmental pragmatics*. New York: Academic Press.
- O'CONNELL, D.C. & KOWAL, S. (2009). Transcription Systems of spoken discourse. In D'HONDT, S., ÖSTMAN, J. & VERSCHUEREN, J. (Eds.), *The pragmatics of interaction*. Amsterdam: John Benjamins Publishing Company, 240-254.
- ORLETTI, F. (2016). La trascrizione delle intercettazioni telefoniche ed ambientali: un esercizio di analisi della conversazione applicata. In GATTA, F. (Ed.), *Parlare insieme. Studi per Daniela Zorzi*. Bologna: Bononia University Press, 49-64.
- ORLETTI, F. (2017). Transcribing intercepted telephone calls and uncovered recordings: an exercise of applied conversation analysis. In ORLETTI F., MARIOTTINI L. (Eds.), *Forensic Communication in Theory and Practice*. Cambridge: Cambridge Scholar Press, 11-26.
- ORLETTI, F. & TESTA, R. (1991). La trascrizione di un corpus di interlingua: problemi teorici e metodologici. In ORLETTI, F. (Eds.), *L'italiano dell'immigrazione: aspetti linguistici e sociolinguistici, Studi italiani di linguistica teorica e applicata*, XX (2), 243-283.
- PAOLONI, A. & ZAVATTARO, D. (2007), *Intercettazioni telefoniche e ambientali. Metodi, limiti e sviluppi nella trascrizione e verbalizzazione*. Torino: Centro Scientifico Editore.

ROMITO, L. & FRONTERA, M. (2017). La trascrizione forense di intercettazioni ambientali: una proposta di metodologia procedurale. In *Quaderni di linguistica*, 5, 105-120.

SACKS, H., SCHEGLOFF, E. & JEFFERSON, G. (1974). A simplest systematic for the organization of turn taking conversation. In *Language*, 50, 696-735.

SCHEGLOFF, E.A. (1986). The routine as achievement. In *Human Studies*, 9(2-3), 111-151.

RAFFAELE MANNA, ANTONIO PASCUCCI, JOHANNA MONTI

## La misurazione stilistica della Falsificazione.

### I comunicati delle Brigate Rosse

#### The Measure of Forgery. The statements of the Red Brigades

In Italy, during the so-called *Years of Lead*, the Red Brigades disseminated several written statements, in which they summarized the ideological intentions and the plan for the *Armed Struggle*. In addition to the propaganda dissemination, the Red Brigades reported about crimes and kidnappings of personalities from industry and politics using the form of written communication. In 1978, the Red Brigades kidnapped Aldo Moro giving news through a first statement, which was followed by nine other statements, one of which represents a fake. In the present study we show the possibilities of linguistic investigation in forensic science using Computational Stylemetry techniques. Taking as an example the statements that appeared during the Moro Case, it will be shown *a posteriori* to what extent the fake statement is different from the original ones. Furthermore, through the stylometric analysis of the statements, we show which are the stylistic and lexical characteristics that distinguish the group of Red Brigades statements from the fake one.

*Keywords:* Computational Linguistics, Computational Stylemetry, Digital Forensics, Red Brigades, Text Analysis.

#### 1. Introduzione

Nel corso degli anni, l'applicazione di metodi e tecniche proprie della linguistica per l'analisi del linguaggio parlato e scritto è stata progressivamente riconosciuta come saliente per fornire ulteriori evidenze e prove da aggiungere ad altri elementi probatori in ambito giurisdizionale. Nello specifico, l'intersezione di linguistica e discipline forensi è stata nominata come Linguistica Forense o, in alcuni casi, Linguistica Giudiziaria (Romito, 2013: 178). In questa, relativamente nuova, disciplina sono state riconosciute una serie di direzioni di ricerca, in gran parte ancorate all'applicazione della metodologia linguistica in diverse dimensioni forensi. In Gibbons e Turell (2008: 1) vengono descritte cinque categorie principali: una prima pertinente all'analisi del linguaggio legislativo; una seconda al linguaggio parlato nelle aule giudiziarie; una terza relativa all'elaborazione di evidenze; una quarta sull'insegnamento e nella ricerca del linguaggio legale scritto e parlato e infine, un'ultima categoria, sulla traduzione e interpretazione del linguaggio in ambito legale.

Nello specifico, per quanto riguarda l'elaborazione di evidenze linguistiche a partire da testi o parlato per fini investigativi e giudiziari, larga parte della ricerca e della pratica investigativa del linguista come perito si è concentrata sull'attribuzio-

ne e sulla descrizione del profilo autoriale a partire dalle caratteristiche linguistiche presenti in un testo e sul riconoscimento del parlante in casi di registrazione vocale (Coulthard, Johnson, 2007). In questo contesto e non restringendo il campo alla sola analisi di dati testuali in pratiche forensi, le metodologie e le tecniche appartenenti alla Linguistica Computazionale e al Trattamento Automatico del Linguaggio Naturale sono risultate adeguate nell'estrazione di caratteristiche linguistiche da testi e affidabili nell'attribuzione di testi ad un autore o ad una serie di autori diversi di paternità incerta (Koppel, Schler & Argamon, 2009; Rocha, Scheirer, Forstall, Cavalcante, Theophilo, Shen, Carvalho & Stamatatos, 2016) come anche nell'estrazione di caratteristiche sociolinguistiche e caratteristiche demografiche dai testi per la conseguente attribuzione degli stessi ad un profilo autoriale (Kocher, Savoy, 2017; Weren, Kauer, Mizusaki, Moreira, De Oliveira & Wives, 2014). Queste tecniche si avvalgono di componenti statistiche per l'analisi e l'estrazione di caratteristiche linguistiche presenti all'interno del testo o dei testi e, spesso, comprendono l'uso di algoritmi di classificazione e *clustering* per identificarne l'autore, con un certo grado di probabilità, in un dato insieme di autori conosciuti e in lizza per l'attribuzione o, nel caso in cui il possibile autore o autori non siano conosciuti, per attribuire il testo ad un profilo autoriale. In particolare, in compiti di attribuzione e profilazione autoriale, vengono spesso utilizzati metodi computazionali in grado di formalizzare ed estrarre caratteristiche stilistiche capaci di descrivere lo stile autoriale, ovvero la marca stilistica con la quale un testo o un insieme di testi sono stati scritti da un certo autore (Daelemans, 2013). Questo insieme di caratteristiche stilistiche viene definito e descritto nella ricerca e nei metodi appartenenti alla stilometria computazionale, ovvero come la misurazione quantitativa delle caratteristiche di stile. In stilometria, queste caratteristiche stilistiche sono intesi come tratti di scrittura fondamentali caratterizzanti lo stile di uno specifico autore. La misurazione quantitativa di caratteristiche stilistiche avviene ad ogni livello di analisi linguistica, a partire da tratti morfologici, lessicali e sintattici fino a tratti strutturali con i quali l'autore organizza la materia linguistica nei testi. Sebbene queste caratteristiche linguistiche e stilistiche compongano la quasi totalità dei testi scritti in una data lingua, un insieme delle stesse è in grado di descrivere e definire lo stile specifico di un autore che lo rende distinguibile da altri (Grant, Baker, 2001). In letteratura, ci si riferisce all'insieme di caratteristiche linguistiche e stilistiche proprie di un autore con espressioni come *impronta stilistica* o *DNA autoriale* (Daelemans, 2013; Eder, 2011) e con il termine *idioletto* (Coulthard, 2004). Oltre a ciò, le stesse caratteristiche stilistiche uniche e proprie di un autore sono state talvolta oggetto di tentativi di imitazione e, spesso, si sono verificati casi di falsificazione e interpolazione di testi (Heiser, 2007; Pearl, Steyvers, 2012).

In questo articolo, si discuteranno e si mostreranno le possibilità di applicazione di metodi e tecniche di stilometria computazionale in uno storico caso di interesse forense risalente a quelli che in Italia sono tristemente noti come *gli anni di piombo* (*fine anni Sessanta – inizio anni Ottanta*). Si prenderanno come caso di studio i nove comunicati delle Brigate Rosse (BR) insieme al falso comunicato, apparso il

18 Aprile 1978, diffusi durante il caso Moro. Si mostreranno *a posteriori* quali siano le caratteristiche stilistiche e linguistiche che distinguono il gruppo dei comunicati autentici delle BR rispetto al comunicato falso (denominato di seguito FBR7), utilizzando tecniche e metriche della stilometria computazionale. Inoltre, tramite un'analisi lessicale, si mostrerà quanto il falso comunicato attesti una lontananza e deviazione nel progresso ideologico e stilistico-linguistico rispetto alle intenzioni programmatiche delle BR.

Il presente articolo è strutturato come segue: nel secondo paragrafo si fornisce una breve narrazione della vicenda storica seguita dalla descrizione delle prime analisi condotte da linguisti ed esperti sui comunicati delle BR; nel terzo paragrafo si presenterà il corpus dei comunicati delle BR, insieme a statistiche descrittive; nel quarto paragrafo si descriverà la metodologia utilizzata per condurre le analisi e si mostreranno le caratteristiche stilometriche rilevate durante la comparazione linguistica del gruppo dei comunicati autentici a confronto con il comunicato falso; poi il paragrafo cinque sarà dedicato alle differenze lessicali tra due gruppi di comunicati autentici delle BR, tenendo come elemento separatore il comunicato falso; infine l'articolo si chiuderà con le conclusioni e la discussione delle possibilità offerte dalle stilometria computazionale in analisi testuali forensi.

## 2. I Comunicati delle Brigate Rosse e le prime analisi linguistiche

Durante la mattina del 16 Marzo 1978, due auto furono aggredite da una banda di terroristi in via Fani a Roma. I terroristi non risparmiarono i passeggeri della seconda auto e stessa sorte toccò ad alcuni passeggeri della prima auto. Su questa, viaggiava il Presidente della Democrazia Cristiana (DC) ed ex Presidente del Consiglio Aldo Moro che fu catturato. Per i successivi 55 giorni, Aldo Moro venne tenuto ostaggio dai rapitori in un luogo segreto.

Il 18 Marzo del 1978 venne diffuso a Roma un comunicato nel quale l'organizzazione terroristica di estrema sinistra, le Brigate Rosse (BR), rivendicò il rapimento di Aldo Moro, dando il via alla lotta armata e allo "scontro di classe".

