SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

BERGAMO UNIVERSITY PRESS
sestante edizioni

UNIVERSITÀ DEGLI STUDI DI BERGAMO | Dipartimento di Lingue, Letterature e Culture Straniere

Sestante Edizioni - Bergamo
www.sestanteedizioni.it

SUPERARE L'EVANESCENZA DEL PARLATO
Un vademecum per il trattamento digitale di dati linguistici
Giuliano Bernini - Ada Valentini - Jacopo Saturno - Lorenzo Spreafico (A cura di)
p. 262 cm. 15,5x22,0
ISBN: 978-88-6642-369-0

Printed in Italy
by Sestanteinc - Bergamo

In copertina: "Evangelista, or letter-writer, and his clients". Immagine tratta da Brown, Robert. 1894. The Countries of the World: being a popular description of the various continents, islands, rivers, seas, and peoples of the globe [with plates]. Londra: Cassell, Petter & Galpin. https://www.flickr.com/photos/britishlibrary/11226480883/
Superare l'evanescenza del parlato: lo sforzo può comportare lo sgomento riflesso nel volto dello scriba di fronte ai modi di parlare di personaggi tanto diversi.

SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

BERGAMO UNIVERSITY PRESS
sestante edizioni

Direttore responsabile
Prof. Giuliano Bernini

Biblioteca di Linguistica e Filologia

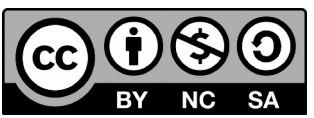
6.

Superare l'evanescenza del parlato
Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

Questo volume è stato stampato con il contributo del Dipartimento
di Lingue, Letterature e Culture Straniere dell'Università degli Studi di Bergamo.

Contributi rivisti dai curatori.



Indice

Introduzione

La disponibilità di attrezzatura informatica e di programmi facilmente accessibili e utilizzabili tramite la rete telematica globale comporta anche la possibilità di fissare su supporto digitale i messaggi linguistici trasmessi naturalmente da locutore a interlocutore attraverso segnali fonico-acustici, la cui realtà fisica in termini di onde sonore svanisce in frazioni di secondo pur non impedendo la loro ricezione e decodificazione tramite l'apparato uditivo del ricevente.

La fissazione su supporto digitale di segnali acustici evanescenti permette poi di ritrasmetterli e riascoltarli quante più volte si voglia e di rappresentarli in un sistema grafico specializzato come quello dell'alfabeto fonetico internazionale (AFI o IPA) o in altri sistemi meno sofisticati ma più adatti alla fruizione che di quei dati si prevede per scopi di ricerca diversi.

Quest'operazione, l'operazione di trascrizione, è un'operazione complessa in cui interagiscono diversi fattori che condizionano la fedeltà del risultato alla realtà fisica che si intende rappresentare. Il fattore più rilevante è rappresentato dallo stesso operatore incaricato della trascrizione, o più precisamente dal suo apparato uditivo, che è il tramite di decodificazione del segnale acustico la cui rappresentazione è quindi – giocoforza – quella di segnali "ascoltati" e non "detti". La terminologia corrente, p.es. "trascrizione di parlato", è il frutto di una metonimia comoda, ma che non deve oscurare la realtà di fatto.

Un secondo fattore rilevante è rappresentato dai segnali trasmessi all'apparato uditivo del trascrittore, che sono condizionati fisicamente dalla strumentazione di registrazione digitale utilizzata e possono condizionare la loro ricezione, decodificazione e, in ultima istanza, rappresentazione. Questi segnali, e siamo al terzo fattore rilevante, sono il prodotto di attività di locuzione di parlanti umani la cui articolazione fonetica è strettamente individuale in termini di caratteristiche della

voce (si pensi solo alle diverse frequenze della voce in relazione all'età e al sesso del parlante), competenza nella lingua in cui si esprimono, rapporto con l'interlocutore o gli interlocutori, scopi del messaggio così formulato.

Un ultimo fattore, esterno all'operazione di trascrizione e di ruolo ausiliario, è poi rappresentato dalla strumentazione digitale che permette l'analisi celere e precisa dei segnali acustici nelle loro dimensioni segmentale e prosodica e mette a disposizione un ambiente di "scrittura", che aiuta l'operatore nel controllo dell'interpretazione e della codificazione di quanto "ascolta" e trascrive.

Il complesso di fattori così sintetizzato è affrontato da diverse angolature in questo volumetto, che si propone come *vademecum*, guida per chi affronta un compito di trascrizione di parlato "ascoltato" per la prima volta o vuole approfondire le singolarità che incontra l'operazione di trascrizione in funzione del parlante e del contesto in cui sono stati prodotti.

In questa prospettiva, il *vademecum* è aperto da un contributo di Lorenzo Spreafico che offre una panoramica anche di ordine storico sulla strumentazione disponibile per la trascrizione e il ruolo che questa può e deve avere rispetto a quello del trascrittore umano per il fine di ottenere una trascrizione il più possibile fedele alla realtà fisica del segnale articolato e udito. Il contributo illustra i problemi teorici, oltre che metodologici, che l'adozione di strumentazione comporta in relazione alle finalità specifiche della linguistica.

Il secondo e il terzo contributo trattano due aspetti relativi alla natura dei segnali acustici e della loro rappresentazione in funzione del parlante che li produce. Nel secondo capitolo Alessandro Vietti prende in considerazione la variabilità intrinseca nei dati di parlato, mostrandone l'estrema rilevanza per capire il funzionamento della componente fonologica a livello di sistema e i meccanismi di produzione e percezione dei suoni del linguaggio. Le sue argomentazioni precisano sul fronte teorico il rapporto tra la disponibilità di strumentazione per l'esatta misurazione dei fenomeni acustici da una parte e, dall'altra, l'individuazione di ciò che è pertinente l'attività cognitiva e sociale del parlante e, in ultima istanza, la comprensione della grammatica del parlato. Nel terzo capitolo, Cinzia Avesani e Barbara Gili Fivela affrontano l'intonazione, fornendo una panoramica critica e ben ancorata empiricamente della metodologia adottata per descriverla e rappresentarla. Come già nel secondo capitolo per il livello segmentale, anche per il livello prosodico si auspi-

ca l'integrazione del livello fonologico e fonetico per ben comprendere i processi di costituenza e prominenza sia presso produzioni di parlanti nativi continui, sia in quella di parlanti una lingua per loro seconda.

Ai problemi di analisi e rappresentazione del parlato non-nativo sono dedicati i due contributi seguenti. Nel quarto capitolo, Sandra Benazzo e Marzena Watorek[1] illustrano le problematiche relative alla trascrizione da adottare nei *corpora* di francese lingua seconda. La scelta di questa lingua è particolarmente significativa per il fatto che la grafia standard del francese, oltre a riprodurre elementi di morfosintassi non presenti nel parlato, presenta una gamma di desinenze verbali diverse di una stessa resa fonetica, ovviamente l'unica a cui l'apprendente di lingua seconda è esposto nell'input. Cruciale è dunque l'adozione di forme di rappresentazione che non siano anche interpretazioni ingiustificate della struttura grammaticale della varietà di apprendimento effettivamente rappresentata, come si mostra nelle scelte operate in diversi *corpora* anche in funzione di diversi orientamenti teorici. Sempre il francese lingua seconda è la base empirica a cui fa riferimento Fabian Santiago nel quinto capitolo, dedicato ai problemi di trascrizione del livello prosodico. Il centro dell'attenzione è posto sugli aspetti metodologici che impone la natura della prosodia per il tipo di dati da raccogliere (spontanei, guidati), per la loro rappresentazione e infine per le procedure di annotazione che li possano rendere comparabili con dati prodotti da parlanti nativi di francese.

Nel penultimo capitolo, il sesto, Luciano Romito considera il prodotto di processi di trascrizione dal punto di vista del loro utilizzo in ambito forense, dove la fedeltà della rappresentazione è spesso drammaticamente cruciale nei procedimenti giudiziari. Nel capitolo si considera la trascrizione a partire da quanto il Codice di Procedura Penale prevede per i risultati di perizie da considerare nel dibattimento e mostra l'intreccio tra la padronanza di tecniche di trascrizione e verbalizzazione e ulteriori competenze, di natura interdisciplinare, che costituiscono il profilo professionale di un "trascrittore forense", una figura di cui si auspica il riconoscimento anzitutto nella formazione universitaria.

La natura di *vademecum* del presente volumetto, consistente nella trattazione aggiornata della gamma di problematiche teoriche e metodologiche nei primi sei capitoli, è consolidata nel capitolo finale, il

[1] Nella grafia originale: Wątorek [maˈzɛ̨na vɔ̃ʷ̃ˈtɔrɛk]

settimo, di ordine applicativo, nel quale Jacopo Saturno introduce all'utilizzo di risorse elettroniche ad accesso libero per l'operazione di trascrizione, con particolare riguardo del programma ELAN, il più sofisticato per le procedure che permettono di tenere conto della diversa natura dei dati di partenza e dei fattori che ne condizionano la rappresentazione trascritta, dando modo sia di verificare autonomamente le proposte dei capitoli precedenti, sia di avviare un *iter* di auto-addestramento al difficile compito del trascrittore.

Il volumetto è plurilingue. La presenza di italiano in quattro capitoli e di francese in due capitoli riflette il seminario all'origine di questa pubblicazione, "Superare l'evanescenza del parlato. Metodi e tecniche di trascrizione e annotazione", tenuto presso il Dipartimento di Lingue, letterature e culture straniere dell'Università di Bergamo nei giorni 8-10 novembre 2018. Oltre che da questo dipartimento nell'ambito del "Progetto di eccellenza", il seminario ha goduto del finanziamento dell'Università italo-francese (1° bando Label scientifico 2018 III; Decreto Rep.18/2018 del 19.7.2018). La lingua inglese scelta per il terzo capitolo riflette invece il suo utilizzo come lingua veicolare della produzione scientifica e l'opportunità di collegamento con una platea di lettori più vasta. L'estesa comunanza delle tre lingue per effetto della derivazione dal latino da una parte e della penetrazione del francese Oltre Manica dall'altra garantiscono comunque una generale possibilità di intercomprensione attraverso tutte le parti del *vademecum*, che potrà stimolare a fissare con piena professionalità dati linguistici di natura per sé evanescente.

Al Dipartimento di Lingue, letterature e culture straniere dell'Università di Bergamo va infine la gratitudine dei Curatori per il finanziamento della composizione e della stampa del *vademecum*.

GB AV JS LS

Lorenzo Spreafico
(Università degli studi di Bergamo)

La trascrizione strumentale del significante: dalle origini alle *digital humanities*[1]

1. *Introduzione*

La trascrizione è la pratica scientifica di trasposizione dei segni linguistici dal canale fonico acustico al canale grafico visivo al fine di facilitarne l'analisi.

Nella maggior parte dei casi la trascrizione prevede che vi sia una fase di ascolto della produzione verbale (o della sua registrazione), seguita da una fase di scrittura in cui si ricorre a segni grafici riconducibili a una ortografia, una ortografia modificata (Edwards 1992), o una scrittura dedicata come quella di un alfabeto fonetico (Wells 2006). Questa fase di ascolto comporta il ricorso ad un senso – quello dell'udito – che nel corso dell'evoluzione umana si è preadattato[2], vale a dire nel volgere di alcuni milioni di anni si è modificato e ottimizzato per assolvere tanto a una funzione uditiva – in particolare quella di localizzare nello spazio le fonti di rumore, così come fatto da tutti i mammiferi viventi (Frings & Müller 2014: 204) – quanto a una funzione percettiva – in particolare quella di discriminare, segmentare, normalizzare e categorizzare i segnali acustici (Raphael 2005). Tale funzione determina l'attivazione di processi di analisi e interpretazione del segnale via via più specializzati a seconda della natura dell'input acustico ricevuto e delle fina-

[1] La redazione di questo contributo è stata resa possibile grazie ai fondi del Dipartimento di Eccellenza di Lingue, Letterature e Culture Straniere dell'Università degli Studi di Bergamo.

[2] Uso qui il termine come traducente dell'inglese *exaptation*, talvolta reso in italiano con exattamento o esattamento, cfr. Simone (2012).

lità dell'ascolto[3]. Nel loro insieme questi processi definiscono la catena psicofisica della percezione uditiva e sono alla base delle differenze tra l'ambiente fisico, dove si generano le onde sonore (stimolo distale); quello sensoriale, dove il segnale sonoro è trasdotto in segnale elettrico (stimolo prossimale); e infine quello psichico, dove gli stimoli distali e prossimali sono elaborati e categorizzati (percetto).

Nella quotidianità della comunicazione parlata, la distinzione tra l'ambiente fisico dell'articolato e l'ambiente psichico del percepito è irrilevante, perlomeno fintanto che gli interlocutori riescano a interagire scongiurando fraintendimenti (Dascal 1999; Pouplier & Goldstein 2005). Tuttavia, per la pratica scientifica della trascrizione, la distinzione tra i due ambienti è epistemologicamente rilevante, perché la differenziazione tra stimoli distali e percetti rimanda alla controversia sull'adeguatezza dei sensi umani per giungere alla comprensione del mondo.

Semplificando molto[4], il dibattito riguarda sia la capacità dei sensi di cogliere tutte le proprietà degli stimoli distali, sia la possibilità che durante il processo di elaborazione degli stessi si generino – in un qualche punto della catena psicofisica della percezione – delle rappresentazioni fuorvianti o non veritiere, come per esempio quando si diano illusioni acustiche[5].

Per tali motivi la discussione sul potenziale dei sensi tutti per la descrizione del mondo ha alimentato una feconda discussione (Lyons 2016), che fra la seconda metà del XVI e la fine del XVII secolo si è arricchita di una prospettiva ancora oggi vitale, quella promossa dagli artefici della cosiddetta "rivoluzione scientifica" (Henry 1997).

In quel contesto, oltre all'affermarsi del metodo sperimentale per lo studio dei fenomeni naturali e della matematica per la loro interpretazione, si è affacciata l'idea che i limiti dei sensi potessero essere superati ricorrendo ad ausili vari che ne potenziassero le capacità, permettendo

[3] Pattamadilok *et al*. (2010) discutono ad esempio di come l'alfabetizzazione influisca sulla percezione del parlato.

[4] Sia permesso rimandare a Spreafico (2020) per ulteriori osservazioni al riguardo.

[5] Le illusioni acustiche sono state indagate in profondità con riferimento agli stimoli musicali, meno in relazione a quelli linguistici. Per una discussione recente cfr. Deutsch (2019).

così di accedere a realtà presenti ma non sensibili[6]; oppure ne oggettivizzassero le percezioni[7].

Seppur parecchi anni dopo, l'idea che i dati sensoriali potessero essere amplificati e oggettivizzati ricorrendo a congegni dedicati si diffuse anche tra gli scienziati interessati alle lingue e al linguaggio, in particolare tra i fonetisti che per primi promossero un approccio metrologico allo studio del parlato, ovvero una sua caratterizzazione basata sull'uso di strumenti e misurazioni e non su percetti e impressioni degli investigatori. Questo approccio, che riguardò dapprima l'indagine della dimensione articolatoria del significante (§2), quindi di quella acustica (§3)[8], è all'origine della trascrizione strumentale del parlato, la cui storia, caratteristiche e usi discuterò brevemente.

2. *Strumenti per indagine articolatoria*

2.1 Le origini

L'approccio strumentale e metrologico al parlato assunse il carattere di sistematico programma di ricerca linguistica a partire dalla metà del XIX secolo quando, nel solco di una tradizione di marca fisiologicista[9] che andava affermandosi anche nelle scienze umane, venne promossa l'adozione di prospettive di laboratorio[10] per lo studio delle facoltà degli individui, inclusa quella di linguaggio.

---

[6] Come nel caso del microscopio, che consentì di avvicinarsi al microcosmo (Smith 2014).

[7] Come nel caso di "termometri e barometri che permisero di quantificare – e nel caso della pressione atmosferica anche di scoprire – grandezze che sino ad allora erano rilevabili soltanto qualitativamente" (Brenni 2013).

[8] Il significante non grafico-visivo dei segni linguistici è in realtà più complesso e può essere indagato strumentalmente con riferimento perlomeno alle dimensioni della neurofonetica (Hertrich & Ackermann 2013), articolatoria, aerodinamica, acustica, e percettiva. Ciononostante, in questo contributo tratterò unicamente della dimensione articolatoria e di quella acustica, perché queste sono quelle cui più comunemente ci si riferisce quando si elabori una trascrizione strumentale.

[9] La marca fisiologicista è stata definita da Brain (2015) come "estetica fisiologica di fine secolo". Per un inquadramento delle ragioni e del contesto culturale in cui è maturata la fonetica strumentale, in particolare la contrapposizione tra la scuola filologica tedesca e l'emergente scuola linguistica francese, può risultare utile anche Brain (1998).

[10] Per la discussione del ruolo dei laboratori nelle scienze tutte si rimanda a James (1989). Per la discussione del concetto di laboratorio come metodo più che come luogo della ricerca fonetica e fonologica, si rimanda invece ai lavori di Pierrehumbert *et al.* (2000) e di Beckman & Kingston (2011, edizione rivista dell'originale del 1990).

In particolare, la rivoluzione fu sostenuta a partire dal 1874 dall'azione combinata di Michel Bréal e Leon Vaïsse, rispettivamente segretario e presidente della *Société de Linguistique de Paris* (SLP), che intendevano promuovere una scuola linguistica francese da contrapporre a quella tedesca, anzitutto abbandonando la strada delle "speculazioni filologiche avventate" (Brain 1998: 258) che, a loro dire, caratterizzavano quest'ultima, soprattutto in conseguenza dell'essere fondate su dati di lingua scritta.

A dimostrazione di ciò, nel suo discorso di insediamento da nuovo presidente della SLP nel 1875, Vaïsse annunciò l'avvio di un programma di ricerca volto a incoraggiare lo studio delle componenti fisiche delle lingue e del linguaggio (Vaïsse 1875). Proprio a tal fine egli promosse la collaborazione tra alcuni membri della SLP e il fisiologo Etienne-Jules Marey. Marey – che dì lì a poco avrebbe dato alle stampe il volume *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine* (1885) in cui riassumeva e sistematizzava oltre quarant'anni di ricerche nel campo – era il principale rappresentante francese delle cosiddette tecniche di "iscrizione" (Clarke & Henderson 2002), vale a dire della rappresentazione grafica dell'andamento di alcune funzioni corporee quali il battito cardiaco, la ventilazione polmonare, il movimento degli arti. Marey accettò di buon grado la proposta di collaborazione avanzata da Vaïsse (Teston 2004), così che in seno alla SLP venne fondato un gruppo di ricerca che si diede quale primo obiettivo la descrizione delle interazioni tra i movimenti della cassa toracica, dei muscoli laringei, delle labbra e dell'aria in transito nelle cavità nasali, il tutto alla luce di dati strumentali. Per riuscire nell'impresa i membri del gruppo costruirono un poligrafo, ovvero uno strumento capace di elaborare tracciati relativi a diversi parametri fisiologici implicati nella produzione del parlato a partire da quanto colto da sensori di diversa natura.

Il poligrafo dedicato allo scopo costituiva una evoluzione del chimografo[11], il dispositivo presentato nel 1840 dal fisiologo tedesco Carl Ludwig (Holmes 2003) che consentiva di trasdurre in movimenti i più svariati fenomeni fisiologici e, quindi, di raffigurarli su un supporto fisico grazie a punteruoli o pennini vincolati a sensori messi a contatto con

[11] Oggigiorno il termine viene talvolta usato per indicare uno strumento dedicato esclusivamente allo studio dei cambiamenti di flussi di aria in transito orale o nasale.

il corpo umano che tracciavano segni su fettucce di carta avvolte su tamburi mossi da meccanismi a orologio, così da pantografare le variazioni nel tempo delle grandezze fisiche colte[12]. Più precisamente, il poligrafo proposto da Marey e colleghi era la sintesi di chimografi che rilevavano quattro parametri: le variazioni di volume della cassa toracica; le variazioni di pressione nelle cavità nasali; le vibrazioni delle corde vocali; i movimenti delle labbra (Brain 1998)[13].

Sebbene gli stessi ideatori del poligrafo mettessero in guardia dal fatto che sia la sensibilità dei sensori utilizzati, sia la frequenza di campionamento del loro strumento potessero essere insufficienti per giungere a una descrizione accurata del parlato (Tillmann 2006), tali apparecchiature ebbero un grande successo e si diffusero tanto in Francia quanto nel resto d'Europa, dove molti altri ricercatori si adoperarono per perfezionarle e per elaborare nuove forme di rappresentazione grafica dei parametri monitorati, così da rendere conto del simultaneo movimento degli articolatori nello spazio e nel tempo[14].

La collaborazione tra Marey e la SLP proseguì a lungo e culminò nel 1897 con la fondazione di un laboratorio di fonetica sperimentale presso il *Collège de France* la cui guida venne affidata a Jean-Pierre Rousselot, che già aveva collaborato con il gruppo di Marey. Nei suoi lavori, Rous-

[12] Il ricorso a strumentazione per la descrizione e l'analisi del parlato non è una innovazione di Marey, perché già altri prima di lui avevano sviluppato strategie per ricavare informazioni articolatorie. Tra gli altri, Erasmus Darwin, nonno di Charles Darwin, che nel 1803 propose di indagare l'articolazione delle vocali inserendo nella bocca dei parlanti dei fogli di stagno che, messi a contatto con la lingua, venivano deformati e plasmati in forme diverse a seconda della vocale pronunciata (Tillmann 2006: 377). Tutta da attribuirsi a Marey fu però l'idea di ricorrere a strumenti che rendessero la dimensione continua e dinamica del parlato, sino a quel momento trascurata in favore di quella statica e discreta indotta dal primato, allora ancora saldo nelle scienze del linguaggio, dello scritto sul parlato.

[13] Le prime due informazioni venivano rilevate ricorrendo a uno pneumografo e a un manometro che generavano due tracce distinte, ad andamento orizzontale quando il torace non si muoveva o dal naso non fuoriusciva aria, oppure parabolico laddove vi fossero stati movimenti del torace o flussi d'aria. Le vibrazioni delle corde vocali venivano invece rilevate grazie a un galvanometro che, in presenza di vibrazioni delle corde vocali, generava una traccia tremolante. Infine, i movimenti delle labbra venivano rilevati tramite un pantografo a molla che con le labbra serrate tracciava una linea, mentre al loro allontanarsi o avvicinarsi riportava due curve divergenti o convergenti.

[14] L'Università tecnica di Dresda (*TU Dresden*) conserva molte di queste apparecchiature nella sua *Historische akustisch-phonetische Sammlung* (HAPS).

selot promosse l'uso sistematico di chimografi e poligrafi per la costruzione di tracciati contenenti informazioni sulla produzione del parlato, che servissero tanto per la documentazione di lingue e, in particolare, la varietà galloromanza da lui parlata (Rousselot 1891a[15]), quanto per la risoluzione di quesiti fonetici teorici, per esempio sulla lunghezza dei foni (Rousselot 1891b). L'applicazione dei vecchi strumenti per nuove finalità ingegnata da Rousselot sancì da un lato la nascita di una nuova disciplina – la fonetica sperimentale – in cui l'indagine degli aspetti fisici del parlato mirava a guadagnare conoscenze linguistiche e non, come fino a quel momento usuale, fisiologiche; dall'altro favorì l'adozione di un nuovo metodo di indagine – quello della trascrizione strumentale – in cui la trasposizione del parlato (o meglio della sua *immagine acustica*, per dirla con De Saussure 1916, che di Bréal in quegli anni fu collaboratore) non era più affidata ai sensi e alle impressioni dei ricercatori, bensì a dispositivi dedicati.

2.2 Le evoluzioni

Sebbene chimografi e poligrafi abbiano costituito le prime attrezzature per la documentazione e la trasposizione strumentale strumentale del parlato, oggi non vengono più impiegati se non per finalità illustrative. Infatti, nel corso del tempo gli strumenti della fonetica si sono evoluti così da rispondere ai bisogni dei ricercatori e permettere di ovviare a una serie di problemi divenuti sempre più pressanti, pena l'impossibilità di produrre nuove conoscenze. Il principale di questi problemi era – e in parte ancora è, come discusso in Stone (2010; 2013) – quello di riuscire a rilevare simultaneamente le proprietà di articolatori molto diversi tra loro, dato che il parlato è il risultato del fine coordinamento spaziale e temporale di più organi, solo alcuni dei quali accessibili dall'esterno del cavo fonoarticolatorio e, quindi, monitorabili direttamente con sensori ottici (per esempio le labbra, Krause *et al*. 2020) o indirettamente con sensori di superficie (per esempio le corde vocali, tramite laringografo, Rothenberg 1992). La maggior parte degli organi di interesse è infatti nascosta e inaccessibile, dunque il loro movimento può essere

[15] La versione digitalizzata del lavoro rintracciabile alla pagina web https://gallica.bnf.fr consente di apprezzare le tracce dei chimografi e dei poligrafi usati da Rousselot.

tracciato solo invasivamente procedendo dall'interno (per esempio, per il palato, tramite elettropalatografo) o dall'esterno del tratto fonoarticolatorio, ma solo utilizzando sistemi diagnostici per immagini.

Tra gli altri problemi – invero non meno rilevanti – che affliggono i fonetisti strumentali vi sono poi: la necessità di visualizzare tessuti dalle proprietà fisiche eterogenee e non comparabili, ad esempio ossa (tracciabili ricorrendo a raggi X – cfr. Sock *et al*. 2011 – ma non a ultrasuoni) e muscoli (visualizzabili con ultrasuoni – cfr. Stone 2005 – o misurabili tramite elettromiografia – cfr. Stepp 2012); la necessità di tracciare lo spostamento di articolatori che si muovono con velocità e gradi di libertà molto diversi tra loro, come nel caso del velo, che ha molti vincoli assiali, o della lingua, che ne presenta assai meno (Badin & Seurrier 2006); infine, la necessità di adottare strumenti minimamente invasivi dell'articolazione, così da scongiurare che il sistema di osservazione modifichi ciò che viene osservato.

Per cercare di risolvere i problemi appena elencati, nel corso degli anni sono stati sviluppati strumenti – spesso adattando tecnologie elaborate in seno all'ingegneria biomedica – costituiti da trasduttori capaci di convertire grandezze fisiche in segnali elettrici contenenti informazioni continue o discrete spendibili anche per l'analisi fonetica[16]. Tra gli strumenti di maggior successo e pertanto più interessanti in prospettiva trascrittoria, almeno tre meritano di essere menzionati: l'elettrolaringografo, l'elettropalatografo e l'articulografo[17]. L'elettrolaringografo è costituito da una coppia di elettrodi da applicare sulla cute in corrispondenza della cartilagine tiroidea, così da ottenere informazioni sul contatto tra le corde vocali e raffigurate come forma d'onda in cui a ciascun picco corrisponde il massimo contatto tra le pliche (Herbst *et al*. 2010). L'elettropalatografo è costituito da un palato artificiale dentro cui sono affogati degli elettrodi che permettono di ottenere informazioni sul contatto tra la lingua e la volta palatina – dunque sul punto di articolazione[18] –

[16] Per una rassegna delle più recenti tecniche cfr. Kochetov (2020a, 2020b) e la s*pecial collection* della rivista *Laboratory Phonology* su *Techniques and Methods for Investigating Speech Articulation* curata da Spreafico & Vietti (2020).

[17] Trascuro qui tecniche quali elettromiografia, pletismografia, pneumotacografia, perché poco diffuse nei laboratori, soprattutto italiani.

[18] Ricorrendo all'elettropalatografo è teoricamente possibile osservare i contatti linguali per almeno sette delle diciassette regioni articolatorie proposte in Ladefoged & Maddieson (1996: 15).

che sono rappresentate discretamente per ogni elettrodo disponibile (Roach & Hardcastle 1976). Infine, l'articulografo è un macchinario che utilizza campi magnetici per tracciare lo spostamento di piccole bobine metalliche posizionate sugli articolatori, permettendo così di ottenere informazioni sulla loro posizione nelle tre dimensioni dello spazio raffigurate come tracciati nel piano cartesiano o euclideo (Perkell 1992).

Sebbene questi tre strumenti abbiano contribuito a migliorare la comprensione di numerosi fenomeni fonetici fondamentali, attualmente la massima espressione della tecnologia strumentale per indagini articolatoria – sia con riferimento alla complessità delle apparecchiature, che alla capacità di rispondere ai *desiderata* dei fonetisti – è rappresentata dalle tecniche diagnostiche per immagini di derivazione medica, in particolare l'ecografia e la tomografia a risonanza magnetica, entrambe innocue e poco invasive del parlato. L'ecografia sfrutta gli ultrasuoni, per lo più per documentare forma e posizione della lingua (Stone 2005). La tomografia sfrutta invece i campi magnetici per visualizzare le strutture anatomiche rigide e molli del cavo orale (Bresch *et al*. 2008), permettendo così di rilevare i movimenti degli articolatori simultaneamente e con risoluzioni temporali e spaziali adeguate per le finalità della fonetica.

Per quanto storicamente prioritario e di grande utilità per la comprensione del funzionamento dei sistemi di significazione umana, non vi è dubbio alcuno che lo studio dell'articolazione non esaurisca quello del significante dei segni linguistici parlati. Prova ne è che inizialmente anche i promotori dell'approccio articolatorio avessero quale loro primo obiettivo non tanto il documentare i movimenti dei diversi organi, quanto l'identificare un equivalente dei fenomeni acustici che potessero fissare per consentire una analisi empirica del parlato, analogamente a quanto perseguito in fisica sperimentale per esempio da Helmholtz (Kursell 2013).

Tuttavia, prima che la fissazione strumentale della dimensione acustica del parlato potesse concretizzarsi sino a diventare, come è oggi, quasi irrinunciabile per la trascrizione del significante fonico-acustico e dei significati ad esso associati, fu necessario si affermasse un altro strumento capace di rivoluzionare tanto l'analisi linguistica, quanto la cultura umana: il registratore audio.

3. *Strumenti per indagine acustica*

3.1 Le origini

Se le prime forme di indagine articolatoria strumentale del parlato si rintracciano negli anni Settanta del XIX secolo (§2.1), le prime ricerche strumentali sulla dimensione acustica datano agli anni Cinquanta dello stesso secolo, ma si consolidano e diffondono solo dallo stesso decennio del secolo successivo.

Infatti, nel 1857 Scott de Martinville brevettò il fonautografo, il primo strumento per la registrazione della voce (Feaster 2010). Questo apparecchio, concettualmente simile a un chimografo, sfruttava un corno alla cui estremità era posizionato un diaframma che vibrava quando colpito da onde sonore. Al diaframma era poi fissata una spazzola che trasferiva le vibrazioni verso un cilindro rotante rivestito di carta affumicata. Questa spazzola rimuoveva il nerofumo dalla bobina di carta trasponendo così le onde sonore – udibili ma invisibili per l'essere umano – dando loro forma di tracce con andamento ondulatorio eventualmente spendibili per l'analisi, in particolare con riferimento agli inviluppi di ampiezza e forma d'onda. Nonostante le forti restrizioni di frequenze campionabili conseguenti ai limiti di sensibilità del diaframma vibrante, queste tracce rappresentano – di fatto – la prima forma di trascrizioni acustica strumentale del parlato.

Purtroppo, il fonautografo non permetteva la riproduzione[19] del suono registrato[20]. Perché ciò diventasse possibile fu necessario attendere una decina di anni (1876) e l'invenzione del fonografo da parte di Thomas Edison (Israel 1998). A differenza del fonautografo, il fonografo era dotato di un cilindro ricoperto da una sottile lamina di stagno che

[19] La riproducibilità del registrato costituiva a quel tempo un limite avvertito ma insuperabile, soprattutto ricorrendo agli strumenti per l'indagine articolatoria. Il problema consisteva nell'impossibilità di costruire un apparato fonatorio artificiale in cui gli articolatori venissero messi in movimento da aste che ripercorressero le tracce elaborate per esempio dai poligrafi. Ciò non era possibile nemmeno ricorrendo ai sistemi allora già disponibili per la sintesi vocale, come la famosa macchina di von Kempelen introdotta nella seconda metà del XVIII secolo e controllata manualmente (Barry & Trouvain 2011).