Fin dal giorno seguente l'apparizione del primo comunicato, la macchina investigativa linguistica si attivò. Il primo a condurre un'analisi di tipo linguistico-filologico fu Tullio De Mauro con un articolo dal titolo "Tentativo di lettura filologica del messaggio BR. Non è come gli altri: sembra tradotto dal francese" pubblicato il 19 Marzo 1978 sul quotidiano *Paese Sera* (Marchetti, 2017: 53). In questo articolo, il celebre linguista ipotizza l'influsso della lingua francese con calchi e interferenze nel lessico usato dalle BR. Successivamente a questo primo articolo di analisi linguistica, un secondo articolo firmato da Arrigo Levi apparve sul quotidiano *La Stampa* dal titolo "L'analisi del volantino dei 'brigatisti' trovato a Roma. C'è un'ipotesi latino-americana" in data 21 Marzo 1978. Come titola quest'ultimo articolo, in questo viene avanzata un'ipotesi di interferenza linguistica con la lingua spagnola.

Durante i giorni del sequestro, a questo primo comunicato seguirono altri nove comunicati, tra i quali uno falso e poi smentito dalle stesse BR. Infatti, il comunicato

numero 7 del 18 Aprile 1978 dava notizia dell'avvenuta esecuzione di Aldo Moro e forniva un'indicazione dettagliata ed esatta del luogo nel quale la salma giaceva, ovvero nei fondali del Lago della Duchessa. Un paio di giorni dopo, il 20 Aprile 1978, un altro comunicato numero 7 apparve a Roma, Genova, Milano e Torino. In questo comunicato, le Brigate Rosse presero le distanze dal comunicato 7 del 18 Aprile e smentirono l'avvenuta uccisione di Aldo Moro allegando una foto dello stesso in vita e con una copia di un giornale datato 19 Aprile. In seguito a questo comunicato, le BR diffonderanno altri tre comunicati. Il 5 Maggio 1978 apparve l'ultimo comunicato, il numero 9, nel quale le BR diedero notizia dell'avvenuta condanna di Aldo Moro, destabilizzante per l'apparato statale e la DC e chiusero così la trattativa con le parti coinvolte. Il 9 Maggio 1978, alla segreteria della DC tramite una telefonata venne riferito il luogo esatto nel quale trovare il cadavere di Aldo Moro. Il cadavere era stato adagiato in una Renault 4 parcheggiata in via Caetani a Roma.

In seguito alla tragica conclusione dei fatti, l'interesse per la materia linguistica contenuta sia nei comunicati sia nelle lettere inviate da Moro nel periodo della prigionia, non si affievolì. Infatti, Leonardo Sciascia propose una lettura e un'analisi delle lettere di Moro ipotizzando l'intenzione del prigioniero di camuffare nel codice linguistico alcune indicazioni sulla posizione della sua prigionia (Sciascia, 1983). Sciascia, poi, fu chiamato a far parte della commissione parlamentare d'inchiesta costituita per il sequestro e l'omicidio di Moro. Proprio alla commissione d'inchiesta, nel 1981, venne inviato un dossier contenente un'analisi linguistica sui comunicati delle BR apparsi durante la prigionia di Moro. L'autore del dossier era Renzo Rota, diplomatico e consigliere all'ambasciata italiana a Mosca dal 1965 al 1972. Nel dossier, Rota ipotizzò che i primi due comunicati brigatisti fossero stati scritti da un sovietico adducendo come prove linguistiche determinati calchi fraseologici dal russo molti comuni nella retorica di alcuni quotidiani sovietici, ovvero *Pravda* e *Kommunist*. Anni dopo, nel 1990, apparve un articolo di Mario Medici (Medici, 1990) nel quale, oltre alla discussione sui diversi comunicati delle BR apparsi prima e dopo il caso Moro, lo studioso propose una breve discussione sul possibile numero di autori e sugli autori stessi dei comunicati ed espone un'analisi sui tipi e sulle principali strategie retoriche messe in atto. Oltre a ciò, propose un'analisi più incentrata sul tipo di linguaggio utilizzato nei comunicati, illustrando alcuni esempi di linguaggio tecnico e settoriale per quanto riguarda i comunicati apparsi durante il caso Moro.

Con riferimento a ricerche apparse nell'ultimo decennio, un articolo (Re, 2013) propone un'analisi linguistica dei comunicati delle BR durante il caso Moro. In questo articolo vengono sottolineate sia caratteristiche di continuità lessicale che di struttura morfosintattica ricorrenti nei comunicati, osservando che i documenti diffusi dalle BR sono dotati di una struttura unitaria. Di unitarietà linguistica e stilistica attraverso i comunicati si discute anche in (Marchetti, 2017). In questo articolo, vengono offerte alcune analisi sul registro linguistico con il quale i comunicati sono scritti e su alcune figure retoriche utilizzate per organizzare l'argomentazione. Inoltre, si propongono alcune analisi lessicali basate sulla frequenza nei comunicati,



ribadendo sia l'uniformità linguistica che l'abilità comunicativa nel gestire toni e registri formali, tecnici ed espressivi.

### 3. *Il Corpus dei comunicati*

Il corpus testuale analizzato in questo articolo è composto dai nove comunicati autentici inviati dalle BR durante il sequestro Moro con l'aggiunta del comunicato 7, prodotto di una falsificazione.

Di seguito, si farà riferimento ai singoli comunicati utilizzando una sigla di riferimento composta da BR seguita dal numero crescente per indicare l'ordine di apparizione dei comunicati (BR1, BR2, ...BR9), mentre (come detto già in precedenza) si indicherà il comunicato falso come FBR7. I testi dei comunicati sono stati estrapolati manualmente<sup>1</sup> e divisi in file testuali singoli con titolo composto dalla sigla di riferimento del comunicato e il numero separato da trattino basso (br\_1, br\_2, etc.). Per tutti i comunicati, sono state conservate le caratteristiche formali, ovvero parole in maiuscolo e spazi tra i paragrafi.

I comunicati sono stati poi analizzati utilizzando il pacchetto *Stylo*<sup>2</sup> (Eder, Rybicki & Kestemont, 2016) disponibile per il linguaggio R. Questo pacchetto offre gli strumenti e le metriche necessarie per condurre analisi e comparazioni di tipo stilometrico su testi.

Il corpus indagato in questo articolo è composto da un totale di 9.332 parole. Nella Tabella 1 si riportano le lunghezze testuali in termini di parole relative ad ogni comunicato, tenendo in considerazione solamente stringhe composte da caratteri alfabetici.

Tabella 1 - *Lunghezza testuale dei comunicati*

<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
739	1.342	1.023	<b>1.355</b>	675	952	1.136	910	1.051	<b>149</b>

In Tabella 1 sono evidenziate, in grassetto, la lunghezza maggiore raggiunta dal comunicato BR4 e la lunghezza minore riferita al FBR7. Già da questa analisi preliminare risulta una disparità in termini di parole utilizzate nel gruppo dei comunicati autentici e il falso comunicato. Infatti, oltre al BR4, anche gli altri comunicati prodotti dalle BR risultano essere composti da più parole rispetto al comunicato falso.

### 4. *Analisi dei Cluster e caratteristiche stilometriche*

In questa sezione si descriveranno gli esperimenti effettuati per analizzare il corpus dei comunicati delle BR e il comunicato FBR7. Nello specifico, si discuterà l'applicazione

<sup>1</sup> Il testo integrale di ogni comunicato è stato estrapolato dal sito web seguente: <http://www.archivio900.it/it/documenti/finestre-900.aspx?c=423>. Per ogni comunicato si è copiato solamente il testo escludendo la glossa iniziale a cura della redazione del sito.

<sup>2</sup> <https://cran.r-project.org/web/packages/stylo/index.html>

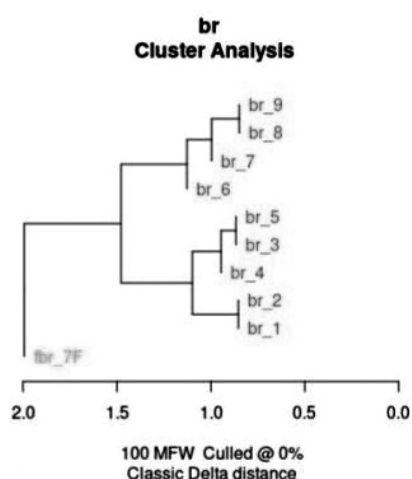
cazione di metodologie stilometriche e tecniche di *clustering* (che consentono di raggruppare i dati presi in analisi in classi omogenee in base alle similarità che i dati stessi presentano) con lo scopo di individuare le caratteristiche stilistico-linguistiche rilevanti per misurare e quantificare il grado di falsificazione e la distanza stilistica del falso comunicato 7 rispetto ai comunicati autentici.

Nel sottoparagrafo 4.1 si forniranno prove quantitative della falsificazione di FBR7 tramite l'utilizzo di singole caratteristiche stilometriche, mentre nel sottoparagrafo 4.2 si analizzeranno le co-occorrenze di parole per misurare la distanza tra il comunicato falso e i comunicati autentici.

#### 4.1 Analisi dei Cluster e caratteristiche stilometriche basate su singole parole

Questo tipo di analisi richiede innanzitutto una pre-elaborazione dei dati. Si è quindi proceduto alla trasformazione di tutti i caratteri in minuscoli, in modo da normalizzare i testi. Successivamente sono state calcolate le frequenze relative alle cento parole singole più comuni. Nessuna parola è stata rimossa in base ad una soglia di frequenza di occorrenza. In seguito, è stata effettuata un'analisi dei cluster utilizzando come caratteristiche testuali dei comunicati le frequenze delle cento parole singole più presenti estratte e come metrica il *Burrow's Delta* – Delta di Burrows – (Burrows, 2002) o *Classic Delta* in *Stylo*. Nello specifico, il Delta di Burrows misura quanto un testo di incerta attribuzione o anonimo, in alcuni casi, e un insieme di testi con autori certi e conosciuti divergono dalla media delle caratteristiche calcolata di questi stessi testi messi insieme e presi in esame. Inoltre, il metodo Delta assegna un peso uguale ad ogni caratteristica stilistico-linguistica che misura, evitando in questo modo che le parole più comuni inficino il risultato. L'analisi dei cluster così condotta è stata poi rappresentata tramite un dendrogramma dei raggruppamenti o cluster. Nella Figura 1 si mostra il risultato di questa prima analisi condotta sui testi.