[20] Ma si consulti l'affascinante percorso di ricerca documentato in http://www.firstsounds.org e mirato a rintracciare fonoautogrammi ancora disponibili e a convertirli in voci e suoni grazie all'impiego di tecniche di informatica umanistica.

poteva essere fatta rototraslare grazie a una manovella. Durante la fase di registrazione la lamina veniva incisa da un punteruolo fissato a una membrana che – come già nel fonautografo – raccoglieva le onde sonore. Invece, durante la fase di riproduzione il solco elicoidale veniva ripercorso dal punteruolo che rimetteva così in vibrazione la membrana cui era vincolato generando un suono e consentendo quindi, per la prima volta nella storia dell'umanità, la riproduzione di quanto registrato.

Quella del fonografo fu una invenzione rilevante anche per il progresso delle tecniche di analisi del significante, perché i cilindri di stagno su cui il parlato veniva registrato poterono essere sfruttati come fossero trasposizioni strumentali del parlato, per esempio indagando l'andamento e la forma dei solchi incisi sulle lamine. Per un breve periodo, questo approccio, introdotto da Ludimar Hermann – che nel 1894 sfruttò un microscopio per studiare le tracce lasciate sui cilindri – e rielaborato da Hector Marichelle – che nel 1897 fotografò, ingrandì, stampò e analizzò le incisioni – costituì la modalità privilegiata per l'analisi acustica del parlato (Marage 1898[21]).

Tuttavia, l'invenzione del fonografo – e quindi degli altri strumenti di registrazione magnetica e digitale del parlato che negli anni si susseguirono – rivoluzionò anche la pratica della trascrizione impressionistica, da quel momento sempre più strumentalmente basata. Infatti, le attrezzature per la registrazione consentirono da un lato di superare i limiti della memoria fonologica (Perrachione *et al*. 2017) che rendono le trascrizioni in tempo reale ricchissime di errori, dunque inaffidabili (Amorosa *et al*. 1985); dall'altro di verificare ogni proposta di trascrizione del parlato alla luce di un terzo di comparazione invariabile, la registrazione, riducendo così il disaccordo tra trascrittori (Shriberg & Lof 1991).

Tuttavia, la trascrizione strumentale del significante fonico-acustico si affermò solo molti decenni dopo l'invenzione del fonografo grazie all'imporsi sul mercato di uno strumento rivoluzionario capace di offrire una nuova modalità di visualizzazione del segnale acustico registrato: lo spettrografo sonoro.

[21] La versione digitalizzata del lavoro rintracciabile alla pagina web https://www.persee.fr/doc/psy_0003-5033_1898_num_5_1_3053 consente di apprezzare la natura delle tracce analizzate da Hermann e Marichelle.

3.2 Le evoluzioni

La rilevanza dello spettrografo sonoro per gli studi di fonetica sperimentale e sulla trascrizione è tutta nella sua funzione, perché lo strumento permette di generare uno spettrogramma che riproduce visivamente la distribuzione dell'energia contenuta nel parlato. Lo spettrografo permette infatti di operare un'analisi del segnale acustico nel tempo e per frequenza e di darne una raffigurazione tridimensionale mostrando simultaneamente le diverse informazioni e riportando su un primo asse il passare del tempo; su un secondo asse la frequenza dei suoni per ciascuno dei punti campionati; su un terzo asse – che nella rappresentazione cartesiana può essere sostituito da una scala di colori – l'ampiezza dei suoni in ciascuno dei punti di intersezione tempo/frequenza. Questa raffigurazione delle informazioni – probabilmente la più influente che la storia dell'elaborazione e analisi dei segnali acustici abbia sinora conosciuto – permette di distinguere e classificare ciascuno degli elementi sonori della lingua parlata che ricada nell'intervallo di sensibilità dello strumento.

Sviluppati nel quadro di programmi di ricerca umanitari[22] e militari nei tardi anni Quaranta del XX secolo (Shankweiler & Fowler 2015)[23], fino alla prima metà degli anni Ottanta gli spettrogrammi venivano generati da spettrografi che analizzavano il segnale sfruttando filtri elettromeccanici analogici, che però peccavano in rapidità e accuratezza di analisi rispetto a quelli elettronici attuali (Farmer 1997). Successivamente gli sviluppi tanto nel campo dell'elaborazione digitale del segnale, quanto in quello della sua analisi numerica, favorirono la trasformazione radicale del processo di produzione degli spettrogrammi che, difatti, oggi vengono elaborati pressoché solo matematicamente ricorrendo alla trasformata di Fourier computata per ciascun istante campionato

[22] Finalizzati a ottenere una macchina che leggesse testi scritti a non vedenti.

[23] Per la precisione, il primo modello di spettrografo fu presentato nel 1939 presso i *Bell Telephone Laboratories* di Murray Hill (New Jersey). Tuttavia, a causa del potenziale bellico dello strumento, il prototipo venne tenuto nascosto sino alla fine della seconda guerra mondiale, quando nel 1946 i piani per la sua costruzione e le possibili applicazioni furono pubblicati nel *Journal of the Acoustical Society of America* (JASA) e resi disponibili all'intera comunità scientifica. Un affascinante riassunto della storia dello spettrogramma è offerto da Shankweiler & Fowler (2015).

del segnale sebbene sarebbe possibile generarli anche ricorrendo ad altri criteri matematici[24].

Ciò permette di concludere questa sezione del contributo osservando quanto la trasposizione strumentale della dimensione articolatoria e di quella acustica siano tra loro teoricamente differenti. Infatti, mentre nel caso del dato articolatorio la trasposizione ha come sua funzione quella di svelare e fissare ciò che, pur ricadendo nella sfera del visibile, è semplicemente inaccessibile perché si muove in spazi o con tempi che l'occhio umano non può cogliere; nel caso del dato acustico la trasposizione ha come sua funzione quella di assegnare una forma visibile a un fenomeno che in realtà sarebbe accessibile al trascrittore ricorrendo anche ad un senso dedicato, l'udito. Di conseguenza, mentre la trasposizione strumentale articolatoria serve anzitutto a fissare un invisibile percetto, la trasposizione strumentale acustica serve invece a dar forma diversa ad uno stimolo distale, trasformandolo da percetto uditivo a percetto visivo, così da consentire la gestione visuale delle numerose ma invisibili informazioni in esso contenute e che costituiscono la base della sua percezione, per esempio, frequenza, intensità, o durata dello stimolo.

4. *Dalla trasposizione strumentale alla trascrizione strumentale*

La comparsa di sistemi per l'indagine strumentale e metrologica del parlato in prospettiva articolatoria o acustica ha rappresentato un punto di svolta per l'analisi del significante fonico dei segni linguistici, sino a quel momento basato per lo più sull'introspezione (Kemp 2006) o su inferenze basate sull'indagine della lingua scritta[25]. Inoltre – e il breve excursus nella storia degli strumenti per la fonetica di laboratorio riportato nei paragrafi precedenti lo vorrebbe documentare – la comparsa di siste-

[24] Va notato che ciò obbligherebbe a una diversa organizzazione visiva delle informazioni, cosa che la comunità dei fonetisti pare poco disponibile a modificare (Fulop & Fitz 2006).

[25] Giova ricordare che per lungo tempo termini quali *lettera*, *suono*, *fono* o *fonema* sono stati impiegati in maniera intercambiabile. Per esempio, Jacob Grimm intitolò il primo capitolo della prima edizione (1822) della sua grammatica tedesca *Die Lehre von den Buchstaben*, modificandolo poi in *Lautlehre* a partire dalla edizione del 1840 (cfr. Haas 1990).

mi per l'indagine strumentale del parlato può essere fatta coincidere con la nascita della trascrizione strumentale del significante non grafico-visivo dei segni linguistici. Infatti, sin dall'ideazione delle prime attrezzature dedicate, la rilevazione strumentale del significante si è accompagnata con l'elaborazione di strategie di raffigurazione – dunque trasposizione – delle informazioni colte dai sensori, vuoi per renderle fruibili in tempi diversi da quello della produzione, vuoi per garantirne un'analisi replicabile. Per tale motivo, indipendentemente da come siano state generate, e da quale forma abbiano assunto, queste rappresentazioni vanno considerate come vere e proprie forme di trascrizione del parlato, seppur caratterizzate da proprietà che permettono di distinguerle da altre forme di trascrizione più diffuse, anzitutto quelle impressionistiche (Kemp 2006: 397) che sfruttano un sistema di rilevazione basato sui sensi e un sistema di notazione basato su un sistema di scrittura alfabetico, come nel caso dell'alfabeto fonetico internazionale (IPA 1999) che probabilmente costituisce la più nota tecnica linguistica di trasposizione del significante fonico-acustico.

La prima differenza[26] tra trascrizioni strumentali e impressionistiche risiede nella natura dei dati selezionati per la trasposizione. Infatti, le trascrizioni strumentali mostrano tutti e solo gli aspetti del segnale che gli strumenti utilizzati dall'elicitatore dei dati siano in grado di presentare in virtù del come siano stati progettati e indipendentemente dal fatto che quei segnali siano parte di un significante linguistico. Al contrario, le trascrizioni impressionistiche mostrano solo, e non necessariamente tutti, quegli aspetti del significante che siano rilevanti per la costruzione di contrasti linguisticamente pertinenti, perché elaborate impiegando l'apparato uditivo-percettivo calibrato sulla comprensione delle informazioni linguistiche. In tal senso si dà un non trascurabile problema di quantità e qualità dell'informazione per l'analisi del significante. Infatti, se è vero che le trascrizioni strumentali contengono solitamente più informazioni di quelle rintracciabili in una trascrizione impressionistica – fosse anche solo perché la capacità di campionamento, ovvero di conversione da continuo a discreto, del segnale da parte degli strumenti è superiore a quella umana – non è ugualmente vero che que-

[26] Baso i confronti sulle categorie presentate in Wells (2006) e rielaborate in Heselwood (2013).

ste siano tutte ugualmente servibili per una indagine linguistica. A titolo di esempio, si può osservare che l'analisi spettrografica del segnale audio di un dialogo – che costituisce una forma di trascrizione strumentale – riferirà di tutte le informazioni contenute nella registrazione, inclusi eventuali rumori ambientali, che evidentemente non contano come significanti di segni linguistici e verranno perciò ignorati o adeguatamente trattati dall'umano che produca una trascrizione impressionistica.

La seconda differenza tra le trascrizioni strumentali e quelle impressionistiche è che – per riprendere le categorie presentate in IPA (1999: 36) – le prime sono orientate all'emittente, mentre le seconde al ricevente. L'obiettivo principale delle trascrizioni strumentali è infatti quello di rendere conto di ciò che i parlanti articolino o pronuncino e non, come nel caso della trascrizione impressionistica, di ciò che gli ascoltatori percepiscono. È questa una differenza particolarmente significativa laddove per la trascrizione impressionistica si ricorra all'alfabeto fonetico internazionale, perché quest'ultimo presuppone che la forma trascritta sia comune a chi parla e a chi ascolta, ovvero che la trascrizione sia fonemica (IPA 1999: 27).

Diversamente, come osservato in Heselwood (2013), le trascrizioni strumentali sono trascrizioni fonetiche, dunque specifiche e capaci di restituire l'analisi di frammenti di significante facilmente identificabili e chiaramente associabili a un emittente nonché a un tempo e a un luogo di produzione; non sono invece spendibili per offrire una descrizione generale del parlato di un individuo o, meno ancora, di una comunità di individui indefinitamente ampia e dalla vaga o idealizzata caratterizzazione sociolinguistica. Peraltro, nel caso delle trascrizioni strumentali, la differenza può essere quantificata e riportata a un *continuum* che tenga conto dei valori assunti dai parametri delle impostazioni adottate per la rilevazione degli osservabili, ad esempio con riferimento alla densità del campionamento spaziale o alla frequenza di quello temporale, valori solitamente tra loro concorrenti quando si usano strumenti per indagine articolatoria.

La terza differenza tra trascrizioni strumentali e trascrizioni impressionistiche riguarda la natura delle raffigurazioni offerta da ciascuna tipologia di trasposizione. Infatti, mentre le trascrizioni strumentali sono analogiche, quelle impressionistiche sono discrete. In primo luogo perché le raffigurazioni prodotte dagli strumenti variano proporzionalmente al variare del fenomeno fisico rilevato dai loro sensori, di cui conser-

vano dunque le proprietà quantitative. Per esempio, al modificarsi della distanza tra le labbra, si modifica conformemente anche la distanza tra le tracce prodotte da un articulografo, e solo i limiti di sensibilità o elaborazione dei dati dell'apparecchiatura alterano il rapporto tra i due valori (Stella *et al*. 2012; Stella *et al*. 2013). In secondo luogo le trascrizioni strumentali sono analogiche perché producono rappresentazioni dinamiche, così come dinamici sono i fenomeni cui si riferiscono. Infatti, le qualità rilevate dagli strumenti vengono visualizzate senza che intervengano fenomeni di discretizzazione tra le entità che le costituiscono che assomiglino a quelli differenziali rintracciati negli ascoltatori umani (Albano Leoni & Maturi 2002: capitolo 4), ma solo quelli – invarianti nell'intero processo – dovuti al campionamento per la produzione del segnale per come reso possibile dall'apparecchiatura impiegata.

Diversamente, come anticipato, le trascrizioni impressionistiche conformi alle indicazioni dell'associazione fonetica internazionale sono rappresentazioni discrete. Anzitutto perché sono fondate su raffigurazioni simboliche degli osservabili, di cui non riportano dunque tutte le variazioni fisiche, bensì solo quelle percepite dal trascrittore. Poi perché i simboli alfabetici con cui sono elaborate sono distinti e immodificabili, pertanto inservibili a rendere la continuità dei fenomeni che si può osservare strumentalmente tanto all'interno di ciascun segmento ipotizzato, quanto tra segmenti. Per esempio, ricorrendo ad una trascrizione impressionistica conforme IPA non vi è alcun modo di segnalare la variazioni delle formanti acustiche rilevabili strumentalmente per il decorso di una stessa vocale o tra vocali in contesti differenti; così come nei segni dell'alfabeto non vi è alcuna relazione tra la larghezza di un glifo tipografico e la durata del segmento cui si riferisca, oppure tra la spaziatura tra due glifi e la dinamica di coarticolazione dei segmenti cui rimandano, perché entrambi i parametri sono definiti esclusivamente dalla natura proporzionale o non proporzionale del tipo di carattere (*font*) selezionato in sede di progettazione del sistema di notazione[27].

[27] Non sono a conoscenza di ricerche sulla percezione di trascrizioni fonetiche o fonologiche elaborate con caratteri tipografici tra loro differenti, ma ipotizzo che così come ve ne sono di emozionali per la scrittura (Juni & Gross 2008) e interpretative per la organizzazione e distribuzione spaziale dei turni nelle trascrizioni del discorso (Edwards 1993), ve ne possano essere anche per le notazioni fonetiche alfabetiche.

Un'ultima differenza tra le trascrizioni strumentali e quelle impressionistiche è che le prime riferiscono univocamente di un livello del significante, mentre le seconde possono essere impiegate per rimandare da uno a tutti i livelli di indagine disponibili. Per esempio, la lettura di una trascrizione strumentale elaborata ricorrendo a un elettropalatografo rimanda chiaramente al piano dell'articolazione, più precisamente a quello del contatto tra la lingua e il palato. Invece, la lettura di una trascrizione impressionistica in cui si faccia uso per esempio della notazione [r] può essere intesa riferirsi tanto alla dimensione articolatoria, quanto a quella aerodinamica, acustica, o percettiva del suono. Sebbene la possibilità di rendere conto di solo una delle dimensioni del significante possa in prima battuta far ipotizzare un limitato potenziale descrittivo e dunque una limitata spendibilità delle trascrizioni strumentali, va notato in primo luogo che anche la rilevazione di fenomeni assai circoscritti può innescare rivoluzioni conoscitive (come avvenuto per il tempo di attacco della sonorità, cfr. Cho *et al*. 2019); e in secondo luogo che la consapevolezza di quali siano i confini del rilevabile ricorrendo allo strumento garantisce certezza di quanto venga trascritto, dunque rigore metodologico.

Quest'ultima specificità delle trascrizioni strumentali, combinata con la vaghezza delle trascrizioni impressionistiche alfabetiche, ha fatto sì che nel tempo venisse accordata alle trascrizioni strumentali una crescente preferenza, anche per indagini non fonetiche.

5. *Dalla trascrizione impressionistica alla trascrizione strumentale*

Sebbene lo sviluppo e l'impiego di apparecchi tecnologici appositamente progettati per lo studio sperimentale del parlato risalga ai primi dell'Ottocento e si sia affermato tra la fine dell'Ottocento e l'inizio del Novecento, è solo a partire dalla seconda metà del XX secolo che questo approccio si è definitivamente imposto, soprattutto grazie alla maggiore diffusone degli strumenti. In particolare, la situazione ha cominciato a cambiare dopo la fine della seconda guerra mondiale quando, in virtù di una riduzione dei costi d'acquisto, il numero di spettrografi resi disponibili nei laboratori di fonetica aumentò tanto da consentire analisi acustiche – dunque trascrizioni strumentali – su larga scala. Tuttavia, è solo dalla fine degli anni Ottanta del secolo scorso che – soprattutto per effetto della rivoluzione digitale che ha interessato le scienze umane –

la tipologia e il numero di dispositivi per la ricerca fonetica strumentale è cresciuto, tanto che oggi la maggior parte dei ricercatori che volessero sfruttarli potrebbero farlo a fronte di investimenti ragionevoli[28].

La diffusione degli strumenti ha avuto diverse ricadute sulla metodologia della ricerca e la teoria linguistica, ovviamente soprattutto per le discipline della fonetica e della fonologia.

Le prime, scontate, conseguenze sono state rappresentate dalla pronta verifica empirica di svariate proposte speculative impossibili da falsificare ricorrendo alle sole risorse prima disponibili, vale a dire a propriocezione ed eterocezione; nonché l'elaborazione di nuove teorie linguistiche strumentalmente informate. Per esempio, grazie alle rilevazioni strumentali è stato possibile validare alcune delle prime ipotesi sulla coarticolazione (Kühnert & Nolan 1999); così come indagare quali aspetti del significante fossero dovuti a proprietà anatomofisiologiche e quali invece alla natura delle rappresentazioni linguistiche, tanto da consentire l'elaborazione di approcci incorporati (*embodied*) alla fonologia (Mompean 2014; Gick *et al*. 2019).

Una seconda ricaduta della diffusione degli strumenti – meno evidente, ma pertinente per la tematica di questo volume – ha riguardato invece la percezione e la fruizione delle trascrizioni impressionistiche. Infatti, sebbene ancora oggi queste trascrizioni identifichino la più diffusa tecnica di trasposizione del significante, e sebbene la loro utilità venga ampliamente riconosciuta – soprattutto per la didattica della linguistica e delle lingue –, il loro sfruttamento – soprattutto nelle discipline più interessate all'indagine del significante – è drasticamente calato, tanto con riferimento alla frequenza, quanto con riferimento ai domini d'uso.

Infatti, le trascrizioni impressionistiche vengono sempre più spesso considerate inadeguate per la descrizione e l'analisi scientifica del significante, principalmente perché l'alta variabilità dei risultati prodotti tanto da uno stesso trascrittore, quanto da diversi trascrittori impegnati a rendere un identico stimolo distale (Bucholtz 2007), ne fa prodotti difficilmente compatibile con il requisito della replicabilità metrologica richiesto dall'applicazione del metodo scientifico alle scienze umane e sociali.

[28] Per esempio, il più diffuso strumento per la trascrizione strumentale acustica, Praat (Boersma 2001), è gratuito e gli unici costi da preventivare per il suo uso sono quelli legati alla formazione del trascrittore.

Per tale motivo, le trascrizioni impressionistiche alfabetiche sono progressivamente passate dall'essere un'unica tecnica di trasposizione del significante, all'essere una delle possibili soluzioni per glossare delle trascrizioni strumentali così da renderle leggibili anche da quanti non conoscano lo strumento e la sua modalità di rappresentazioni dei dati[29].

Per quanto sempre più diffusa tale pratica è rischiosa per l'elaborazione di valide osservazioni e teorie. Infatti, come anticipato in §4, è possibile che gli autori e i lettori di una glossa redatta ricorrendo per esempio ai simboli dell'alfabeto fonetico internazionale implicitamente ritengano che le entità riferite nella rappresentazione strumentale e quelle riferite nella notazione alfabetica coincidano, invece che complementarsi a vicenda, perché mentre le trascrizioni strumentali rendono conto del valore assunto da alcune variabili fisiche rilevate dagli strumenti fonetici prima, durante o dopo la produzione del parlato, le trascrizioni impressionistiche riportano graficamente l'esito dell'analisi di alcuni oggetti percettivi operata dal trascrittore sintetizzando la sua esperienza di ascolto dei segnali acustici (le cui proprietà fisiche possono essere misurate dagli strumenti) e la sua adesione a una pratica – più o meno teoricamente fondata – di resa di quell'esperienza.

In termini metodologici, ciò impone una riflessione sulla differenza tra gli oggetti riferiti sia con riferimento al piano della relazione tra articolazione e acustica da un lato e tra acustica e percezione dall'altro; sia con riferimento al piano della capacità degli strumenti e dei trascrittori di riportare fedelmente quanto rilevato a livello di stimolo distale o percetto.

Pur senza poter qui approfondire le diverse tematiche, va notato che per quanto concerne la relazione tra il piano articolatorio e quello acustico il problema principale è rappresentato dall'impossibilità di postulare una corrispondenza biunivoca tra comportamenti articolatori ed esiti acustici, tant'è che a uno stesso gesto articolatorio possono corrispondere effetti acustici differenti (cfr. Stevens & Hanson 2010 per una introduzione alla tematica e Ximenes *et al*. 2017 per una verifica delle conseguenze sul piano della descrizione di lingue).

[29] Oppure – ma solo nel caso non siano disponibili apparecchiature adeguate – in tecnica di fortuna per l'annotazione temporanea di dati da sottoporre a successiva verifica strumentale.

Per quanto riguarda invece la relazione tra il piano acustico e quello percettivo, è stato da tempo e da più parti dimostrato – anche in prospettiva trascrittoria – che, poiché i processi di interpretazione del significante linguistico sono finalizzati all'estrazione di significati, gli ascoltatori possono percepire elementi che non siano effettivamente presenti nel segnale acustico, se utile per dare un senso a quanto udito; o, al contrario e per la stessa ragione, non percepirne altri che siano ad esempio rintracciabili strumentalmente (Oller & Eilers 1975; cfr. anche Engstrand *et al*. 2007).

Per quanto riguarda infine la capacità – o, meglio, da presumersi incapacità – di strumenti e trascrittori di riportare esattamente quanto rilevato, occorre rimandare anzitutto al problema della irreversibile degradazione del segnale nelle trasmissioni analogiche o digitali causate da rumore o distorsioni che comporta una certa infedeltà tra lo stimolo distale e la sua rappresentazione strumentale finale. Inoltre, merita sottolineare che probabilmente il trascrittore umano opera il riconoscimento dei suoni alla luce della complessa interazione tra la sua memoria semantica, che gli permette di segmentare il flusso del significante trovando unità di senso (Mitterer & Cutler 2006); la sua memoria ecoica, che gli permette di mantenere attivo lo stimolo uditivo anche dopo che sia terminato (Johnson 2007); la sua memoria dichiarativa, che gli permette di confrontare lo stimolo uditivo mantenuto attivo con i prototipi interiorizzati e le forme grafiche per esplicitarli, così che in ciascuno degli stadi di memoria si possono dare delle modificazioni della percezione, interpretazione o rappresentazione tali per cui lo stimolo distale, il percetto e la sua resa grafica in trascrizione impressionistica non coincidano (Knight 2011).

Se da un lato ciò conferma che, pur muovendo da uno stesso fenomeno, gli oggetti finali della trascrizione strumentale e di quella impressionistica pertengono livelli distinti – rispettivamente quello fisico e fonetico ancorati nella fisiologia del parlante, e quello sensoriale e fonologico ancorati nella percezione dell'ascoltatore – dall'altro ciò induce a riconoscere che il giudizio di chi trascriva impressionisticamente va sempre considerato come legittimo[30], anche quando sia disallineato da

[30] Ovviamente a patto che la traduzione dell'esperienza percettiva venga fatta in conformità alle regole del modello pratico con cui la si voglia comunicare, per esempio quelle elaborate in IPA (1999).

ciò che viene rilevato strumentalmente. Per questo motivo concordo con i molti sostenitori della trascrizione impressionistica – per esempio Howard & Heselwood (2011) – che non solo stigmatizzano la pratica di convalida delle trascrizioni impressionistiche tramite verifica strumentale, ma anche promuovono la documentazione e la verifica di tutte le incongruenze tra i due tipi di trascrizione, così da giungere a una migliore comprensione del processo di trasformazione del parlato articolato in parlato trasmesso, del parlato trasmesso in parlato udito e, infine del parlato udito in parlato percepito. Per lo stesso motivo ritengo poi che l'uso della trascrizione automatica[31] – che equiparo a una trascrizione strumentale perché si basa sull'analisi automatica di un segnale, solitamente quello acustico, cui viene data forma ortografica – sia da utilizzare solo se si sia maturata la consapevolezza che a partire dall'ascolto dello stesso stimolo prossimale un trascrittore umano – in fin dei conti l'unico davvero significativo ai fini di una indagine linguistica – e un trascrittore automatico potrebbero giungere a rappresentazioni diverse del significante.

Per poter procedere ad ulteriormente esplicitare le ragioni di questa posizione, è utile discutere di come le trascrizioni strumentali possano essere impiegate.

6. *Usi della trascrizione strumentale*

Secondo Heselwood (2013: capitolo 6), indipendentemente da come siano generate, le trascrizioni strumentali possono essere impiegate in tre maniere differenti, che prevedono un progressivo distanziamento dal dato originale. Il primo modo prevede di elaborare l'analisi a partire direttamente dalla rappresentazione del segnale strumentale (§6.1). Il secondo vuole invece che la rappresentazione del segnale venga glossata ricorrendo a un sistema di notazione discreto, per esempio lo IPA (§6.2). Il terzo modo prevede infine che la rappresentazione venga impiegata per dirimere una trascrizione impressionistica (§6.3).

[31] Quello della trascrizione automatica è un fenomeno sempre più diffuso stante da un lato la possibilità di trascrivere un numero crescente di parole nell'unità di tempo, dunque di creare basi di dati più ampie in minor tempo, dall'altro la necessità di ridurre i costi delle ricerche, tipicamente alti nel caso di trascrizioni impressionistiche.

6.1 Trascrizioni strumentali in senso stretto

Il primo modo è il più radicale dei tre, perché prevede di accettare che la resa grafica di quanto colto e visualizzato dagli strumenti costituisca a pieno titolo una forma di trascrizione. In termini analitici la peculiarità di questa trascrizione – strumentale in senso stretto – è quella di fornire una raffigurazione del parlato coerente in ogni suo punto, perché elaborata ricorrendo a sensori e non a sensi, ovvero senza che si attivino processi interpretativi capaci di modificare la resa degli stimoli distali, se non come detto per via di difetti nella esecuzione della rilevazione e/o della trasduzione delle informazioni colte dal sensore. La coerenza della misurazione, tuttavia, non implica la sua oggettività, perché tanto le scelte costruttive dello strumento e della visualizzazione delle informazioni, quanto il suo impiego risultano sempre da scelte dell'operatore, per esempio in termini di posizionamento delle sonde, oppure di impostazioni dei parametri di analisi e resa visiva del segnale.

Inoltre, l'adesione a questo tipo di trascrizione obbliga a una radicale trasformazione del processo di analisi dei dati, perché impone di fare della ricerca linguistica muovendo da dati non scritti, in qualche modo così rivedendo uno degli approcci della linguistica moderna che, per quanto abbia promosso la priorità del parlato, lo ha indagato per lo più muovendo dalla sua forma (tra)scritta – meglio alfabetizzata – come emerge da Hjelmslev che nota che "è importante per la teoria linguistica che si riesca ad affinare l'idea che soggiace all'invenzione della scrittura, cioè l'idea di fornire un'analisi che porta ad entità di estensione minima e di numero infimo" (1968: 47).

Se si escludono le proposte di Hermann e Marichelle citate in §3.1, la trasformazione del processo di analisi che consegue dall'intendere la visualizzazione strumentale come una forma di trascrizione, è ancora agli esordi, perciò sfrutta in larga parte conoscenze elaborate al di fuori della linguistica, dunque dalla spendibilità da validare. Per esempio, nell'ambito della trascrizione strumentale di dati articolatori, l'ultimo lustro ha visto un significativo incremento delle informazioni generate ricorrendo alla tomografia a risonanza magnetica. Queste differiscono dalla gran parte delle altre trascrizioni strumentali e non strumentali[32],

[32] Con la sola eccezione forse delle trascrizioni di marca iconica – che tuttavia si basavano su propriocezioni e/o speculazioni - sviluppate a partire dal XVII secolo e promosse per esempio negli alfabeti organici di John Wilkins (1668) e Alexander Melville Bell (1867), per cui cfr. Kemp (2006).

perché consentono una rappresentazione realistica o, meglio, iconica[33] degli articolatori osservati. Se dal punto di vista della valutazione in tempo reale dei fenomeni – per esempio a fini diagnostici e terapeutici per quelle che sono le applicazioni patolinguistiche della trascrizione strumentale (Ball & Code 1997; Heselwood & Howard 2008) – tale modalità di rappresentazione costituisce un vantaggio, la loro analisi anche in tempo differito può invece risultare problematica. Anzitutto perché richiede un cambiamento del paradigma di osservazione, poiché in termini generali la modalità di visualizzazione delle informazioni scientifiche ne influenza la percezione in conseguenza di quei processi di "elaborazione preventiva" – ovvero della capacità del sistema visivo umano di identificarne le informazioni di base (Healey & Enns 2011) – che hanno una ricaduta anche per l'analisi delle trascrizioni, come dimostra la diversa percezione delle informazioni a seconda della loro organizzazione spaziale (Du Bois 1991). Poi perché richiede di trovare una soluzione tanto per l'estrazione delle sole informazioni foneticamente pertinenti[34], quanto per una loro analisi che passi attraverso l'identificazione di pratiche – preferibilmente automatizzabili così da garantirne la comparabilità – accettabili per la segmentazione degli oggetti visivi e la loro classificazione (Deserno 2011). Ciò può per esempio avvenire attraverso l'identificazione di biomarche visive (Kessler *et al*. 2014) oppure di indici anatomici statici o dinamici (Töger *et al*. 2017) che permettano di discriminare i singoli articolatori (Carignan *et al*. 2020) alla luce di pratiche ormai diffuse in medicina, ma la cui rappresentatività – tanto in termini metrologici, quanto linguistici – è dibattuta anche perché rimanda a paradigmi diversi da quelli cui per decenni si è fatto ricorso per l'indagine di trascrizioni impressionistiche alfabetiche, ad esempio quelli delle discipline di laboratorio.

Nonostante le difficoltà attuali, l'investimento di risorse necessario per concludere questo processo verrà forse compensato dalla possibilità di eliminare ogni intervento manipolativo su base linguistica dei dati fi-

[33] In effetti non vi è identità di forma o struttura tra realtà e trascritto, bensì similarità. Per esempio, la policromia originale è sostituita da immagini a livelli di grigi, e i contrasti di questi ultimi sono modificabili a piacimento dall'utente dello strumento per favorire la discriminazione delle strutture anatomiche, cosa impossibile nella realtà.

[34] Per esempio, le immagini del cavo orale prodotte ricorrendo a tomografia a risonanza magnetica spesso includono anche la visualizzazione di cranio ed encefalo, non pertinenti per la trascrizione e l'indagine articolatoria, dunque da ignorare.

sici, quindi di quelle ambiguità necessariamente indotte dai processi di *textualization*[35] (Bauman & Briggs 1990) e *retextualization*[36] (Haberland & Mortensen 2016) che il ricorso a notazioni alfabetiche comporta, e di cui tratterò nel prossimo paragrafo.