Figura 1 - *Analisi dei Cluster sulle frequenze delle 100 parole più comuni*



Come riportato in Figura 1, il comunicato FBR7 sembra essere molto distante rispetto alla totalità dei comunicati autentici. Infatti, il dendrogramma mostra molto chiaramente la distanza stilistico-linguistica che esiste nei testi componenti il corpus. Oltre alla contrapposizione tra comunicati autentici e comunicato falso, in Figura 1 si può notare anche la formazione di due gruppi omogenei all'interno dei comunicati autentici: un primo che raggruppa i testi dei primi cinque comunicati, un secondo che avvicina i comunicati dal BR6 al BR9. Si descriverà con maggiori dettagli questo aspetto nel paragrafo 5.

L'ispezione manuale delle caratteristiche stilistiche e lessicali utilizzate durante questo tipo di analisi stilometrica, volta a misurare la distanza testuale tra i comunicati, ci mostra un insieme di caratteristiche rilevanti nel determinare il grado di falsificazione del comunicato FBR7. Nell'ispezione, oltre alla lunghezza minore in termini di parole del comunicato in esame rispetto ai comunicati autentici, già rilevata in precedenza, tra le caratteristiche stilistico-linguistiche che ci permettono di determinare il grado di falsificazione c'è la minore frequenza della congiunzione "e" rispetto ai comunicati autentici. Come anche rilevato in (Marchetti, 2017: 58), nei comunicati autentici sembra, infatti, prevalere una preferenza per la paratassi o coordinazione rispetto al comunicato falso. Come è evidente in Tabella 2, i valori relativi alla congiunzione *e* per il comunicato falso (segnalati in grassetto) sono distanti dal gruppo dei comunicati autentici. Nello specifico, in Tabella 2 si riportano i valori riguardanti sia la frequenza, calcolata come la "quota" appartenente ad ogni autore o comunicato per una data caratteristica per il numero totale di parole, sia quelli relativi allo *z-score* che rappresenta e descrive quanto sia distante una data caratteristica dalla norma dell'uso in un dato corpus in riferimento ad ogni autore o comunicato.

Tabella 2 - *Preferenza di coordinazione nei comunicati autentici*

	<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
<i>e</i>	3.51	2.75	5.37	4.72	4.44	4.51	3.34	4.95	4.94	<b>2.01</b>
<i>e (z-sc)</i>	-0.49	-1.18	1.20	0.60	0.35	0.41	-0.65	0.80	0.81	<b>-1.86</b>

In entrambi i casi, i valori di frequenza e *z-score* sono inferiori e si allontanano dalla norma d'uso dell'intero corpus dei comunicati BR. Tra i comunicati autentici, sembra mostrare una preferenza per la coordinazione il BR3 che supera di poco le 1.000 parole in lunghezza testuale. Il BR4, il comunicato più esteso testualmente, ha un valore leggermente minore rispetto ad altri due comunicati BR8 e BR9, il primo con una lunghezza testuale inferiore sia al BR9 che al BR4, il secondo con una lunghezza testuale superiore alle 1.000 parole.

Un secondo insieme di caratteristiche stilometriche utilizzate durante l'analisi fa riferimento ai connettivi di causa ed effetto come "quindi", all'avverbio di comparazione "come" e alla congiunzione avversativa "ma".

In Tabella 3 si riportano i valori di frequenza in riferimento a queste tre caratteristiche.

Tabella 3 - *Uso di connettivi, avverbi e congiunzioni*

	<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
quindi	0.13	0	0.09	0.22	0	0.21	0	0.21	0.19	<b>0</b>
come	0.14	0.74	0.48	0.29	0.14	0.31	0.96	0.32	0.20	<b>0</b>
ma	0.27	0.22	0.58	0.36	0.59	0.21	0.52	0.33	0.38	<b>0</b>

Anche in Tabella 3 sono segnalati in grassetto i valori del comunicato falso. Come è evidente dai risultati, il FBR7 non mostra nessun utilizzo del connettivo “quindi”, dell’avverbio di comparazione “come” e della congiunzione avversativa “ma”. Ciò ancora a dimostrazione della lontananza stilistico-linguistica rispetto ai comunicati autentici. Infatti, in questi ultimi sembra esserci una maggiore attenzione e raffinatezza nella strutturazione della materia linguistica e delle argomentazioni sviluppate attraverso l’utilizzo di strumenti stilistici come connettivi, avverbi e congiunzioni. Come anche discusso da Marchetti (2007: 59), l’organizzazione dei periodi nei comunicati BR sembra riflettere una padronanza della lingua scritta e uno stile analitico nel discutere ed organizzare il programma ideologico brigatista.

Inoltre, nel comunicato FBR7 non vi è alcun utilizzo di parole proprie del vocabolario militante del programma brigatista come “movimento”, “partito”, “controrivoluzione”, “proletariato”, “governo”, “potere”, “classe”, “comunista”, “multinazionali”.

#### 4.2 Analisi dei Cluster e caratteristiche stilometriche basate su co-occorrenze di parole

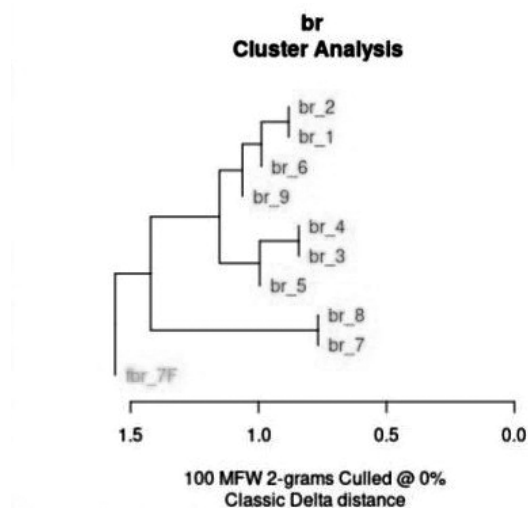
Di seguito, sarà descritta una seconda analisi stilometrica effettuata sui testi del corpus preso in esame con lo scopo di rilevare ulteriori caratteristiche in grado di quantificare la falsificazione. In questo secondo caso, saranno estratte e misurate le coppie di parole più frequenti nei comunicati componenti il corpus. Nello specifico, sono state calcolate le co-occorrenze delle parole più comuni per misurare la distanza stilistica tra i comunicati. Oltre ad una differenza stilistica tra i comunicati, in questa analisi, data la finestra testuale più ampia, sono state anche rilevate le differenze lessicali e di contenuto tra i comunicati.

A tal fine, sono state calcolate ed utilizzate le prime cento coppie di parole (2-grams) più frequenti all’interno del corpus e nessun *culling* è stato applicato all’estrazione delle caratteristiche testuali. Anche in questo esperimento, è stata effettuata un’analisi dei cluster utilizzando come metrica il Delta di Burrows (Burrows, 2002).

In Figura 2 si riporta il dendrogramma generato da questa seconda analisi. Come si evince dalla Figura 2, anche nella misurazione della distanza testuale utilizzando coppie di parole, il comunicato FBR7 risulta discostarsi dal gruppo di comunicati autentici. Inoltre, da un punto di vista linguistico-stilistico, nella formazione dei gruppi omogenei testuali, i comunicati autentici risultano aggregati secondo una diversa organizzazione della materia linguistica rispetto all’organizzazione mostrata in Figura 1. Infatti, i risultati di questa seconda analisi non suggeriscono la stessa contrapposizione stilistica tra i primi cinque comunicati e gli ultimi quattro, come invece evidenzia l’analisi sulle parole singole (cfr. Figura 1), bensì si dispongono in

funzione di una diversa organizzazione dei temi trattati in ciascun comunicato. Ciò è rilevabile dall'ispezione manuale delle caratteristiche testuali utilizzate nell'analisi.

Figura 2 - *Analisi dei Cluster sulle frequenze delle 100 coppie di parole più frequenti*



In Tabella 4 si riportano le prime cinque coppie di parole estratte e utilizzate durante questa seconda analisi. I valori relativi al comunicato falso sono segnalati in grassetto.

Tabella 4 - *Co-occorrenze rilevanti nel determinare la falsificazione*

	<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
stato imperialista	0.13	0.07	0.39	0.51	0.14	0.21	0.17	0.22	0.28	<b>0</b>
controrivoluzione imperialista	0.13	0.14	0.19	0.14	0.14	0.10	0	0	0.28	<b>0</b>
comunista combattente	0.13	0.14	0.19	0.22	0.29	0.10	0.08	0.11	0.19	<b>0</b>
borghesia imperialista	0.27	0.14	0.09	0	0	0.21	0	0.22	0.09	<b>0</b>
(stato) imperialista (delle) multinazionali	0.27	0.07	0.39	0.36	0.29	0.31	0	0	0.28	<b>0</b>
il movimento (rivoluzionario)	0.40	0.14	0.19	0.29	0.14	0.21	0.17	0.22	0.28	<b>0</b>
<b>la dc</b>	0.27	0.07	0	0.07	0	0.63	0.96	0.99	0.19	<b>0</b>
<b>della dc</b>	0.13	0.14	0	0.07	0.14	0.21	0.08	0.33	0.57	<b>1.35</b>
Aldo Moro	0.54	0.52	0.29	0.07	0.14	0.63	0.70	0.66	0.57	<b>2.02</b>

Nell'analisi delle co-occorrenze di parole si possono notare gli intenti programmatici e il linguaggio formulaico, come "stato imperialista", ideologicamente motivato dei comunicati autentici. Proprio per la presenza e l'utilizzo di formule ed espres-

sioni ideologiche, come “movimento rivoluzionario” o “stato imperialista”, il comunicato FBR7 viene a trovarsi in una posizione distante rispetto ai comunicati più programmatici. Infatti, in questo falso non compare nessun riferimento al lessico militante diffuso nella totalità dei comunicati autentici. Come anche il riferimento, oggetto o soggetto, all'*establishment* rappresentato da “la DC”, ha una frequenza nulla rispetto ai comunicati autentici. Sebbene in questi “la DC” venga maggiormente rappresentata linguisticamente nel gruppo di comunicati che va dal BR6 al BR8 rispetto a quanto sia rappresentato nei primi comunicati, nel comunicato FBR7 il riferimento alla Democrazia Cristiana è presente sotto forma di complemento (“della DC”), con una frequenza nettamente superiore rispetto alla media attestata nei comunicati autentici.