6.2 Trascrizioni strumentali glossate

Come rilevato da Heselwood (2013: §6.2.10), sebbene le trascrizioni strumentali in senso stretto siano di per loro validi strumenti di lavoro, spesso vengono accompagnate da notazioni alfabetiche, tipicamente secondo le norme dell'alfabeto fonetico internazionale, che ne favoriscano la leggibilità da parte di utenti non esperti della tecnica strumentale. Tale pratica si è notevolmente diffusa da quando gli ambienti informatici per l'analisi strumentale del significante hanno reso disponibili ambienti di trascrizione multilineare che consentono di accompagnare la riga dedicata alla visualizzazione delle informazioni colte dai sensori con una linea per le annotazioni.

Poiché le notazioni alfabetiche impiegano glifi discreti, solitamente questi ambienti informatizzati permettono di segmentare la linea di annotazione. Questa operazione di discretizzazione può essere operata automaticamente oppure impressionisticamente a partire dalla linea strumentale principale. Nel primo caso l'operazione solitamente si basa sull'identificazione di regolarità e irregolarità nel segnale o nella sua rappresentazione visiva, così da garantire la sistematicità della segmentazione, come fatto ad esempio da alcuni programmi informatici per la trascrizione, sottotitolatura o allineamento forzato di registrazione audio e testi scritti (Kisler *et al*. 2017). Nel secondo caso, invece, l'operazione è affidata al trascrittore che può decidere di basarsi vuoi sull'interpretazione visiva del segnale; vuoi sull'interpretazione uditiva della registra-

[35] Con riferimento alla trascrizione, la *textualization* è definita da Bauman & Briggs (1990: 73) come "the process of rendering discourse extractable, of making a stretch of linguistic production into a unit – a text - that can be lifted out of its interactional setting" e, più chiaramente, da Park & Bucholtz (2009: 485) come: "the process by which circulable texts are produced by extracting discourse from its original context and reifying it as a bounded object".

[36] Con riferimento alla trascrizione, il processo di *retextualization* è definito da Haberland & Mortensen (2016: 585) come "the process by which the text (in our case the transcript) is brought to life again by being read, either aloud or silently".

zione del parlato sincronizzata al segnale visualizzato; vuoi – meno auspicabilmente – su una combinazione delle due strategie. Sebbene ricorrendo tanto alla discretizzazione visiva, quanto a quella uditiva si giunga alla segmentazione del segnale, il valore e la spendibilità dei due tipi di risultato sono profondamente differenti. Infatti, nonostante entrambe le operazioni siano basate sui sensi, solo la seconda viene operata rifacendosi a processi di discretizzazione di marca fonetica definiti dall'insieme dei modelli fonologici interiorizzati o ritenuti possibili dal trascrittore e che mirano all'identificazione di contrasti spesi o spendibili per veicolare significati alla luce di valori definiti. Al contrario, durante la segmentazione visiva del segnale, l'operazione viene effettuata identificando somiglianze e differenze nella forma grafica del trascritto strumentale, dunque trattando l'informazione disponibile come se ciascuna sua parte fosse funzionalmente identica alle altre, ovvero in maniera linguisticamente neutra, analogamente a quanto fatto durante la segmentazione automatica delle immagini tomografiche per l'identificazione degli articolatori.

Quale che sia la modalità di identificazione dei confini, l'operazione di segmentazione determina un disaccoppiamento tra la natura continua del segnale fisico e quella discreta della sua percezione uditiva o visiva. Inoltre, a segmentazione conclusa, a ciascun frammento identificato viene combinata un'annotazione alfabetica che funga da glossa della visualizzazione strumentale dell'osservabile. Il limite principale dell'operazione è che, come già accennato, glifi che dovrebbero essere impiegati solo per rimandare alla dimensione fonemica, dunque percettiva, vengono riciclati per trattare di fenomeni fonetici ai più diversi livelli. Pertanto, in fase di ricostruzione del valore associato a ciascun simbolo dall'estensore delle glosse (*retextualization*), l'analista può essere indotto a fraintendere o neutralizzare i valori fonetici riferiti da ciascun glifo in conseguenza del loro presentarsi in forme identiche pur essendo espressione di analisi di marca differente[37]. Due sono i possibili effetti. Da un lato, l'interpretazione delle glosse non più alla luce della prospettiva con cui siano state originariamente elaborate, bensì di quella che il lettore attribuisca loro, perché più familiare o utile. Dall'altro, la compara-

[37] Nei termini di Edwards (2005: 325) si tratterebbe di una violazione del criterio di *visual separability of unlike events* per cui "types of information which are qualitatively different from each other […] tend to be encoded in distinctly different ways".

zione delle trascrizioni di fenomeni erroneamente ritenuti equivalenti in virtù dell'essere state riportate a un sistema di notazione che impieghi i medesimi glifi. Purtroppo esemplari in tal senso sono i confronti tra gli esiti di analisi acustiche e articolatorie su fenomeni fonemici coincidenti, non legittime essendo al più i due tipi di analisi complementari.

Nel complesso, la pratica di glossatura alfabetica di trascrizioni strumentali può comportare problemi di sovra-interpretazione o di sotto-rappresentatività della trascrizione (Heselwood 2013: 226). Il primo caso si verifica quando una glossa viene interpretata alla luce del valore che ha nella notazione fonetica includendo proprietà assenti nella trascrizione strumentale, ad esempio perché non rilevabili dall'attrezzatura impiegata. Un esempio di tale errore si ritrova nelle discussioni di rilevazioni elettropalatografiche annotate con i simboli IPA in cui – di fatto speculando – si facciano affermazioni su quali porzioni linguali sarebbero state coinvolte nel contatto, aspetto non rilevabile dallo strumento che può solo registrare le porzioni del palato interessate da un contatto. Il secondo caso – quello della sotto-rappresentatività della trascrizione – si verifica invece quando la trasposizione strumentale contenga informazioni per cui non si dispone di adeguata possibilità di notazione nell'alfabeto di riferimento, per esempio sempre con riferimento a osservazioni elettropalatografiche, la asimmetrica distribuzione sul piano coronale di un contatto tra la lingua e il palato.

6.3 Trascrizioni strumentali ancillari

La terza modalità di impiego delle trascrizioni strumentali ipotizzata da Heselwood (2013) è quello che le vede sfruttate per confermare la validità di una trascrizione impressionistica. Tale uso è molto frequente soprattutto al di fuori della comunità dei fonetisti, che spesso vengono interpellati dai linguisti empirici perché mettano loro a disposizione strumentazione e conoscenze per dirimere la trascrizione di segmenti difficili da classificare alla luce del solo ascolto. Tale richiesta – che solitamente si risolve nell'applicazione di tecniche di analisi acustica a registrazioni audio di qualità più o meno adeguata – comporta che nella trascrizione la componente strumentale assuma un ruolo secondario rispetto a quello dell'ascolto, ovvero che si elabori una trascrizione per lo più impressionistica, ma integrata in alcuni suoi passaggi dai risultati di una qualche analisi strumentale.

Come anticipato, a mio giudizio questo tipo di trascrizione è da evitare, perché più che la complementarietà, promuove la confusione di due tecniche che – per i motivi detti nelle sezioni precedenti e per quelli che presenterò nelle conclusioni – sarebbe bene mantenere distinte, così da ricorrere o a una trascrizione strumentale in senso stretto o a una trascrizione impressionistica.

7. *Conclusione*

In questo contributo ho trattato di trascrizione strumentale, spesso contrastandola con la trascrizione impressionistica che utilizzi il sistema di notazione dell'alfabeto fonetico internazionale. Nel farlo, non ho inteso argomentare a favore del primato dell'una sull'altra, quanto piuttosto chiarire le peculiarità – e dunque la spendibilità – di ciascuna nel quadro teorico cui ogni trascrittore voglia aderire. In particolare, non ho voluto sostenere che la crescente disponibilità di strumenti di osservazione delle proprietà fisiche del significante linguistico debba portare all'esclusione della trascrizione impressionistica dal novero dei metodi scientifici.

Infatti, da un lato gli oggetti di interesse della trascrizione strumentale e di quella impressionistica sono tra loro differenti e riportabili graficamente solo ricorrendo a metodologie distinte, dedicate e appropriate. Dall'altro, mentre le trascrizioni impressionistiche hanno dimostrato di essere valide tecniche di descrizione dei fenomeni linguistici, quelle strumentali necessitano ancora di alcune messe a punto. Per esempio con riferimento al problema della normalizzazione – ovvero del fatto che produzioni fonemicamente identiche possono mostrare grandi differenze e variazioni fisiche (Johnson 2005; cfr. anche Vietti in questo volume) – di fatto irrilevante per la trascrizione impressionistica, ma ancora di là dall'essere risolto per molte tecniche di trascrizione strumentale, come dimostrano due grandi linee di sviluppo divergenti, quella di normalizzazione *ex-ante* che promuove la collocazione dei sensori in punti anatomicamente (Rebernik *et al*. 2021) o funzionalmente equivalenti, o quella di normalizzazione *ex-post*, che insegue tecniche di confronto statistico tra gli elementi (Wang *et al.* 2014).

Anche per questi motivi, nel contributo ho inteso incoraggiare il ricorso alla trascrizione strumentale ancillare soprattutto quale strategia per colmare le lacune conoscitive relative all'interfaccia tra il piano fisi-

co della produzione e trasmissione del significante e quello psichico e simbolico della sua percezione simbolica. Tale avanzamento viene – e presumo verrà ulteriormente – facilitato grazie allo sfruttamento del potenziale reso disponibile dagli ambienti informatici per trascrizione elaborati in seno all'informatica umanistica che consentono di creare trascrizioni multilivello, dunque di evidenziare convergenze e divergenze tra i piani di trascrizione.

Infine, ho ritenuto utile trattare di queste tematiche perché la crescente disponibilità di tecniche per il trattamento strumentale automatico del parlato fa delle trascrizioni oggetti tecnologici sempre più facilmente producibili ma, forse, meno adatti all'indagine linguistica.

Riferimenti bibliografici

Albano Leoni, Federico & Maturi, Pietro. 2002. *Manuale di fonetica*. Roma: Carocci.

Amorosa, Hedwig & von Benda, Ursula & Wagner, Edith & Keck, A. 1985. Transcribing detail in the speech of unintelligible children: A comparison of procedures. *British Journal of Disorders of Communication* 20. 281–287.

Badin, Pierre & Serrurier, Antoine. 2006. Three-dimensional modeling of speech organs: Articulatory data and models. *Technical Committee of Psychological and Physiological Acoustics* 36(5). 421–426.

Ball, Martin & Code, Chris (eds.). 1997. *Instrumental Clinical Phonetics*. London: Whurr Publishers.

Barry, William & Trouvain, Jürgen. 2011. *Phonus 16, In memoriam Wolfgang von Kempelen*. Saarbrücken: Institute of Phonetics, Saarland University.

Bauman, Richard & Briggs, Charles. 1990. Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19. 59–88.

Beckman, Mary & Kingston, John. 2011. Introduction. Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech (Reprint). In Cohn, Abigail & Fougeron, Cécile & Huffman, Marie (eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199575039.001.0001/oxfordhb-9780199575039 (consultato il 29.10.2020).

Boersma, Paul. 2001. Praat: A system for doing phonetics by computer. *Glot International* 5(9/10). 341–345.

Brain, Robert. 1998. Standard and semiotics. In Lenoir, Timothy (ed.), *Inscribing science: Scientific texts and the materiality of communication*, 249–284. Stanford: Stanford University Press.

Brain, Robert. 2015. *The Pulse of Modernism: Physiological Aesthetics in Fin-de-Siècle Europe*. Washington: University of Washington Press.

Brenni, Paolo. 2013. Gli strumenti della scienza e la loro produzione. In *Il Contributo italiano alla storia del Pensiero – Tecnica*. Roma: Istituto della Enciclopedia italiana fondata da Giovanni Treccani. Versione telematica: http://www.treccani.it/enciclopedia/gli-strumenti-della-scienza-e-la-loro-produzione_%28Il-Contributo-italiano-alla-storia-del-Pensiero:-Tecnica%29/ (consultato il 29.10.2020).

Bresch, Eeik & Yoon-Chul, Kim & Krishna, Nayak & Dani, Byrd & Shrikanth, Narayanan. 2008. Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging. *IEEE Signal Processing Magazine* 25(3). 123–132.

Bucholtz, Mary. 2007. Variation in transcription. *Discourse Studies* 9(6). 784–808.

Carignan, Christopher & Hoole, Phil & Kunay, Esther & Pouplier, Marianne & Joseph, Arun & Voit, Dirk & Frahm, Jens & Harrington, Jonathan. 2020. Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI. *Laboratory Phonology* 11(1). http://doi.org/10.5334/labphon.214.

Cho, Taehong & Whalen, Douglas & Docherty, Gerard. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72. 52–65.

Clarke, Bruce & Henderson, Linda. 2002. Introduction. In Clarke, Bruce & Henderson, Linda (eds.), *From energy to information: Representation in science and technology, art, and literature*, 1–18. Stanford: Stanford University Press.

Dascal, Marcelo. 1999. Misunderstanding. *Journal of Pragmatics* 31(6). 753–863.

De Saussure, Ferdinand. 1916. *Corso di linguistica generale*. *Versione italiana, introduzione, traduzione e commento di Tullio De Mauro*. Bari: Laterza.

Deserno, Thomas. 2011. Fundamentals of Biomedical Image Processing. In Deserno, Thomas (ed.), *Biomedical Image Processing*, 1–51. Berlin: Springer.

Deutsch, Diana. 2019. *Musical Illusions and Phantom Words: How Music and Speech Unlock Mysteries of the Brain*. New York/Oxford: Oxford University Press.

Du Bois, John. 1991. Transcription Design Principles for Spoken Discourse Research. *Pragmatics* 1(1). 71–106.

Edwards, Jane. 1992. Transcription of discourse. In Bright, William (ed.), *International Encyclopedia of Linguistics*, 370–371. New York/Oxford: Oxford University Press.

Edwards, Jane. 1993. Principles and contrasting systems of discourse transcription. In Edwards, Jane & Lampert, Martin (eds.), *Talking Data, Transcription and Coding in Discourse Research*, 3–31. Hillsdale: Lawrence Erlbaum.

Edwards, Jane. 2005. The Transcription of Discourse. In Schiffrin, Deborah, & Tannen, Deborah & Hamilton, Heidi E. (eds.), *The Handbook of Discourse Analysis*, 321–348. Malden: Wiley.

Engstrand, Olle & Frid, Johan & Lindblom, Björn. 2007. A perceptual bridge between coronal and dorsal /r/. In Solé, Maria & Beddor, Patrice & Ohala, Manjari (eds.), *Experimental approaches to phonology*, 175–191. Oxford: Oxford University Press.

Farmer, Alvirda. 1997. Spectrography. In Ball, Martin & Code, Chris (eds.), *Instrumental Clinical Phonetics*, 22–63. London: Whurr Publishers.

Feaster, Patrick. 2010. *The Phonautographic Manuscripts of Édouard-Léon Scott De Martinville*. Bloomington, Indiana: First Sounds. http://www.firstsounds.org/publications/articles/Phonautographic–Manuscripts.pdf. (consultato il 29.10.2020).

Frings, Stephan & Müller, Frank. 2014. *Biologie der Sinne*. Berlin: Springer.

Fulop, Sean & Fitz, Kelly. 2006. A Spectrogram for the Twenty-First Century. *Acoustics today* 2(3). 26–33.

Gick, Bryan & Schellenbery, Murray & Stavness, Ian & Taylor, Rayan. 2019. Articulatory phonetics. In Katz, William & Assmann, Peter (eds.), *The Routledge Handbook of Phonetics*, 107–125. Abingdon-on-Thames: Routledge.

Haas, Walter. 1990. Jacob Grimm und die deutschen Mundarten. *Zeitschrift für Dialektologie und Linguistik* 65.

Haberland, Hartmut & Mortensen, Janos. 2016. Transcription as Second-Order Entextualization: The Challenge of Heteroglossia. In Capone, Alessandro & Mey, Jacob (eds.), *Interdisciplinary Studies in Pragmatics, Culture and Society, Perspectives in Pragmatics, Philosophy & Psychology 4*, 581–600. Basel: Springer.

Healey, Christopher & Enns, James. 2011. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE transactions on visualization and computer graphics* 18. 1170–1188.

Henry, John. 1997. *The scientific revolution and the origins of modern science*. New York: St. Martin's Press.

Herbst, Christian & Fitch, Tecumseh & Švec, Jan. 2010. Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively. *The Journal of the Acoustical Society of America* 128(5). 3070–3078.

Hermann, Ludimar. 1894. Phonophotographische Untersuchungen. *Archiv für die gesamte Physiologie des Menschen und der Tiere* 58. 264–279.

Hertrich, Ingo & Ackermann, Hermann. 2013. Neurophonetics. *WIREs Cognitive Science* 4. 191–200.

Heselwood, Barry. 2013. *Phonetic Transcription in Theory and Practice*. Edinburgh: Edinburgh University Press.

Heselwood, Barry & Howard, Sara. 2008. Clinical Phonetic Transcription. In Ball, Martin & Perkins, Michael & Müller, Nicole & Howard, Sara (eds.), *The Handbook of Clinical Linguistics*, 381–399. Malden: Blackwell.

Hjelmslev, Louis. 1968. *I fondamenti della teoria del linguaggio. Introduzione e traduzione di Giulio Lepschy*. Torino: Einaudi.

Holmes, Frederic. 2003. L'Ottocento: biologia. Fisiologia e medicina sperimentale. Roma: Istituto della Enciclopedia italiana fondata da Giovanni Treccani. Versione telematica:: https://www.treccani.it/enciclopedia/l-ottocento-biologia-fisiologia-e-medicina-sperimentale_%28Storia-della-Scienza%29/ (consultato il 29.03.2020).

Howard, Sara & Heselwood, Barry. 2011. Instrumental and perceptual phonetic analyses: The case for two-tier transcriptions. *Clinical Linguistics and Phonetics* 25. 940–948.

IPA. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Israel, Paul. 1998. *Edison: a Life of Invention*. New York: Wiley.

James, Frank. 1989. *The Development of the Laboratory: Essays on the Place of Experiments in Industrial Civilization*. London: Palgrave Macmillan.

Johnson, Keith. 2005. Speaker Normalization in Speech Perception. In Pisoni, David & Remez, Robert (eds.), *The Handbook of Speech Perception*, 363–389. Malden: Wiley.

Johnson, Keith. 2007. Decisions and mechanisms in exemplar-based phonology. In Sole, Maria & Beddor, Patrice & Ohala, Manjari (eds.), *Experimental Approaches to Phonology*, 25–40. Oxford: Oxford University Press.

Juni, Samuel & Gross, Junie. 2008. Emotional and Persuasive Perception of Fonts. *Perceptual and Motor Skills* 106(1). 35–42.

Kemp, Alan. 2006. Phonetic Transcription: History. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*, vol. 9, 395–410. Boston: Elsevier.

Kessler, Larry & Barnhart, Huiman & Buckler, Andrew & Choudhury, Kingshuk & Kondratovich, Marina & Toledano, Alicia & Guimaraes, Alexander & Filice, Ross & Zhang, Zheng & Sullivan, Daniel. 2014. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Statistical Methods in Medical Research* 24(1). 9–26.

Kisler, Thomas & Reichel, Uwe & Schiel, Florian. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326–347.

Knight, Rachael-Anne. 2011. Towards a cognitive model of phonetic transcription. In Przedlacka, Joanna & Maidment, John & Ashby, Michael (eds.), *Proceedings of the Phonetics Teaching and Learning Conference 2011*, 17–20. London: University College London.

Kochetov, Alexei. 2020a. Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass* 14(4). https://onlinelibrary.wiley.com/doi/abs/10.1111/lnc3.12368 (consultato il 29.10.2020).

Kochetov, Alexei. 2020b. Research methods in articulatory phonetics II: Studying other gestures and recent trends. *Language and Linguistics Compass* 14(6). https://onlinelibrary.wiley.com/doi/abs/10.1111/lnc3.12371 (consultato il 29.10.2020).

Krause, Peter & Kay, Christopher & Kawamoto, Alan. 2020. Automatic Motion Tracking of Lips using Digital Video and OpenFace 2.0. *Laboratory Phonology* 11(1). https://www.journal-labphon.org/articles/10.5334/labphon.232/# (consultato il 29.10.2020).

Kühnert, Barbara & Nolan, Francis. 1999. The origin of coarticulation. In Hardcastle, William & Hewlett, Nigel (eds.), *Coarticulation. Theory, Data and Techniques*, 7–30. Cambridge: Cambridge University Press.

Kursell, Julia. 2013. Experiments on Tone Color in Music and Acoustics: Helmholtz, Schoenberg, and Klangfarbenmelodie. *Osiris* 28(1). 191–211.

Ladefoged, Peter & Maddieson, Ian. 1996. *The Sounds of the World's Languages*. Malden: Blackwell.

Lyons, Jack. 2016. Epistemological Problems of Perception. In Zalta, Edward (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/perception-episprob/ (consultato il 29.10.2020).

Marage, Dr. 1898. Les phonographes et l'étude des voyelles. *L'année psychologique* 5. 226–244.

Marichelle, Hector, 1897. *La parole d'après le tracè du phonographe*. Paris: Delagrave.

Mitterer, Holger & Cutler, Anne. 2006. Speech perception. In Brown, Keith & Anderson, Anne (eds.), *The Encyclopaedia of Language & Linguistics*, 770–782. Amsterdam: Elsevier.

Mompean, Jose. 2014. Cognitive linguistics and phonology. In Littlemore, Jeannette & Taylor, John (eds.), *The Bloomsbury Companion to Cognitive Linguistics*, 253–276. London: Bloomsbury Publishing.

Oller, Kimbrough & Eilers, Rebecca. 1975. Phonetic expectation and transcription validity. *Phonetica* 31. 288–304.

Park, Joseph & Bucholtz, Mary. 2009. Introduction. Public transcripts: Entextualization and linguistic representation in institutional contexts. *Text and Talk* 29. 485–502.

Pattamadilok, Chotiga & Knierim, Iris & Kawabata Duncan, Keith & Devlin, Joseph. 2010. How Does Learning to Read Affect Speech Perception?. *Journal of Neuroscience* 30(25). 8435–8444.

Perkell, Joseph & Cohen, Marc & Svirsky, Mario & Matthies, Melanie & Garabieta, Iñaki & Jackson, Michel. 1992. Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92. 3078–3096.

Perrachione, Tyler & Ghosh, Satrajit & Ostrovskaya, Irina & Gabrieli, John & Kovelman, Ioulia. 2017. Phonological Working Memory for Words and Nonwords in Cerebral Cortex. *Journal of Speech, Language, and Hearing Research* 60(7). 1959–1979.

Pierrehumbert, Janet & Beckman, Mary & Ladd, Robert. 2000. Conceptual Foundations of Phonology as a Laboratory Science. In Burton-Roberts, Noel & Carr, Philip & Docherty, Gerard (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*, 273–303. Oxford: Oxford University Press.

Pouplier, Marianne & Goldstein, Louis. 2005. Asymmetries in the perception of speech production errors. *Journal of Phonetics* 33(1). 47–75.

Raphael, Lawrence. 2005. Acoustic Cues to the Perception of Segmental Phonemes. In Pisoni, David & Remez, Robert (eds.), *The Handbook of Speech Perception*, 182–206. Malden: Wiley.

Rebernik, Teja & Jacobi, Jidde & Jonkers, Roel & Noiray, Aude & Wieling, Martijn. 2021. A review of data collection practices using electromagnetic articulography. *Laboratory Phonology,* 12(1), 6. http://doi.org/10.5334/labphon.237 (consultato il 09.04.2021).

Roach, Peter & Hardcastle, William. 1976. A computer system for the processing of electropalatographic and other data. In Tatham, Mark (ed.), *Proceedings of the Vth Phonetics Symposium*, 127–142. Colchester: University of Essex.

Rothenberg, Martin. 1992. A Multichannel Electroglottograph. *Journal of Voice* 6(1). 36–43.

Rousselot, Jean-Pierre. 1891a. Les modifications phonétiques du langage. *Revue des patois gallo-romans* 4. 65–208.

Rousselot, Jean-Pierre. 1891b. La méthode graphique appliquée à la recherche des transformations inconscientes du langage. *Revue des patois gallo-romans* 4. 209–213.

Shankweiler, Donald & Fowler, Carol. 2015. Seeking a reading machine for the blind and discovering the speech code. *History of Psychology* 18(1). 78–99.

Shriberg, Lawrence & Lof, Gregory. 1991. Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics* 5(3). 225–279.

Simone, Raffaele. 2012. *Presi nella rete*. Milano: Garzanti.

Smith, Mark. 2014. *From Sight to Light: The Passage from Ancient to Modern Optics*. Chicago: University of Chicago Press.

Sock, Rudolph & Hirsch, Fabrice & Laprie, Yves & Perrier, Pascal & Vaxelaire, Béatrice. 2011. An X-ray database, tools and procedures for the study of speech production. In Ostry, David & Baum, Shari & Ménard, Lucie & Gracco, Vincent (eds.), *ISSP2011. Proceedings of the 9th International Seminar on Speech Production*, 41–48. Montréal: ISSP.

Spreafico, Lorenzo. 2020. Corpora di parlato o corpora di ascoltato?. *Rivista italiana di dialettologia* 44. 38–51.

Spreafico, Lorenzo & Vietti, Alessandro (eds.). 2020. Techniques and Methods for Investigating Speech Articulation. *Special collection of Laboratory Phonology*. https://www.journal-labphon.org/collections/special/techniques-and-methods-for-investigating-speech-articulation/ (consultato il 09.04.2021).

Stella, Massimo & Bernardini, Paolo & Sigona, Francesco & Stella, Antonio & Grimaldi, Mirko & Gili Fivela, Barbara. 2012. Numerical instabilities and three-dimensional electromagnetic articulography. *The Journal of the Acoustical Society of America* 132(6). 3941–3949.

Stella, Massimo & Stella, Antonio & Sigona, Francesco & Bernardini, Paolo & Grimaldi, Mirko & Gili Fivela, Barbara. 2013. Electromagnetic Articulography with AG500 and AG501. *INTERSPEECH 2013*. https://www.isca-speech.org/archive/interspeech_2013/i13_1316.html (consultato il 29.10.2020).

Stepp, Cara. 2012. Surface Electromyography for Speech and Swallowing Systems: Measurement, Analysis, and Interpretation. *Journal of Speech, Language, and Hearing Research* 55(4). 1232–1246.

Stevens, Kenneth & Hanson, Helen. 2010. Articulatory-Acoustic Relations as the Basis of Distinctive Contrasts. In Hardcastle, William & Laver, John & Gibbon, Fiona (eds.), *The Handbook of Phonetic Sciences*, 2nd ed., 424–453. Malden: Wiley.

Stone, Maureen. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics* 19. 455–501.

Stone, Maureen. 2010. Laboratory Techniques for Investigating Speech Articulation. In Hardcastle, William & Laver, John & Gibbon, Fiona (eds.), *The Handbook of Phonetic Sciences*, 9–38. Malden: Wiley.

Stone, Maureen. 2013. Laboratory Techniques for Investigating Speech Articulation. In Hardcastle, William & Laver, John & Gibbon, Fiona (eds.), *The Handbook of Phonetic Sciences,* 2nd ed., 7–38. Malden: Wiley.

Teston, Bernard. 2004. L'œuvre d'Etienne-Jules Marey et sa contribution à l'émergence de la phonétique dans les sciences du langage. *Travaux Interdisciplinaires du Laboratoire Paroleet Langage d'Aix-en-Provence (TIPA), Laboratoire Parole et Langage* 23. 237–266.

Tillmann, Hans. 2006. Experimental and Instrumental Phonetics: History. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*, vol. 9, 374–389. Boston: Elsevier.

Töger, Johannes & Sorensen, Tanner & Somandepalli, Krishna & Toutios, Asterios & Lingala, Sajan & Narayanan, Shrikanth & Nayak, Krishna. 2017. Test-retest repeatability of human speech biomarkers from static and real-time dynamic magnetic resonance imaging. *The Journal of the Acoustical Society of America* 141(5). 3323–3336.

Vaïsse, Leon. 1875. Discours du Président. *Bullettin de la Société de linguistique de Paris* 1-5. clj–civij.

Wang, Jun & Samal, Ashok & Green, Jordan. 2014. Across-speaker Articulatory Normalization for Speaker-independent Silent Speech Recognition. *INTERSPEECH 2014*. http://www.isca-speech.org/archive/interspeech_2014. (consultato il 29.10.2020).

Wells, John. 2006. Phonetic transcription and analysis. In Brown, Keith (ed.), *Encyclopedia of Language and Linguistics*, vol. 9, 386–396. Boston: Elsevier.

Ximenes, Blackwood & Shaw, Jason & Carignan, Christopher. 2017. A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *The Journal of the Acoustical Society of America* 142(1). 363–377.

Alessandro Vietti
(Libera Università di Bolzano)

Il ruolo della variabilità acustica nella costruzione del dato linguistico

1. *Introduzione*

Lo studio strumentale della dimensione fonetica del parlato condotta negli ultimi decenni ha messo in nuova luce il ruolo della variabilità nello sviluppo e nell'organizzazione delle categorie fonologiche (Cohn *et al*. 2012; Hinskens *et al.* 2014). In particolare, il paradigma di ricerca della *Laboratory Phonology* si propone di costruire una teoria del funzionamento dei sistemi fonologici che comprenda al proprio interno i processi di variazione connessi ai meccanismi di produzione e percezione dei suoni linguistici (Pierrehumbert *et al.* 2012). In altri termini, si può sostenere che ci sono delle buone ragioni empiriche per stabilire una relazione molto stretta tra fonologia, fonetica e ambiente sociale esterno (Docherty & Foulkes 2014).

Gli obiettivi primari di questo contributo sono pertanto di illustrare quali siano i tipi di variabilità connessi con la dimensione fonetica del parlato e di discuterne la rilevanza sul piano teorico. Gli studi esemplificati e le riflessioni che ne discendono hanno la funzione di chiarire l'importanza del ruolo della variabilità acustica per lo studio della fonologia e, di conseguenza, mostrare la necessità metodologica di includere questa dimensione nella costruzione di dati di parlato (Voghera 2017: cap. 2). In questo modo, e solo a un livello molto generale, si deriveranno alcune indicazioni per la pratica metodologica.

Volendo essere molto sintetici, si può ridurre il contenuto di questo articolo a poche righe: la variabilità fonetica è importante per la comprensione della grammatica della lingua parlata e per questo è consigliabile non eliminarla nella costruzione del dato.

Come ultima avvertenza, poiché il contributo ha l'obiettivo di portare alcune riflessioni teoriche nella pratica metodologica della trascrizione e costruzione dei dati linguistici, la sua natura non è certo quella di

una rassegna sistematica dei temi e dei risultati scientifici contenuti nella vastissima letteratura in fonetica e fonologia sperimentali o in sociolinguistica variazionista.

2. *Assunti relativi alla variazione fonetica*

Come osservato, il tema della variabilità del parlato trova uno spazio particolare all'interno della prospettiva teorica di *Laboratory Phonology*. In questo contesto la variabilità non viene trattata unicamente in termini di variazione sociale, per così dire esterna al sistema linguistico, ma riveste un'importanza teorica per la formulazione di una teoria fonologica che sia empiricamente fondata (cfr. Foulkes & Docherty 2006 e, secondo una prospettiva di fonologia ottimalista, Anttila 2002).

Vorrei introdurre a questo punto tre proposizioni intorno alla variazione fonetica: le prime due hanno una natura di assunti di base (quasi assiomatica), mentre la terza mantiene ancora uno status di ipotesi di ricerca, piuttosto forte, che necessita ulteriori prove.

Le tre proposizioni sono le seguenti:

P1 il parlato è estremamente variabile;

P2 la variabilità rende disponibili delle informazioni indessicali relative al parlante e al contesto d'uso che sono compresenti a quelle linguistiche in senso stretto;

P3 l'informazione indessicale veicolata è integrata nella competenza fonologica.