Nel comunicato FBR7, pur non essendoci un uso del lessico militante votato alla propaganda rivoluzionaria, l'intenzione di diffusione di false notizie intorno al sequestrato come tematica principale di tutto il comunicato può essere letto nell'uso di “Aldo Moro”, maggiore rispetto alla media dei comunicati autentici.

In maggior misura, in questa seconda analisi, proprio l'assenza di un vocabolario marcato dal punto di vista ideologico e la conseguente interruzione di una narrazione militante nel testo del comunicato FBR7, allontana quest'ultimo dai comunicati apparsi in precedenza.

##### 5. *Differenze lessicali e tematiche nei comunicati*

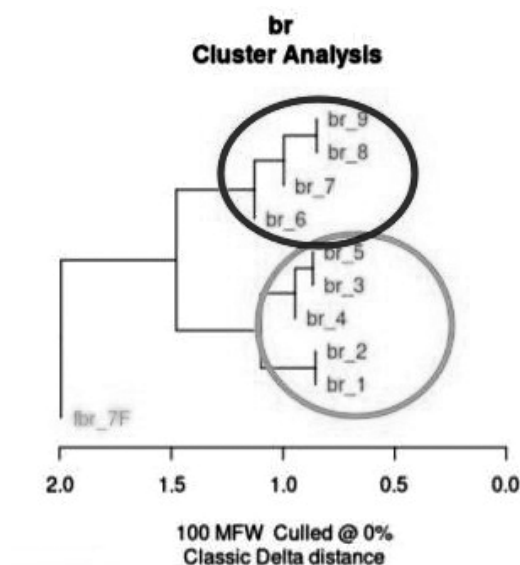
Nella Figura 1 è stato illustrato il dendrogramma generato dall'analisi delle caratteristiche stilometriche in riferimento alle cento parole più frequenti utilizzate nel corpus. Come anticipato nella sezione 4.1, nella suddetta figura si può notare la formazione di due gruppi omogenei e ben definiti di comunicati, accomunati da alcune caratteristiche lessicali ricorrenti nei due rispettivi gruppi. Ciò fornirà un'ulteriore misura della falsificazione e della lontananza ideologica del comunicato falso rispetto alle intenzioni programmatiche mostrate nei primi sei comunicati autentici (Re, 2013).

In questo paragrafo, si descriverà un metodo per misurare le differenze tra testi o gruppi di testi e, in generale, per misurarne la progressione temporale in termini lessicali. Per condurre ciò, utilizzando le caratteristiche lessicali estratte in precedenza in riferimento alle cento parole più frequenti, si è utilizzato un metodo basato sulla funzione *Craig's Zeta* (Burrows, 2007; Craig, Kinney, 2009).

Questa funzione non si basa sulla frequenza delle parole ma sul loro uso da parte di un autore in diversi testi. A tal fine, i testi vengono divisi in set di uguale dimensioni e grazie a questa funzione si confronta quante sezioni di ogni set contiene ciascuna parola. Il metodo combina la percentuale delle sezioni di un set di testi in cui compare una determinata parola con la percentuale delle sezioni di un set di testi in cui tale parola è assente in un'unica misura distintiva per ciascuna parola. In questo modo, si producono due liste di parole, la prima delle parole usate e preferite nel primo gruppo di testi e la seconda lista con le parole assenti nel primo gruppo di testi e presenti nel secondo gruppo.

Come mostrato in Figura 3, si prenderanno in esame i due gruppi segnalati e le rispettive differenze con il comunicato FBR7. I due gruppi segnalati sono disposti in ordine cronologico secondo la progressione temporale di apparizione dei comunicati. Infatti, il primo gruppo fa riferimento ai primi cinque comunicati, mentre il secondo gruppo si riferisce ai comunicati BR6 e successivi. Applicando la funzione Zeta di Craig, si rilevano le aree tematiche racchiuse intorno al lessico utilizzato nei diversi periodi di apparizione dei comunicati distinguendo così, dal lessico presente nei due gruppi, la progressione temporale e linguistica delle intenzioni e delle svolte programmatiche nelle BR.

Figura 3 - *Gruppi omogenei di comunicati*



Il risultato di questo metodo segnala due aree tematiche nel gruppo dei primi cinque comunicati e tre aree nel secondo gruppo. Il lessico utilizzato nella stesura dei primi cinque comunicati è nettamente più votato alla diffusione e alla chiamata verso una lotta e una rivoluzione contro l'establishment. Infatti, le parole maggiormente presenti nei primi cinque comunicati ruotano intorno all'area tematica coerente con azioni e concetti come: "Armata", "rivoluzionaria", "controrivoluzione", "proletariato", "iniziativa", "militari", "ristrutturazione", "popolo", "guerra", "terrore", da un lato; mentre la seconda area tematica si forma intorno all'oggetto contro il quale detta guerra rivoluzionaria deve essere svolta: "regime", "classe", "potere", "paese", "strutture", "multinazionali", "imperialista", "democristiani", "stampa".

Al contrario, nel secondo gruppo di comunicati che va dal BR6 al BR9, l'attenzione e le intenzioni delle BR si spostano sulla lotta alla contrattazione e all'accusa verso la classe politica del tempo come sicari e "complici" nella "condanna" di Aldo Moro. In questo gruppo di comunicati, dall'analisi vengono segnalate tre aree tematiche. Una prima sulle complicità nel sequestro e nella futura condanna: "complici",

“governo”, “morte”, “DC”, “condannato”, “Moro”, “sentenza”; una seconda sulle possibilità di accordo: “rilascio”, “libertà”, “risposta”, “Andreotti”; e una terza intorno alla liquidazione del sequestrato come problema non risolvibile: “scaricare”, “partito”, “prigioniero”, “problema”, “genocidio”, “condanna”.

## 6. Conclusioni

In questo articolo sono state presentate le analisi condotte con metodi di stilometria computazionale in grado di misurare la falsificazione a posteriori di un documento apparso durante il caso Moro. Durante l’analisi dei cluster, sono state presentate due applicazioni di metodi stilometrici utilizzate per quantificare le caratteristiche linguistico-stilistiche nei comunicati autentici e in quello falso, calcolando una finestra testuale differente con la metrica Delta di Burrows, al fine di estrarre e computare la distanza stilistica tra i comunicati e, di conseguenza, in grado di misurarne la falsificazione nel comunicato FBR7.

Inoltre, tramite l’applicazione di un’ulteriore misura stilometrica, nota come Craig’s Zeta, si è illustrato un metodo per comparare l’uso lessicale tra testi o gruppi di testi, in questo caso tra i comunicati, fornendo risultati che dimostrano le differenze lessicali tra testi e in grado di monitorare la progressione temporale stilistica e contenutistica in testi.

Il risultato di questi esperimenti condotti con tecniche stilometriche ha mostrato la capacità di estrarre tali evidenze testuali e, in generale, di analisi di dati testuali per l’identificazione e riconoscimento di caratteristiche linguistico-stilistiche in casi di falsificazione testuale.

## Ringraziamenti

Questa ricerca è stata svolta nell’ambito di un progetto di dottorato innovativo a caratterizzazione industriale finanziato dal fondo POR Campania FSE 2014/2020. Gli autori hanno contribuito alla stesura delle sezioni come segue: Raffaele Manna è l’autore dei paragrafi 3, 4 e 5. Antonio Pascucci è l’autore del paragrafo 1, mentre i paragrafi 2 e 6 sono in comune tra Raffaele Manna e Antonio Pascucci. Si ringrazia la Professoressa Johanna Monti per aver guidato la fase di ricerca e la stesura di questo articolo.

## Riferimenti bibliografici

BURROWS, J.F. (2002). ‘Delta’: a measure of stylistic difference and a guide to likely authorship. In *Literary and linguistic computing*, 17(3), 267-287.

BURROWS, J.F. (2007). All the way through: testing for authorship in different frequency strata. In *Literary and Linguistic Computing*, 22(1), 27-48.



- COULTHARD, M. (2004). Author identification, idiolect, and linguistic uniqueness. In *Applied linguistics*, 25(4), 431-447.
- COULTHARD, M. & JOHNSON, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Londra-New York: Routledge.
- CRAIG, H. & KINNEY, A.F. (Eds.). (2009). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.
- DAELEMANS, W. (2013). Explanation in computational stylometry. In *International Conference on Intelligent Text Processing and Computational Linguistics*. Berlin/Heidelberg: Springer, 451-462.
- EDER, M. (2011). Style-markers in authorship attribution: a cross-language study of the authorial fingerprint. In *Studies in Polish Linguistics*, 6(1), 99-114.
- EDER, M., RYBICKI, J. & KESTEMONT, M. (2016). Stylometry with R: a package for computational text analysis. In *The R Journal*, 8(1), 107-121.
- GIBBONS, J. & TURELL, M.T. (Eds.). (2008). *Dimensions of Forensic Linguistics*. Amsterdam Filadelfia: John Benjamins.
- GRANT, T. & BAKER, K. (2001). Identifying reliable, valid markers of authorship: a response to Chaski. In *Forensic Linguistics*, 8, 66-79.
- HEISER, M.S. (2007). The Majestic Documents: A Forensic Linguistic Report. <https://drmsb.com/michaelseiser/MJ%20Test%20article.pdf/> Accessed 07.01.2022
- KOCHER, M. & SAVOY, J. (2017). Distance measures in author profiling. In *Information processing & management*, 53(5), 1103-1119.
- KOPPEL, M., SCHLER, J., & ARGAMON, S. (2009). Computational methods in authorship attribution. In *Journal of the American Society for information Science and Technology*, 60(1), 9-26.
- MARCHETTI E. (2017). "Boia imperialisti, spie di regime e corrotti buffoni". La lingua dei comunicati delle brigate rosse durante il Sequestro Moro. In *L'analisi linguistica e letteraria*, 2, 51-70.
- MEDICI, M. (1990). Nel ventre del mostro. Le caratteristiche stilistiche e le fonti di ispirazione dei volantini prodotti dal terrorismo italiano. In *Italiano e oltre*, 1(5), 17-30.
- SCIASCIA, L. (1983). *L'affaire Moro. Con aggiunta la Relazione Parlamentare*. Palermo, Sellerio.
- PEARL, L. & STEYVERS, M. (2012). Detecting authorship deception: a supervised machine learning approach using author writeprints. In *Literary and linguistic computing*, 27(2), 183-196.
- RE, M. (2013). Il linguaggio dei 55 giorni che cambiarono l'Italia. Analisi dei documenti scritti dalle Brigate Rosse durante il sequestro di Aldo Moro. In *Revista de la Sociedad Española de Italianistas*, 9, 251-76.
- ROMITO, L. (2013). *Manuale di Linguistica Forense*. Roma: Bulzoni.
- ROCHA, A., SCHEIRER, W.J., FORSTALL, C.W., CAVALCANTE, T., THEOPHILO, A., SHEN, B., CARVALHO, A.R.B., & STAMATOS, E. (2016). Authorship attribution for social media forensics. In *IEEE transactions on information forensics and security*, 12(1), 5-33.
- WEREN, E.R., KAUER, A.U., MIZUSAKI, L., MOREIRA, V.P., DE OLIVEIRA, J.P.M., & WIVES, L.K. (2014). Examining multiple features for author profiling. In *Journal of information and data management*, 5(3), 266-266.