La prima proposizione appare evidentemente come una sorta di ovvia constatazione se non se ne fornisce un'adeguata specificazione. Si tratta piuttosto del riconoscimento di una variazione intrinseca al parlato legata, da una parte, alle caratteristiche fisiologiche di chi produce e percepisce i segnali e, dall'altra, a quelle fisiche dei segnali stessi che si propagano attraverso il mezzo acustico e quello visivo (P1). Le informazioni veicolate dai segnali sono tradizionalmente divise in quelle linguistiche, corrispondenti grosso modo alle unità discrete del linguaggio, e altre non-linguistiche, maggiormente legate alle caratteristiche continue del segnale. Queste ultime sembrano fornire ai parlanti-riceventi dati utili sulle caratteristiche sociali di chi parla o, in genere, sul contesto

comunicativo (P2). L'idea alla base è di comprendere se queste informazioni vengano completamente scartate nella processazione del linguaggio, e quindi nella rappresentazione delle unità, oppure, al contrario, svolgano un ruolo di facilitazione del processo di comprensione del messaggio e, più in generale, siano integrate nella competenza fonologica (P3).

Sebbene tutti e tre questi punti verranno affrontati di seguito, non saranno tuttavia trattati in modo ugualmente esteso, il punto 3 sarà infatti quasi completamente sacrificato.

2.1 Presenza di variabilità

Nella tradizione italiana di ricerca in sociolinguistica, al tema della variabilità della lingua è associata prototipicamente la nozione di architettura della lingua nella sua rappresentazione geometrica fornita da Berruto (2012). In questa visione possiamo quindi definire le varietà come oggetti di analisi e osservarne la disposizione relativa nello spazio multidimensionale definito sulla base di Coseriu (1969; e di recente discusso teoricamente in Berruto 2015).

Alternativa meno frequente nella sociolinguistica italiana è quella di osservare il comportamento di elementi della grammatica che variano in relazione all'azione sinergica di fattori linguistici e sociali. Questi elementi, denominati variabili (socio)linguistiche (a partire da Labov 1963), sono tendenzialmente unità discrete della grammatica, più frequentemente fonetico-fonologiche, che svolgono una medesima funzione sul piano strutturale, ma sono suscettibili di ricevere un differente valore sociale all'interno della comunità dei parlanti (Vietti 2019).

Osservando il parlato più da vicino, nei suoi meccanismi di produzione e percezione, per intenderci, possiamo rilevare un ulteriore dominio di variabilità della lingua. Per raggiungere questo micro-dominio di variabilità non è sufficiente l'uso dell'orecchio, dobbiamo infatti dotarci di strumenti di osservazione e misurazione[1]. Si entra in questo modo in un ambito della realtà in cui fenomeni continui e discreti si congiungo-

[1] Con questo non si vuole certo implicare che i fenomeni in questo dominio non siano percepibili (altrimenti non avrebbe neppure senso studiarli); al contrario lo sono ma, come esseri umani, non siamo in grado di osservarli in modo consapevole ed esplicito, se non utilizzando degli strumenti.

no e intrecciano. Da un lato, abbiamo i suoni del linguaggio che si realizzano attraverso dimensioni fisiche continue[2] come per esempio il tempo, la frequenza e l'intensità, dall'altra invece troviamo, per semplificare, degli oggetti discreti e delimitati come i segmenti che, a loro volta, fanno riferimento ai fonemi.

Su questo piano microscopico, gli oggetti di indagine sono dunque i segnali continui usati per veicolare il messaggio linguistico. O meglio, non i segnali continui in quanto tali, ma in quanto portatori di pacchetti discreti di informazione. Limitiamoci in questo contesto a osservare il campo dei suoni e, all'interno di questo, quello che potremmo, per semplicità, definire dei segmenti (opposto a quello di unità prosodiche più ampie).

Su questo piano di osservazione ci sono diversi fattori che possono indurre variabilità:

1. i meccanismi e i processi di produzione e percezione (es. controllo del sistema motorio periferico, percezione e *mapping* del segnale acustico, Hoole *et al.* 2012; Hoole & Pouplier 2015; Stevens & Hanson 2010; Moore 2010; Whalen 2019);
2. la variabilità fonologica contestuale, determinata da processi di coarticolazione e adattamento contestuale (Farnetani & Recasens 2010; Recasens 2018);
3. l'uso in un contesto comunicativo (per esempio la distinzione tra parlato spontaneo, non pianificato, e perciò tendenzialmente ipoarticolato, e controllato, pianificato e di norma iperarticolato, cfr. Ernestus 2012; Lindblom 1990);
4. le caratteristiche individuali della voce di un parlante (da quelle fisiologiche alle caratteristiche sociali; cfr. Johnson 1997; Foulkes & Docherty 2006).

In questo modo, il campo dei fenomeni osservabili non solo si estende, ma intreccia anche domini di natura differente: in generale, quello continuo e quello discreto e, più in particolare, quello acustico relativo ai suoni e quello simbolico rappresentato dalle categorie linguistiche[3].

[2] In senso intuitivo e impreciso con continuo possiamo intendere un fenomeno che non ha interruzioni al suo interno.

[3] In una prospettiva di comprensione multimodale del processo di comunicazione si dovrebbero integrare anche altri ambiti dell'esperienza, in primo luogo quello motorio.

Qui già si può intravedere l'insorgere di un problema che non è solo di natura teorica, ma si presenta anche sul piano metodologico, ovvero quello della relazione tra domini nella rappresentazione del dato trascritto. Infatti, se ipotizziamo che le categorie discrete che compongono le strutture del linguaggio nulla abbiano a che fare con i fenomeni continui, allora tutto questo discorso è di fatto inutile. Ma se ammettiamo, e ci sono buone prove empiriche per farlo, che il mondo continuo e quello discreto siano in stretto contatto, o in altri termini che le categorie discrete contengano tracce del mondo continuo, o addirittura poggino su di esso, allora è necessario comprendere meglio come siano legati e cosa li leghi. Alcune implicazioni di questa ipotesi saranno trattate più avanti.

2.2 Variabilità come fonte di informazione

Una volta individuate le possibili dimensioni di variabilità si può passare alla seconda proposizione (P2), ovvero esaminare come la (micro)variabilità presente nel segnale nell'atto di produzione e percezione del parlato costituisca una fonte di informazioni da integrare con quelle linguistiche. Se adottiamo la prospettiva di un programma di ricerca, si tratta a questo punto di comprendere (a) che tipo di informazioni si rendano disponibili ai parlanti-ascoltatori, (b) come queste informazioni vengano utilizzate all'interno del processo di comunicazione e infine (c) che posto occupino questi elementi nella rappresentazione delle categorie linguistiche (in senso stretto).

Nei due sottoparagrafi che seguono verrà precisata la natura delle informazioni veicolate dal parlante e dal parlare in contesto, ovvero la quarta e la terza dimensione di variazione individuate nel paragrafo 2.1.

2.2.1 Informazioni legate al parlante

Il medium fonico-acustico utilizzato per trasmettere un messaggio linguistico rappresenta anche un vettore di informazioni di varia natura sul parlante. Come ascoltatori siamo di norma in grado di inferire, se è pertinente, delle caratteristiche sociali del parlante – come l'area geografica di provenienza e/o l'appartenenza a gruppi sociali, linguistici o culturali – oppure dei tratti più individuali come l'età, il genere o lo stato emotivo (Abercrombie 1967: 5-9; Foulkes & Docherty 2006).

Questo tipo di variabilità viene definita indessicale poiché il meccanismo semiotico di associazione tra le caratteristiche del segnale e il contenuto veicolato si basa sulla compresenza fisica di significato e significante (Foulkes 2010; Vietti 2014). Per variabilità indessicale si intende perciò l'insieme delle informazioni immediatamente associabili alle caratteristiche del parlante: da quelle più idiosincratiche (che ci permettono per esempio di riconoscere una specifica voce), fino a quelle associate a significati sociali condivisi da un'intera comunità.

Il dominio dei tratti linguistici utilizzabili per indicizzare dei contenuti extra-linguistici coincide quasi completamente con le possibilità fenomenologiche dell'interfaccia fonetico-fonologica. La gamma degli elementi variabili è definita da un insieme che comprende fenomeni di variabilità sub-segmentale, allofonica, fonemica, fonotattica, sillabica, accentuale, ritmica fino a raggiungere la realizzazione degli schemi intonativi nei costituenti prosodici maggiori e la qualità della fonazione (Foulkes & Docherty 2006: 412-419).

L'estensione del dominio dei fatti osservabili al di sopra e al di sotto del segmento rappresenta una novità nello studio della variazione introdotta dall'uso sistematico dell'analisi strumentale negli studi (socio)fonetici. In questo modo oltre alla variazione tra categorie segmentali è possibile documentare anche aspetti più dettagliati del parlato in termini di parametri continui di durata, intensità, frequenza, nonché di evoluzione temporale dei suoni. In una prospettiva metodologica, questo aspetto riveste un particolare rilievo. Per determinare differenze gradate tra quantità è necessario che queste trovino uno spazio nella rappresentazione del dato, in modo da poterle richiamare nell'analisi e collegare con unità discrete di tipo fonologico o di livello superiore.

Questa recente tradizione di indagine sugli attributi indessicali del parlato ha origine in una serie di studi condotti nell'ambito della psicolinguistica tra i quali si possono sicuramente segnalare Palmeri *et al.* (1993), Goldinger (1996) e Goldinger (1998). Questi studi avevano l'obiettivo di comprendere se e come la memoria episodica, ovvero quella che immagazzina dettagli di specifiche esperienze, interagisse con le rappresentazioni linguistiche astratte. In particolare, gli esperimenti condotti mostrano come le informazioni acustiche legate idiosincraticamente a una specifica voce vengano trattenute nella memoria della pronuncia di una parola e ne facilitino l'identificazione e il riconoscimento anche a distanza di tempo. In altre parole, a parità di condizioni sperimentali, se una parola viene pro-

nunciata da una voce già sentita in precedenza, questa viene identificata e riconosciuta più facilmente di altre, pronunciate con voci nuove e mai sentite. I risultati di questi esperimenti pongono perciò un problema nuovo nella rappresentazione fonologica del lessico, ovvero indicano che, accanto a una rappresentazione lessicale astratta, ne può esistere un'altra, costituita da un insieme di memorie dettagliate (chiamate esemplari) di un dato item, e che questi due piani di rappresentazione interagiscono tra di loro quando circostanze e obiettivi comunicativi lo richiedano.

Su queste prime prove sperimentali si sono poi accumulate evidenze empiriche sempre più consistenti – nell'ambito della fonologia sperimentale e della sociofonetica (tra questi cfr. Johnson 1997; Hawkins 2003; Hay *et al.* 2006; Pierrehumbert 2001; Todd *et al.* 2019) – relative al ruolo delle informazioni legate alle caratteristiche specifiche del parlante. Questo genere di informazioni, che possono sembrare a prima vista irrilevanti se si osserva un sistema linguistico nel suo insieme, costituiscono tuttavia il ponte attraverso il quale i parlanti di una comunità possono formulare delle generalizzazioni di tipo sociale a partire da informazioni linguistiche. Rappresentano, per così dire, il punto di innesco di un processo di costruzione di varianti e variabili linguistiche alle quali vengono associate informazioni di contesto, prima di portata locale e via via più generali e condivise.

Come gli insiemi di esemplari siano organizzati, come siano collegati alle categorie astratte, in che modo e in quali circostanze interagiscano con tali categorie restano tutti problemi aperti da risolvere nei prossimi anni all'interno di questo paradigma di ricerca fonologica.

2.2.2 Parlare: informazioni legate al contesto d'uso della lingua

Lo studio dell'uso della lingua in un contesto comunicativo spontaneo, secondo un approccio strumentale, ha amplificato la consapevolezza dell'esistenza di una estrema variabilità interna al parlante indotta dai meccanismi di produzione e ricezione del parlato spontaneo[4] (Ernestus

[4] In questo senso l'unione di approccio strumentale e osservazione di contesti comunicativi spontanei ha determinato una visione arricchita della gamma di fenomeni ascrivibili alla natura della lingua parlata che si inserisce, per l'italiano, in un quadro empirico e teorico già ben delineato dagli studi a cavallo tra gli anni Ottanta e Novanta come per esempio da Bazzanella (1994), Berretta (1994), Berruto (1985), Sornicola (1981), Voghera (1992). L'attenzione per l'uso della lingua parlata è di recente al centro di un rinnovato interesse come testimoniato anche da Voghera (2017).

2012). Il tema della variabilità presente nel parlato connesso non rappresenta un tema nuovo nella ricerca fonetica[5], o linguistico più in generale. La novità risiede piuttosto nelle nuove possibilità di acquisizione, conservazione e analisi di ampie basi di dati di parlato spontaneo e dialogico raccolto al di fuori dei consueti *task* comunicativi da laboratorio. Il quadro che si delinea permette di osservare una variabilità fonologica quantitativamente e qualitativamente differente (Ernestus & Baayen 2011).

La variabilità che emerge dall'analisi di corpora di parlato non appare casuale, non si tratta di "rumore" non sistematico generato dal meccanismo di produzione e ricezione, ma piuttosto di un processo che, su ampia scala e a lungo termine, ha una ricaduta sull'organizzazione delle categorie linguistiche (Bybee 2001). Si pone pertanto la necessità di trovare uno spazio nella rappresentazione fonologica per questi processi o, anche per questa fonte di variabilità, di determinare quale legame sussista tra la variabilità fonetica e le categorie fonologiche.

All'interno del dominio segmentale si assiste per esempio a un aumento dei fenomeni di assimilazione o di coarticolazione (dovuta anche a nuovi contesti fono-sintattici creati dalla catena parlata) e di riduzione della forma. La reale portata di questi processi negli stili più spontanei e informali rimane ancora in larga parte inesplorata.

Emblematici in questa direzione sono gli studi sul parlato conversazionale spontaneo condotti da Ernestus (a partire da Ernestus 2000) che rivelano l'esistenza di un'ampia gamma di forme disponibili per ogni lessema, distanti dalla forma canonica (tipica della produzione in isolamento). Queste forme sono il risultato di un generale processo di riduzione della forma della parola che può assumere l'aspetto di lenizione, assimilazione parziale o totale e di cancellazione del segmento. La combinazione di questi processi di riduzione produce un ampio insieme di possibili varianti della stessa parola, come per esempio nella parola nederlandese *natuurlijk* 'naturalmente' dove accanto alla forma di citazione [natyrlək] troviamo pronunce come [natyləkˌ], [ntyləkˌ], [ntykˌ], [tyrləkˌ], [tyləkˌ], [tylkˌ], [tykˌ], [tygˌ], [dykˌ], e [dygˌ] (Ernestus 2000: 137-143). Questa ampia variabilità solleva evidentemente molti interrogativi sulle possibili cause e sollecita perciò ulteriore ricerca per comprendere:

[5] Si pensi per esempio alla tradizione di ricerca sul continuum di ipo- e iper-articolazione del parlato elaborato nella teoria H&H di Lindblom (1990).

(a) quali processi fonetici e fonologici determinino le specifiche forme, ovvero quali proprietà segmentali debbano essere conservate per rendere la parola utilizzabile;
(b) quale effetto esercitino i diversi contesti prosodici in cui le parole compaiono sulla loro riduzione;
(c) quale ruolo svolgano la riconoscibilità, la prevedibilità in contesto e la frequenza delle parole sulla riduzione;
(d) se si tratti di un fatto esclusivo e periferico di implementazione del parlato, tendente a svanire, oppure di un processo che, attraverso l'uso ripetuto, incide sulle rappresentazioni lessicali e/o assume un significato indessicale di spontaneità e informalità.

Insomma, come si può facilmente osservare le questioni aperte da ricerche in questo campo non sono certo di poco conto per la piena comprensione del funzionamento della lingua parlata.