DONATELLA CURTOTTI, GABRIELLA DI PAOLO, WANDA NOCERINO

## La traccia vocale nelle indagini penali in Italia

### The vocal track in digital investigations in Italy

This paper offers an overview of the state of the art in relation to the use and procedural use of the vocal track. In particular, the Authors question the scientific nature of the sound track, suggesting corrective measures from a *de jure condendo* perspective.

*Keywords:* vocal track; scientific proof; criminal trial; guidelines; protocols.

#### 1. *La voce come traccia*

Da ormai qualche decennio, nel procedimento penale, le caratteristiche fisiche e comportamentali dell'individuo assumono un ruolo dirimente per l'accertamento dei fatti: gli inquirenti ricorrono sempre più spesso alle tecniche di biometria forense al fine di determinare o verificare l'identità del soggetto mediante l'uso automatizzato di parametri fisiologici<sup>1</sup>.

Tra gli strumenti che consentono l'individuazione di un soggetto per apportare un contributo conoscitivo alle indagini, assume particolare rilievo l'impronta vocale (*rec-tius*: la traccia fonica)<sup>2</sup>, dotata di un'autentica attitudine identificativa (Alesci, 2017).

Certamente, l'impulso al ricorso alle investigazioni vocali si è progressivamente amplificato in ragione dell'incremento di dispositivi per la comunicazione a distanza<sup>3</sup> e tramite la rete *Internet*: non può sottacersi, infatti, come la rivoluzione informatica degli ultimi tempi abbia profondamente modificato le abitudini degli individui, alterandone il modo di vivere, comunicare, interagire e intendere i rapporti interpersonali. Relazioni affrancate dalla dimensione fisica e materiale che cede il posto a quella eterea; intersezioni di sguardi, gesti, parole sostituite da algide digitazioni su scatole meccaniche che sembrano rappresentare l'unica interfaccia dell'uomo moderno. Una realtà, questa, che inevitabilmente involge e travolge prepotentemente anche il mondo del diritto, determinando un effetto domino che si ripercuote sulle più o meno tradizionali tecniche investigative, imponendone una

<sup>1</sup> L'affinamento delle tecniche identificative conduce all'utilizzazione di numerose tipologie di dati biometrici, distinti in dinamici (la grafia, la tonalità della voce, l'analisi dell'andatura, i movimenti labiali) e statici: questi ultimi sono basati sul riconoscimento di caratteri fisici tendenzialmente immutabili (impronte digitali, iride, conformazione delle orecchie, odore del corpo, reticolo venoso del polso). Sulle tecniche di biometria forense cfr. Belfatto (2015).

<sup>2</sup> Sul tema, approfonditamente, si vedano La Regina (2018) e Ciampini (2009).

<sup>3</sup> Sul tema della remotizzazione, approfonditamente, cfr. Curtotti (2006).

furente modernizzazione nell'ottica della creazione di una «giustizia penale 2.0» (Lorusso, 2019: 821; Di Paolo, 2013: 736).

Non solo. La “digitalizzazione” delle informazioni è inevitabilmente destinata ad amplificare i suoi effetti nel momento storico che si vive; momento in cui il processo penale, al pari di ogni altro settore della vita, esige che le attività, le comunicazioni, i rapporti, avvengano in modalità remota quale strumento di contenimento dell'epidemia da Covid-19. Il che porta agevolmente a far pensare in via preliminare che, mai come ora, possa farsi strada un'apertura culturale inedita, una sorta di *favor* da parte del legislatore, della dottrina e della giurisprudenza, verso l'impiego più generalizzato delle indagini a distanza e verso una stabilizzazione di quelle misure emergenziali nate “a tempo”<sup>4</sup>, con lo scopo di imprimere un'accelerazione alla macchina giudiziaria<sup>5</sup>.

Si potrebbe arrivare ad immaginare che l'attenzione rivolta negli scorsi anni a tutte le forme di investigazione a distanza (nel tentativo di fornire adeguate risposte sotto il profilo della compatibilità degli esiti investigativi con il sistema costituito, ricorrendo molto spesso alla magmatica categoria della prova atipica) arrivi ad assumere direzioni nuove, inclini a riconoscere un'autonomia concettuale e una più solida tenuta rispetto ai tradizionali valori del processo penale.

In questo contesto e, più in generale, tutte le volte in cui non sia possibile pervenire all'identificazione di un soggetto ricorrendo ad ulteriori elementi probatori, la traccia vocale rappresenta un importante punto di partenza per gli inquirenti.

Da quanto detto, emerge la centralità della linguistica forense nel circuito processual-penalistico, non solo nelle aule di giustizia, quale tecnica di retorica dibattimentale o di analisi del linguaggio legislativo e giurisprudenziale, ma anche e soprattutto quale strumento investigativo utile durante le indagini di polizia.

Si tratta di una problematica ben nota sia nel settore delle intercettazioni telefoniche, in cui si riscontra di frequente l'utilizzo di *sim card* acquisite con documenti contraffatti (o tramite prestanome) che nel campo delle intercettazioni ambientali, a contrasto delle quali si adottano contromisure che rendono impossibile procedere ad attività di osservazione e identificazione di soggetti che si trovano in luogo di monitoraggio.

In questi casi, la traccia vocale ottenuta mediante l'intercettazione rappresenta sicuramente un buon punto di partenza per le investigazioni.

Non solo: la traccia fonica, infatti, assume un ruolo centrale anche nelle investigazioni di *intelligence* (Curtotti, 2018).

<sup>4</sup> Solo a titolo esplicativo, si pensi all'implementazione dei sistemi di celebrazione delle udienze da remoto e alla predisposizione di un complesso di regole per condurre attività di indagine a distanza. Cfr. art. 83, comma 12 *quater*, l. 24 aprile 2020, n. 27, di conversione, con modificazioni, del d.l. 17 marzo 2020, n. 18; art. 221, d.l. 19 maggio 2020, n. 34, convertito con modificazioni dalla l. 17 luglio 2020, n. 77; art. 23, d.l. 28 ottobre 2020, n. 137, convertito con modificazioni dalla l. 18 dicembre 2020, n. 176.

<sup>5</sup> In questo senso depono anche la l. 23 settembre 2021, n. 134 (c.d. “Legge Cartabia”). Si pensi alle norme relative alla stabilizzazione delle disposizioni relative all'implementazione del deposito telematico degli atti processuali penali, nonché all'introduzione di norme atte a legittimare l'utilizzo delle videoregistrazioni delle prove dichiarative e dei collegamenti a distanza.

Come noto, il terrorismo internazionale di matrice islamica e le minacce cibernetiche stanno determinando un arretramento della risposta statale verso inedite forme di prevenzione del crimine, con il precipuo intento di neutralizzare l'offesa ed evitare che danni devastanti si producano<sup>6</sup>.

Tra le nuove tecniche di investigazione preventiva (o, più correttamente, proattiva)<sup>7</sup>, la linguistica forense occupa un ruolo centrale per gli organi di *intelligence* governativa. Tra gli altri compiti, assai rilevanti per il lavoro di *intelligence* sono: a) l'identificazione automatica della lingua materna, utile per individuare il Paese di origine dei rifugiati politici che non sono in possesso di documenti; b) la previsione della radicalizzazione attraverso l'atteggiamento linguistico; c) l'identificazione del potere (dominante-dominato) attraverso lo studio della sovrapposizione e dei cambi di intensità della voce (Zambonini, 2013).

## 2. Le investigazioni per l'acquisizione delle tracce foniche

Prima ancora di addentrarsi nei meandri dell'impiego delle tracce vocali nel processo penale e, dunque, interrogarsi sulla spendibilità processuale della traccia vocale, sembra doveroso soffermarsi sul *modus operandi* degli inquirenti per l'acquisizione di tali elementi di prova.

Uno degli obiettivi delle indagini foniche è quello di identificare il parlatore attraverso la comparazione della voce anonima (ossia quella delle conversazioni captate o ascoltate sulla scena del crimine ovvero ottenuta nel corso di una qualsivoglia attività investigativa) con la voce della persona sottoposta alle indagini o dell'imputato, già nota agli inquirenti<sup>8</sup>.

Il *match*, ovvero l'identificazione dell'imputato mediante comparazione tra le differenti tracce vocali acquisite, è il risultato di una serie di passaggi, strettamente interconnessi tra loro, che così possono essere idealmente scomposti (Ajili, 2017; Foulkes et al., 2012: 557):

1. acquisizione di un campione vocale (c.d. saggio fonico);
2. attribuzione dell'identità del parlatore, avvalendosi delle capacità amnestiche dell'investigatore e, dunque, delle abilità sensoriali degli inquirenti o di un esperto.

Per quanto attiene all'acquisizione del saggio fonico, le possibili modalità che conducono all'apprensione di un campione vocale, sono due: la prima, presuppone la collaborazione fattiva del parlatore, che – ove richiesto – può scegliere se prestare la propria voce per la predisposizione del saggio fonico; la seconda, si attiva allorché il soggetto l'interessato rifiuti di collaborare. L'ostruzionismo eventualmente frapposto

<sup>6</sup> Sulle indagini preventive, volendo, cfr. Nocerino (2019).

<sup>7</sup> Così definite nella Risoluzione del XVIII Congresso internazionale di diritto penale, Istanbul, 20–27 settembre 2009, in *Rivista di diritto processuale*, 2010, 333 ss.

<sup>8</sup> Non solo. Perché nel caso in cui viene registrata la voce di un sospetto ma il colpevole non è ancora stato identificato e arrestato, le caratteristiche della voce possono essere esaminate per acquisire informazioni sull'origine regionale, sociale ed etnica del parlatore. Tale tecnica è nota come profilazione degli oratori (cfr. Jessen, 2008).

all'esecuzione di una ricognizione auditiva o all'esecuzione di un accertamento fonico è, tuttavia, aggirabile attraverso l'impiego di "sucedanei", potendo utilizzare un saggio fonico già in possesso degli inquirenti, ad esempio perché acquisito nel corso di una intercettazione o di un interrogatorio documentato attraverso mezzi di riproduzione fonografica. Del pari, nulla preclude di intraprendere uno specifico itinerario investigativo volto all'acquisizione di una traccia vocale da utilizzare per l'accertamento fonico o per l'esecuzione di un riconoscimento vocale. Il primo pensiero va naturalmente alle intercettazioni, allo stato consentite anche attraverso il captatore informatico<sup>9</sup>, il quale darà sicuramente un notevole impulso allo sviluppo delle investigazioni vocali.

Con riferimento alla verifica di compatibilità tra le due voci (quella nota e quella anonima), può dirsi che in dottrina esiste una bipartizione fondamentale relativamente ai metodi per l'attribuzione dell'identità a partire dalla voce (La Regina, 2018: 77): un metodo soggettivo, ovvero la tecnica che sfrutta la capacità di ciascun individuo di riconoscere una persona sentendola parlare<sup>10</sup>, e un metodo oggettivo che fonda l'identificazione sulla base di analisi strumentali del segnale acustico. In questo contesto, si distingue ulteriormente tra metodi automatici e semi-automatici a seconda che si effettui esclusivamente ricorso ad un *software* per confrontare i campioni vocali oppure sia necessario l'intervento di un esperto per selezionare il materiale, controllare e valutare i risultati (Grimaldi et al., 2014: 4).

### 3. *La spendibilità processuale della traccia vocale*

Una volta identificato il parlatore, attraverso la verifica di compatibilità tra la traccia vocale acquisita e quella dell'imputato già nota agli investigatori, ci si è posti l'interrogativo relativo alla valenza processuale degli elementi di prova ottenuti con le modalità sopra descritte<sup>11</sup>.

In questo senso, si procede ad analizzare il complesso di istituti che consentono l'ingresso della traccia vocale nel processo penale per permettere al giudice di valutarla al pari degli altri elementi di prova raccolti.

Lo si dirà immediatamente. Nonostante la centralità della linguistica e della fonetica forense nelle investigazioni penali, la legislazione nazionale appare poco avanguardista: a dispetto di quanto accade negli altri Paesi, non esiste alcuna regolamentazione dell'indagine fonica, determinando un ostacolo all'impiego processuale dei risultati probatori acquisiti. A ben guardare, infatti, il legislatore nazionale regola le tecniche di identificazione vocale solo quando le stesse vengono impiegate nel processo penale attraverso l'espletamento di perizie e consulenze tecniche quando è necessario un accertamento di tipo tecnico (artt. 220 ss. c.p.p.), ovvero ricorrendo alla ricognizione effettuata dall'investigatore che procede all'identificazione

<sup>9</sup> Sul tema, si consenta il richiamo a Nocerino (2021).

<sup>10</sup> La principale obiezione che si muove contro l'impiego di questi tipi di tecniche in ambito forense «è proprio la loro soggettività: i risultati ed i giudizi che da esse derivano, non possono essere quantificati con delle metriche riproducibili ed indipendenti dal soggetto che li ha espressi» (Bove et al., 2002: 479).

<sup>11</sup> Nella letteratura straniera, cfr. Saks, Koehler (2005).

di una voce registrata o di un suono che comunque custodisce nella sua memoria (Bontempelli, 2013).

Al di là dell'ipotesi – sicuramente più affidabile – della verifica di compatibilità da parte di un esperto<sup>12</sup>, si profilano criticità in rapporto alla “altre forme di ricognizione” (art. 216).

Seppur è vero che la ricognizione vocale non si differenzia dall'omologo utilizzabile per riconoscere il volto o l'aspetto di una persona (*ex* art. 213 c.p.p.), deve ammettersi che attraverso la tipizzazione della ricognizione vocale si assiste ad una consistente riduzione del «tasso di tipicità» (Felicioni, 2019: 266) dello strumento disciplinato dall'art. 216 c.p.p. rispetto alla ricognizione di persone.

Come precisato,

mentre in materia di ricognizioni personali il legislatore si è preoccupato di definire il rapporto tra cadenze acquisitive e attendibilità del risultato, indicando un percorso funzionale a contenere il rischio di errori nel riconoscimento [...], in materia di riconoscimento vocale, e più in generale rispetto alle “altre ricognizioni” [...], l'unico percorso di salvaguardia è tracciato con riferimento alle operazioni da compiere per sondare l'attendibilità del ricognitore [...] il quale – [...] invitato dal giudice a descrivere la voce da riconoscere, indicando tutti i particolari che ricorda – rievcherà le proprie impressioni uditive attraverso il ricorso a descrizioni – voce bassa, stridula, profonda, acuta, baritonale e così via – difficilmente decodificabili attraverso parametri oggettivi e, in quanto tali, verificabili (La Regina, 2018).

Ancor di più, le problematiche relative all'impiego processuale delle tracce foniche attengono ad alcune forme di riconoscimento del parlatore non tipizzate dal legislatore Fraser, 2018: 129; French et al., 2018: 298; Gold et al., 2019: 1).

Si tratta delle attività investigative basate sul c.d. riconoscimento informale o “atipico” effettuato dagli investigatori nel corso della testimonianza, bypassando il ricorso allo strumento disciplinato dall'art. 216 c.p.p.

In questi casi, ancor più che altrove, vengono adottate tecniche ricognitive “soggettive” fondate sull'ascolto ripetuto delle voci dei soggetti coinvolti nell'inchiesta. Lungi dal rappresentare un ostacolo all'ingresso di una simile “prova” nel processo – risultando solo subordinata alla testimonianza degli operanti<sup>13</sup> – non può non rilevarsi come l'elemento di prova così raccolto stenti ad essere affidabile, perché affidato esclusivamente alle capacità sensoriali degli ascoltanti. In questo caso, «non si richiede una valutazione desunta da canoni tecnico-scientifici» ma più semplicemente un apprezzamento «permeato di soggettivismo» (Melchionda, 1990: 553).

<sup>12</sup> Per dovere di chiarezza, si precisa che la trattazione della questione è rinviata al paragrafo successivo posto che l'identificazione della voce attraverso la perizia pone una serie di questioni attinenti all'impiego della prova scientifica nel processo penale.

<sup>13</sup> Così Cass., sez. II, 27 gennaio 2017, n. 12858, *CED Cass.*, n. 269900. In questo senso, la Suprema Corte ha ritenuto che la ricognizione di voce costituisce un valido indizio quando è ritenuto attendibile la disposizione di colui che, avendo ascoltato la voce dell'imputato, afferma di identificarlo con sicurezza. Cass., sez. I, 8 maggio 2013, n. 35011, in *CED Cass.*, n. 257209.

D'altra parte, si tratta di operazioni poco giustificabili sotto il profilo dei principi fondamentali in materia di prova. Proprio con riferimento ai riconoscimenti informali, la dottrina ha enucleato dal sistema un principio di infungibilità che discende dal dovere del giudice «di osservare l'ordine normativo che fa corrispondere ad ogni tipo di esigenza probatoria uno specifico mezzo per soddisfarla» (Rafaraci, 1998: 1743; Cavini, 2015). Di conseguenza, il riconoscimento vocale svolto secondo il modulo procedimentale della testimonianza deve considerarsi inutilizzabile *ex art.* 191 c.p.p., in quanto acquisito eludendo le garanzie pretese dall'art. 216 c.p.p.<sup>14</sup>.

#### 4. I rischi probatori della traccia fonica

Molteplici i rischi connessi al riconoscimento vocale: come sostenuto da autorevole dottrina, «[S]ono basi magmatiche quelle su cui si fonda il fenomeno ricognitivo, le quali, inevitabilmente, determinano un elevato tasso di soggettivismo del riconoscimento, oltre che un consistente grado di fallibilità» (La Regina, 2018: 3).

Intanto, sono le caratteristiche ontologiche della traccia vocale a far dubitare della sua stessa attendibilità (Albano Leoni & Maturi, 1991: 316). La voce, infatti, è un bioindicatore dotato di una capacità “caratterizzante imperfetta” (Biral, 2015), ciò non solo perché essa è variabile a livello inter-individuale, variando da persona a persona, ma anche intra-individuale, dal momento che la voce è soggetta a cambiamenti a breve termine (si pensi alle alterazioni determinate da stati di ansia, di salute, dal fumo di una sigaretta, dall'assunzione di bevande alcoliche) o anche a lungo termine determinati dal trascorrere del tempo. Determinante è poi il canale di trasmissione, perché la voce di un medesimo individuo è soggetta a modifica ove veicolata, ad esempio, da un telefono cellulare. Inoltre, altri aspetti del discorso e del linguaggio sono modellati dal contesto situazionale in cui il discorso ha luogo. Molte persone, ad esempio, parlano istintivamente a un volume più alto quando le condizioni di ascolto non sono ottimali, come quando si parla con un forte rumore di fondo: l'aumento dell'ampiezza del parlato ha numerosi effetti sul segnale acustico.

Da ciò discende che la voce: a) non è immutabile, potendo subire variazioni in forza di diversi fattori; b) non possiede le caratteristiche sufficientemente univoche da consentire la distinzione tra un individuo e un altro (Nobile, 2016); c) non esiste “una soglia limite superata la quale il livello di compatibilità possa considerarsi un valore tendenzialmente assoluto (Chimichi, 2011: 383).

Inoltre, anche i metodi di riconoscimenti fondati su analisi strumentali e parametri oggettivi si servono dell'intervento attivo di un operatore che dovrebbe essere opportunamente addestrato ad effettuare il processo di estrazione del parlato e, conseguentemente, garantire l'affidabilità del risultato. Allo stato, tuttavia, non viene riconosciuta in Italia una simile figura professionale e il tutto viene gestito dagli investigatori, spesso poco formati all'esecuzione di tali delicate operazioni.

<sup>14</sup> Sostiene tale posizione La Regina (2018: 122).



Ma ciò che ha caratterizzato il dibattito nazionale in tema di “traccia fonica” è la possibilità che essa riesca a superare il c.d. *Daubert test*<sup>15</sup>, ossia lo *standard* di validazione giuridica della scienza incerta<sup>16</sup>.

Come noto, perché una prova possa definirsi “scientifica”<sup>17</sup> deve fondarsi su coperture generali o statistiche con un coefficiente di probabilità pari a “1”<sup>18</sup>. La certezza processuale si fonda sul confronto dell’ipotesi ricostruttiva con l’evidenza disponibile ed è fondamentale che essa resista ai tentativi di falsificazione.

Al di là delle metodologie impiegate, l’esperto deve chiarire la tecnica prescelta, esplicitandone il tasso di errore, rendere pubblici i *test* effettuati per sondare l’affidabilità degli strumenti usati, nonché l’ampiezza della banca dati che si consultata e i motivi per cui si sono privilegiate alcune parti del segnale e ritenute aleatorie altre.

Proprio sulla base delle indicazioni provenienti dalla standardizzazione del *Dauber test*, in passato si è ritenuto che la traccia fonica non fosse idonea a superare le rigide regole sulla scientificità del metodo impiegato, posto che a) non sussiste un preciso limite che imponga di non utilizzare il materiale sonoro che superi alcuni limiti qualitativi; b) a differenza degli accertamenti aventi ad oggetto le impronte digitali, in materia di riconoscimento vocale non esiste neppure una soglia limite superata, quale il livello di compatibilità, che possa considerarsi un valore tendenzialmente assoluto; c) i risultati del metodo soggettivo, fondati esclusivamente sul giudizio uditivo o visivo dell’esperto e privi di criteri e procedure standardizzate, non sono riproducibili, ne è calcolabile il margine di errore e ciò inficia in punto di attendibilità l’impiego probatorio del risultato.

In altri termini, si è detto che, in questo settore, la compatibilità è sempre un valore relativo (Chimici, 2011).

### 5. *L’importanza della valutazione giudiziale*

Sebbene l’analisi delle caratteristiche vocali non possa determinare da sola l’identità di un parlante, essa può fornire una vasta gamma di informazioni sul parlante, anche se con vari gradi di precisione e sicurezza.

<sup>15</sup> Il primo criterio di valutazione consiste nella possibilità di testare l’ipotesi scientifica avanzata, di sottoporla a verifica empirica (Hempel), di falsificarla e confutarla. Il secondo criterio considera se la teoria sia stata oggetto di *peer review* e di pubblicazioni. Il terzo elemento è rappresentato dalla percentuale di errore, nota o potenziale, della teoria, ed un quarto risalente al *Frye standard* che, pur perdendo il proprio carattere vincolante – “*general acceptance is not a necessary precondition to the admissibility of scientific evidence*” –, riconosce il valore della comunità scientifica rilevante e della scienza normale. Sul punto, esaustivamente, si veda Curtotti et al. (2019: 7).

<sup>16</sup> Sui caratteri della prova scientifica, per tutti, cfr. Dominioni (2005).

<sup>17</sup> Con questa espressione non si designa un certo tipo di *thema probandum*, né una certa fonte o un certo mezzo di prova e nemmeno un particolare metodo di valutazione delle informazioni raccolte nel processo; bensì ci si riferisce, empiricamente, al sempre più diffuso fenomeno dell’impiego, nella formazione del giudizio di fatto, di nozioni e metodi cognitivi che esorbitano dalla comune esperienza e dalla cultura media della società alla quale il giudice appartiene.

<sup>18</sup> Cass., sez. un., 10 luglio 2002, n. 30328, *CED Cass.*, n. 22213.

Ecco la ragione per cui la dottrina più avanguardista ha ipotizzato ulteriori soluzioni per consentire l'ingresso delle più innovative tipologie di indagine tecnica che, seppur non perfettamente confacenti ai criteri Daubert, risultano idonee per garantire la scientificità del metodo adoperato.

In primo luogo, va precisato che alcuni recenti studi fanno riferimento alla necessità di procedere ad analisi statistica allo studio della voce mediante il calcolo del rapporto di verosimiglianza LR (*like-lihood ratio*) con l'applicazione del teorema di Bayes (cfr. Grimaldi, 2019). In altri termini, nel moderno approccio alla TFSI (*Technical Forensic Speaker Identification*) l'identificazione del parlante si ispira alla identificazione del DNA, cioè assumendo una prospettiva probabilistica.

In tale prospettiva,

l'esperto forense non deve e non può fornire la probabilità che il parlato registrato dell'anonimo sia stato prodotto dal sospettato. In altre parole, lo scienziato forense non deve presentare la probabilità di colpevolezza o di non colpevolezza. È compito del giudice giungere a queste probabilità e decidere sulla base di tutte le evidenze forensi (e non) che emergono durante il processo. Allo scienziato forense deve essere solo richiesta la forza dell'evidenza (Grimaldi et al., 2014: 4).

Perché ciò accada, l'esperto è tenuto a considerare due aspetti: la similarità, dei campioni di parlato dell'anonimo e del sospettato rispetto ai parametri di interesse e la tipicità delle caratteristiche fonetiche tra i due campioni di parlato rispetto a una popolazione di riferimento.

Proprio sulla base di tale assunto, la dottrina nazionale si è concentrata sull'importanza del contenuto della decisione giudiziale quale strumento di garanzia di attendibilità della tecnica impiegata.

Per intenderci, al fine di distinguere la "*Junk Science*" dalla "*Good Science*", si ricorre non solo e non tanto a canoni prestabili e predeterminati, quanto al contenuto del provvedimento decisionale del giudice.

In particolare, la dottrina ricorre alla c.d. "motivazione rafforzata" (Cecchi, 2021), definibile come

una formula con la quale, da un lato, si esorta la cautela decisionale verso specifici profili giuridici e, dall'altro, si pretende l'elaborazione di un impianto motivazionale irrobustito rispetto a tali questioni, la cui verifica si ritiene imprescindibile ai fini legittimità del provvedimento emanato. [...] La peculiare caratteristica di questa metodica di giudizio e di giustificazione si rinviene nel fatto che il giudice è tenuto a percorrere una serie di *step*, di passaggi obbligati, costituiti da argomenti che concernono aspetti salienti della fattispecie in esame che devono essere apprezzati alla luce di parametri criteri condivisi e consolidati, non che intersoggettivamente verificabili (Cecchi, 2021: 437).

Così, una volta affrontati tutti i passaggi valutativo motivativi obbligati (ad esempio il tasso di errore della teoria scientifica, la sperimentata abilità e la sottoposizione a tentativi di falsificazione, la condivisione della teoria scientifica nella comunità degli esperti, il *curriculum* del perito del consulente tecnico che la sostiene), l'autorità

giudiziaria potrà scegliere, concretizzando ciascun singolo argomento sulla scorta delle peculiarità del caso di specie, se accogliere o meno la teoria.

### 6. *Le aperture del mondo forense all'impiego della prova linguistica*

Alla luce delle considerazioni esposte, al fine di superare le criticità derivanti dall'ontologia stessa della traccia fonica, potrebbe essere auspicabile l'introduzione – e la conseguente dotazione alle forze di polizia – di protocolli operativi standardizzati da applicare ai casi concreti, i quali, nella duplice veste di garanti della corretta valutazione dello stato *de quo* da parte degli operatori del settore e di guida per il successivo controllo giurisdizionale, pur non vincolando l'attività della p.g., consentono di delineare un assetto normativo “stabile” e generalmente condiviso, contenendo il rischio che la discrezionalità degli interpreti, prima, e degli operanti, poi, sfoci in arbitrio.

In questo senso, anche il compito di valutazione del giudicante potrebbe essere facilitato. Come noto, seguendo l'insegnamento della sentenza c.d. Cozzini<sup>19</sup>, il giudice, oltre al rispetto delle regole di valutazione della prova scientifica, deve anche vagliare i criteri seguiti nella formazione della stessa, verificando la compatibilità con i requisiti di affidabilità delle informazioni rese dagli esperti nel processo penale. Di qui, la predisposizione di protocolli operativi standardizzati sulle modalità di acquisizione non già della traccia vocale ma dello stesso saggio fonico (e, dunque, non solo in fase di comparazione ma anche di acquisizione), potrebbe essere utile per abbattere il rischio di valutazioni – più o meno colposamente – errate.

In questo senso, non possono essere sottaciute le importanti iniziative messe in campo dall'Associazione Italiana di Scienze della Voce proprio in merito alla predisposizione di adeguate linee guida in materia<sup>20</sup>.

Inoltre, non può non rilevarsi come il primo e più serio investimento in materia di identificazione del parlatore andrebbe effettuato nel campo della formazione, con il supporto delle associazioni scientifiche che, come accade in altri Paesi, dovrebbero coadiuvare l'azione degli esperti predisponendo apposite risoluzioni e linee guida, di ausilio anche per l'attività di controllo del giudice, da esplicitare non solo sulle metodologie impiegate per l'accertamento ma anche, e prima ancora, sulla qualificazione dell'esperto da chiamare per effettuarlo.

Un dato va comunque precisato. Deve ammettersi che negli ultimi anni sono assolutamente numerosi i passi in avanti effettuati sul campo e notevoli sono le possibili evoluzioni della materia. Tra queste si segnala il progetto di biometria vocale dell'Arma dei Carabinieri (13 luglio 2017), volto alla condivisione dei risultati ottenuti con tecniche di riconoscimento semi-automatiche per la creazione di una banca dati da adottare secondo il modello giuridico già previsto per la banca dati del DNA.

<sup>19</sup> Cass., sez. VI, 13 dicembre 2010, n. 43786, in *CED Cass.*, n. 248944.

<sup>20</sup> Cfr. Proposta di linee guida dell'Associazione Italiana di Scienze della Voce, 2019. In ambito sovranazionale, va rimarcato l'impegno dell'Associazione Internazionale di Fonetica Forense ed Acustica (<https://www.iafpa.net/>) e della rete Europea degli Istituti di Fonetica Forense (ENFSI – <https://enfsi.eu/>).

*Riferimenti bibliografici*


- AJILI, M. (2017). *Reliability of voice comparison for forensic applications. Artificial Intelligence*. Université d'Avignon: PhD thesis in Philosophy.
- ALBANO LEONI, F., MATURI, P. (1991). Fonetica sperimentale e fonetica giudiziaria. In *Giustizia penale*, 96(10), 316-320.
- ALESCI, T. (2017). *Il corpo umano fonte di prova*. Padova: Cedam.
- BELFATTO, E. (2015). *La biometria applicata alla sicurezza ed al contesto forense*. Bologna: FDE Institute Press.
- BIRAL, M. (2015). L'identificazione della voce nel processo penale: modelli, forme di accertamento, tutela dei diritti individuali. In *Rivista italiana di diritto e procedura penale*, 58(4), 1842-1879.
- BONTEMPELLI, M. (2013). La ricognizione. In FERRUA, P., MARZADURI, E., SPANGHER, G. (Eds.). *La prova penale*. Torino: Giappichelli, 479-534.
- 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*, LXII (3), 475-490.
- CECCHI, M. (2021). *La motivazione rafforzata del provvedimento. Un nuovo modello logico argomentativo di stilius curiae*. Milano: Wolters Kluwer.
- CAVINI, S. (2015). *Le ricognizioni e i confronti*. Milano: Giuffrè.
- CHIMICHI, S. (2011). Profili giuridici del riconoscimento del parlante. In CONTI, C. (Eds.). *Scienza e processo penale. Nuove frontiere e vecchi pregiudizi*. Milano: Giuffrè, 119-142.
- CIAMPINI, C. (2009). Indagini foniche. In PICOZZI, M., INTINI, A. (Eds.). *Scienze forensi. Teoria e prassi dell'investigazione scientifica*. Torino: Utet, 405-415.
- CURTOTTI, D. (2006). *I collegamenti audiovisivi nel processo penale*. Milano: Giuffrè.
- CURTOTTI, D. (2018). Procedimento penale e intelligence in Italia: un'osmosi inevitabile, ancora orfana di regole. In *Processo penale e giustizia*, 3, 435-448.
- CURTOTTI D., FISHER B.A.J., HOUCK M.M., SPANGHER G. (2019). Diritto e scienza: un rapporto in continua evoluzione. In CURTOTTI, D., SARAVO, L. (Eds.). *Manuale delle investigazioni sulla scena del crimine*. Torino: Giappichelli, 1-35.
- DI PAOLO, G. (2013). Prova informatica (diritto processuale penale). In *Enciclopedia del diritto*. Milano: Giuffrè, VI vol., 736-762.
- DOMINIONI, O. (2005). *La prova penale scientifica*. Milano: Giuffrè.
- FELICIONI, P. (2019). Il riconoscimento del parlante tra prassi e modelli normativi. In SCALFATI A. (Eds.). *Le indagini atipiche*. Torino: Giappichelli, 259-292.
- FOULKES, P. & FRENCH, P. (2012). Forensic speaker comparison: the linguistic-acoustic perspective. In L. SOLAN & P. TIERSMA (Eds.). *Oxford Handbook of Language and Law*. Oxford: Oxford University Press, 557-572.
- FRASER, H. (2018) "Assisting" listeners to hear words that aren't there: dangers in using police transcripts of indistinct covert recordings. In *Australian Journal of Forensic Sciences*, 50(2), 129-139.
- FRENCH, P. & FRASER, H. (2018). Why 'ad hoc experts' should not provide transcripts of indistinct forensic audio, and a proposal for a better approach. In *Criminal Law Journal*, 42(5), 298-302.


- GOLD, E., & FRENCH, P. (2019). International practices in forensic speaker comparisons: second survey. In *International Journal of Speech Language and the Law*, 26(1), 1–20.
- GOLD, E. (2014). Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters. University of York: PhD Dissertation.
- GRIMALDI, M. (2019). Al di là di ogni ragionevole dubbio: voce, scienza e legge nella fonetica forense. In *Questione giustizia*.
- GRIMALDI, M., D'APOLITO, S., GILI FIVELA, B., SIGONA, F. (2014). Illusione e scienza nella fonetica forense: una sintesi. In *Mondo digitale*, 13(53), 1–9.
- LA REGINA, K. (2018). *L'identificazione della voce nel processo penale*. Padova: Cedam.
- LORUSSO, S. (2019). Digital evidence, cybercrime e giustizia penale 2.0. In *Processo penale e giustizia*, 4, 821-828.
- JESSEN, M. (2008). Forensic Phonetics. In *Language and Linguistics Compass*, 2, 671–711.
- MELCHIONDA, A. (1990). Sub art. 216. In CHIAVARIO, M. (Eds.). *Commentario al nuovo codice di procedura penale*. Torino: Utet, 552.
- NOBILE, E. (2016). Le comparazioni vocali. In BARBARO, A., LA MARCA, A., NOBILE, E., ROMEO, P. (Eds.). *La prova tecnica nel processo penale. Aspetti pratico scientifici*. Milano: Key Editore, 135-154.
- NOCERINO, W. (2019). *Le intercettazioni e i controlli preventivi sulle comunicazioni. Riflessi sul procedimento probatorio*. Padova: Cedam.
- NOCERINO, W. (2021). *Il captatore informatico nelle indagini penali interne e transfrontaliere*. Padova: Cedam.
- RAFARACI, T. (1998). Ricognizione informale dell'imputato e (pretesa) fungibilità delle forme probatorie. In *Cassazione penale*, 1743.
- SAKS, M.J. AND KOEHLER, J.J. (2005). The coming paradigm shift in forensic identification science. In *Science*, 309, 892–5.
- ZAMBONINI, G. (2013). Metodi di riconoscimento della voce. In VALLI, R.V.O. (Eds.). *Le indagini scientifiche nel procedimento penale*. Milano: Giuffrè, 749.





## Autori

IACOPO BENEVIERI – Camera Penale di Roma, Responsabile Commissione sulla Linguistica Giudiziaria  
iacopobenevieri@hotmail.com


LUCIA BUSO – Aston Institute for Forensic Linguistics, Aston University, Birmingham, UK  
l.busso@aston.ac.uk;  <https://orcid.org/0000-0002-5665-771X>

SONIA CENCESCHI – Digital Forensic Service, Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Switzerland  
sonia.cenceschi@supsi.ch;  <https://orcid.org/0000-0002-4145-9593>


ELEANOR CHODROFF – Department of Language and Linguistic Science, University of York, United Kingdom & Department of Computational Linguistics, University of Zurich  
eleanor.chodroff@uzh.ch;  <https://orcid.org/0000-0003-4549-5919>


CLAUDIA ROBERTA COMBEI – Dipartimento di Studi Umanistici, Università di Pavia  
claudiaroberta.combei@unipv.it;  <https://orcid.org/0000-0003-1884-8205>


DONATELLA CURTOTTI – Dipartimento di Giurisprudenza, Università di Foggia  
donatella.curtotti@unifg.it


SONIA D'APOLITO – Dipartimento di studi umanistici, Università Del Salento, Lecce  
sonia.dapolito@unisalento.it;  <https://orcid.org/0009-0002-5593-5076>


GABRIELLA DI PAOLO – Dipartimento di Giurisprudenza, Università di Trento  
gabriella.dipaolo@unitn.it


PETER FRENCH – JP French Associates, United Kingdom and Department of Language and Linguistic Science, University of York, United Kingdom  
johnpeter.french@uzh.ch;  <https://orcid.org/0000-0001-7124-8896>


BARBARA GILI FIVELA – Dipartimento di studi umanistici, Università Del Salento, Lecce  
barbara.gilifivela@unisalento.it;  <https://orcid.org/0000-0002-4694-8652>


MIRKO GRIMALDI – Dipartimento di studi umanistici, Università Del Salento, Lecce  
mirko.grimaldi@unisalento.it;  <https://orcid.org/0000-0002-0940-3645>


COSIMO IAIA – Dipartimento di studi umanistici, Università Del Salento, Lecce  
cosimo.iaia@unisalento.it;  <https://orcid.org/0000-0002-4317-4540>


RAFFAELE MANNA – Dipartimento di Studi Letterari, Linguistici e Comparati,  
Università degli Studi di Napoli “L’Orientale”  
rmanna@unior.it;  <https://orcid.org/0009-0006-6285-8557>


CHIARA MELUZZI – Dipartimento di Studi Letterari, Filologici e Linguistici,  
Università degli Studi di Milano  
chiara.meluzzi@unimi.it;  <https://orcid.org/0000-0002-2291-006X>


JOHANNA MONTI – Dipartimento di Studi Letterari, Linguistici e Comparati,  
Università degli Studi di Napoli “L’Orientale”  
jmonti@unior.it;  <https://orcid.org/0000-0002-4563-5988>


UMAR MUHAMMAD-GOMBE – Department of Language and Linguistic Science,  
University of York, United Kingdom  
ugm501@york.ac.uk;  <https://orcid.org/0000-0002-6846-4704>

WANDA NOCERINO – Dipartimento di Giurisprudenza, Università di Foggia  
wanda.nocerino@unifg.it;  <https://orcid.org/0000-0002-3751-9928>

FRANCA ORLETTI – Professore Ordinario Onorario di Glottologia e Linguistica –  
Dipartimento di Studi Umanistici Università Roma Tre  
franca.orletti@uniroma3.it; [www.chiareparole.it](http://www.chiareparole.it);  <https://orcid.org/0000-0001-6925-7742>

ANTONIO PASCUCCI – Dipartimento di Studi Letterari, Linguistici e Comparati,  
Università degli Studi di Napoli “L’Orientale”  
apascucci@unior.it;  <https://orcid.org/0000-0002-5960-7371>

LUCIANO ROMITO – Dipartimento di Culture, Educazione e Società, Università  
della Calabria  
luciano.romito@unical.it;  <https://orcid.org/0000-0003-4681-5865>

FRANCESCO SIGONA – Dipartimento di studi umanistici, Centro di Ricerca  
Interdisciplinare sul Linguaggio, Università Del Salento, Lecce  
francesco.sigona@unisalento.it;  <https://orcid.org/0000-0003-2939-0009>