2.2.3 Variabilità e competenza fonologica

Come anticipato, non vi è in realtà spazio qui per trattare la terza proposizione riportata in §1 che riguarda l'integrazione della variabilità acustica all'interno della competenza fonologica. Si tratta di un compito che eccede gli obiettivi di questo scritto e, d'altra parte, bisogna pure constatare come molti dei temi di ricerca emersi sin qui siano ben lontani dall'aver trovato una sistematizzazione teorica. L'unico elemento che appare sufficientemente certo è che la presenza della variabilità non sia un fenomeno esclusivamente di esecuzione, ma vi sia una relazione strutturale con gli elementi costitutivi del sistema fonologico. La natura della relazione rimane oggetto di speculazioni e di proposte teoriche ancora in attesa di validazione. Tra queste la più promettente è senza dubbio la teoria degli esemplari così come presentata, per esempio, in una delle sue più recenti proposte modellistiche formali in Todd *et al.* (2019).

A riassumere brevemente quanto sin qui illustrato nel paragrafo 2, è importante sottolineare come il contributo della variabilità acustica al processo di costituzione delle categorie fonologiche non debba essere intesa come una proposta alternativa e sostitutiva alla presenza e alla funzione svolta dalle categorie simboliche, astratte e completamente indipendenti dal contesto. Al contrario il grande interesse scientifico, almeno dalla prospettiva della *Laboratory Phonology*, risiede proprio nel-

la complessa impresa di dimostrare come questi due livelli di rappresentazione possano coesistere e interagire. A mo' di chiusura di questo paragrafo è opportuno riportare la riflessione di Pierrehumbert (2012: 174) sulla natura della competenza fonologica:

> According to the phonological principle, forms of words (word forms) are combinations of basic building blocks, which are characteristic of any individual language, meaningless in themselves, but meaningful in combination. Evidence has recently accumulated that, in addition to this abstract level of characterization, lexical entries also include density distributions over detailed phonetic or socio-indexical properties. I accordingly view word forms as both detailed and abstract.

3. *Questioni metodologiche*

Dalla rapida rassegna di problemi tracciata fin qui, appare chiaro come (a) esistano diverse dimensioni di variabilità legate al segnale acustico e (b) la variabilità risultante renda disponibili delle informazioni legate al parlante e al processo di comunicazione. Dunque, nella prospettiva del ricercatore, il problema da porsi a questo punto riguarda le modalità di conservazione, trascrizione e rappresentazione di questo tipo di informazioni. La soluzione a questo problema si orienta tra due grandi possibilità: la normalizzazione e la conservazione. Se si effettua un'operazione di normalizzazione, il segnale viene filtrato eliminando le informazioni di dettaglio considerate di livello troppo locale e non generalizzabili sul piano linguistico, in poche parole si va alla ricerca degli elementi invarianti e si getta via quanto non è pertinente sul piano linguistico (in questo caso fonologico). La strategia per certi versi opposta è quella di conservare una memoria fedele di tutto quanto accade, immagazzinando gli esempi di comunicazione in grandi raccolte sempre disponibili.

Nessuna di queste due vie, se intese in modo esclusivo, rappresenta una soluzione al problema della rappresentazione del dato linguistico: da un lato, avremmo, per esempio, un corpus di dati rappresentati unicamente tramite una codifica fonetica o fonologica discreta (o addirittura ortografica) senza più traccia del segnale acustico continuo; dall'altro, avremmo invece una collezione di registrazioni acustiche senza nessuna forma di codifica o annotazione e, di conseguenza, senza possibilità di stabilire dei legami tra oggetti interni alla raccolta.

Queste due operazioni, seppure distinte, non sono in realtà in contraddizione. Sia sul piano cognitivo, sia su quello più semplicemente metodologico, un misto di normalizzazione e conservazione non solo è possibile, ma notevolmente utile. Le possibilità di osservazione strumentale e le capacità di conservazione digitale dell'informazione rendono infatti possibile conciliare astrazione (normalizzazione) e memoria dettagliata (conservazione).

La via per trovare un bilanciamento tra normalizzazione e conservazione passa attraverso un processo di selezione delle diverse dimensioni di informazione e, successivamente, di individuazione delle strategie di rappresentazione. Si tratta di rispondere operativamente a due istanze:

(a) Quali tipi di informazioni costituiscono il dato (socio)fonetico?

(b) Come e "dove" si trascrivono queste informazioni?

La prima domanda pone diversi problemi. Il primo, evidente, è quello dell'individuazione delle dimensioni di variazione. Fin qui abbiamo affrontato le diverse dimensioni connesse con le caratteristiche del segnale acustico, ma questo fa parte di un sistema integrato di segnali multimodali prodotti attraverso il corpo e finalizzato alla trasmissione di informazioni (cfr. Levinson & Holler 2014). Le diverse modalità possono perciò idealmente entrare a far parte del dato linguistico rendendo sempre più complesso il compito di delimitare il dato grezzo, di partenza.

Da questa constatazione deriva un secondo problema che riguarda invece la delimitazione del campo di interesse, sia in ampiezza che in profondità. Se osserviamo il piano fonetico acustico, le possibilità di parametrizzazione dei suoni linguistici sono davvero molteplici: dall'estrazione di singole misure a intere matrici di misure distribuite nel tempo. Tuttavia, come vedremo nei prossimi sottoparagrafi, un criterio per la delimitazione delle informazioni da considerare è quella di valutare sia il ruolo svolto nel processo di comunicazione, sia la presenza di una funzione linguistica[6].

[6] Volendo derivare le implicazioni più critiche contro questa linea di ragionamento bisognerebbe dire che, da un lato, c'è circolarità nell'argomentazione (per sapere cosa è pertinente è necessario includere tutto), dall'altro ci sono dati che possono essere irrilevanti per l'utente umano, ma ottimali per le capacità iper-descrittive della macchina (in questo caso è implicata evidentemente anche la finalità per la quale il dato viene costruito).

La seconda domanda richiama invece un problema molto pratico che riguarda il modo in cui queste informazioni possano essere integrate in un dato in modo da ottenere una rappresentazione ricca ma, allo stesso tempo, indagabile. Questo aspetto si articola in due componenti: la prima riguarda la scelta del livello di granularità nella discretizzazione del dato acustico (strategie di rappresentazione) e la seconda riguarda invece la modalità di connessione delle categorie discrete con il dato acustico continuo (allineamento temporale).

Una rapida discussione di questi aspetti è contenuta rispettivamente nei paragrafi 3.1 e 3.2.

3.1 Dato fonetico di partenza

Come abbiamo potuto osservare l'analisi strumentale del parlato prevede la misurazione di variabili continue sia nell'ambito delle onde acustiche, sia in quello motorio, legato primariamente alla coordinazione dei movimenti degli articolatori. Se pensiamo al primo ambito, registrando digitalmente la nostra voce possiamo facilmente disporre di descrizioni dettagliate dei cambiamenti continui nel tempo delle caratteristiche spettrali. La ricchezza di queste descrizioni nel tempo, in frequenza e in intensità supera persino le capacità umane di percezione. Per comprendere l'ampiezza di tali descrizioni possiamo pensare al fatto che un segmento [a] di circa 140 ms (Fig. 2) può, per esempio, essere definito da una matrice di 23 rilevazioni temporali per 12 coefficienti[7].

Questi coefficienti rappresentano dei parametri acustici che sintetizzano proprietà rilevanti del segnale e, anche se sembra incredibile, costituiscono già una riduzione dell'informazione complessiva disponibile. Si tratta evidentemente di descrizioni non solo molto ricche nel tempo e in frequenza, ma anche molto distanti da quelle fonologiche come, per esempio, quelle per tratti distintivi alle quali è abituato il linguista. Il segmento [a] può infatti essere ricondotto al fonema /a/ e specificato con uno o pochi tratti, p.e. [+ basso] e [+ arretrato], a seconda del sistema fonologico in cui compare[8]. Il nodo però non è decidere quale sia la rappresentazione migliore tra queste, ma comprendere come questi due

[7] Si tratta per esempio di coefficienti cepstrali usati tradizionalmente nei sistemi di riconoscimento automatico del parlato e sono il risultato di un processo di elaborazione numerica a partire dall'informazione relativa all'onda acustica (digitalizzata; cfr. Iskarous 2018).

[8] E della teoria fonologica di specificazione che vogliamo adottare.

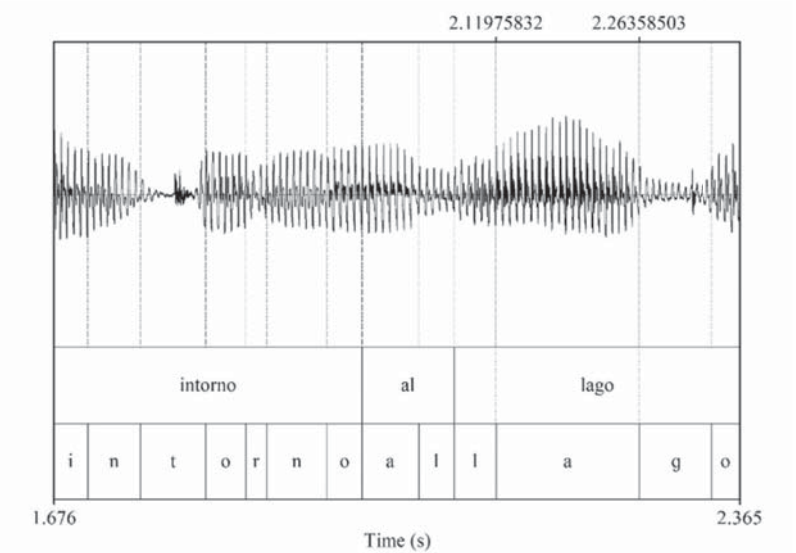


Fig. 1. Forma d'onda (in alto) e trascrizione ortografica e fonologica (in basso) tratta dal corpus DIA (Mereu & Vietti, 2021).

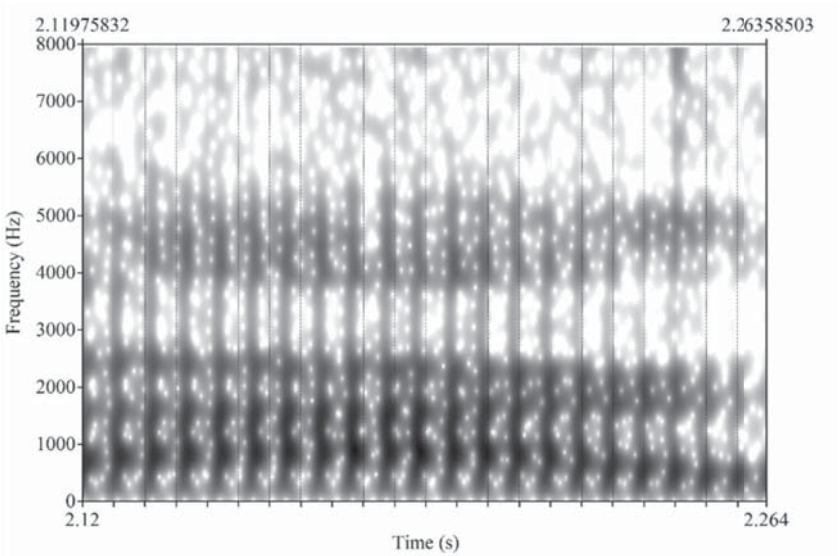


Fig. 2. Spettrogramma di [a] *in* [ˈla:go]*, le barre verticali indicano i 23 istanti in cui sono stati estratti i coefficienti.*

livelli di descrizione possano essere collegati cognitivamente, ovvero cercare di definire quale possa essere un modello cognitivamente realistico della competenza fonologica in grado di spiegare la compresenza di questi due livelli di rappresentazione.

Se a questa considerazione aggiungiamo che la variabilità intra- e intersoggettiva nella realizzazione di quella matrice di 23 x 12 osservazioni svolge un ruolo rilevante nel processo di produzione e comprensione, dovrebbe apparire chiaro come le modalità di codificare il dato costituiscano un aspetto decisivo per lo studio della fonologia in una prospettiva empirica.

La ricchezza informativa del dato acustico porta con sé anche il rischio di una iper-descrizione della realtà. Questa possibilità è stata ben sottolineata da Labov nella sua critica all'analisi della variabilità sociofonetica. La ricerca del dettaglio nell'individuazione di suoni e voci deve arrestarsi a un certo punto sulla base di una buona giustificazione scientifica altrimenti, sostiene Labov (2006: 508), ci si imbarcherebbe in un "endless pursuit of detail".

Sulla relazione tra necessità di un limite alla ricchezza e precisione della descrizione fonetica va operata però una distinzione cruciale tra la rappresentazione del dato e la sua descrizione. La prima può essere infatti anti-economicamente ricca, ampia e vicina alla realtà fisica, mentre la seconda deve essere necessariamente più sintetica e ristretta. Il discrimine di quanto dovrebbe entrare in una descrizione linguistica è rappresentato dalla presenza di una chiara funzione socio-cognitiva svolta dagli elementi analizzati nel sistema linguistico o nel processo comunicativo.

Come linguista interessato alla descrizione di una varietà di lingua devo innanzitutto comprendere qual è il livello di variabilità fonetica che riveste un valore all'interno del sistema. Per esempio, nella descrizione della variabilità fonetica di un fonema nella varietà A sarà utile arrestarsi in prima battuta alla variazione nelle forme fonetiche che rivela una relazione con i contesti fonetici in cui il segmento compare, determinando così una variabilità allofonica nel sistema[9].

[9] Ogni affermazione fatta andrebbe noiosamente precisata, qui sì in una regressione infinita, perché: (a) potremmo voler distinguere tra una allofonia posizionale, determinata dalla coarticolazione, e una di tipo più fonologico dal contesto segmentale; (b) la relazione allofonica rivelata è di norma di natura probabilistica e non deterministica; (c) si dovrebbe distinguere tra sistema e processazione (la variabilità coarticolatoria delle forme potrebbe non svolgere un ruolo sul piano del sistema fonologico, ma di sicuro essere fondamentale nella comprensione online del linguaggio, in quanto la coarticolazione favorisce la previsione di sillabe e parole successive).

Avendo però sostenuto fin qui che le informazioni sulla variabilità sociale del segnale acustico non sono secondarie nel processo di produzione e comprensione della lingua parlata è chiaro che nel descrivere una varietà non potrò limitarmi unicamente al valore "interno" al sistema[10], ma dovrò anche includere nella descrizione la variabilità tra le forme che è suscettibile di ricevere un significato sociale.

In conclusione, poiché la funzione linguistica e il valore sociale delle diverse forme non sono conosciute a priori, ma sono il risultato della descrizione e analisi del ricercatore, la rappresentazione del dato deve necessariamente comprendere un insieme più ampio di informazioni rispetto a quelle derivate dall'osservazione e dall'analisi del ricercatore. Insomma, escludere per principio le informazioni sulla variabilità acustica significa prendere una decisione teorica che preclude a monte la possibilità che questa dimensione eserciti un effetto sul funzionamento del linguaggio[11].

3.2 Rappresentazione del dato e allineamento temporale

Se l'obiettivo è di prevedere un ruolo per la variabilità fonetica all'interno dello studio del linguaggio, in che modo questa può essere catturata sul piano metodologico?

Una prima semplice risposta è che per svolgere una qualsiasi analisi su un segnale acustico dobbiamo poterlo rappresentare sotto forma di categorie discrete, dobbiamo infatti trovare un modo per classificare questi oggetti continui e trasformarli in categorie, poiché solo in questo modo possiamo confrontarli tra loro. Si tratta, in primo luogo, di segmentare il segnale e, successivamente di assegnare un'etichetta al segmento identificato. L'attuazione di questo processo implica il problema della selezione della "taglia" delle etichette, ovvero del grado di precisione dell'etichettatura in termini di proprietà acustiche e articolatorie che vogliamo siano rappresentate nella trascrizione.

Una soluzione unica a questo problema non esiste, poiché la scelta ottimale dipende dalle specifiche esigenze di ricerca. Per esempio, uno studio sull'acquisizione fonologica della prima lingua da parte di bam-

[10] Anzi, volendo assumere sul serio la prospettiva sociofonetica non potrò più nemmeno distinguere con assoluta certezza tra ciò che è "interno" e fonologico e ciò che è "esterno" e sociofonetico.

[11] Posizione non nuova per altro, ma che si scontra con l'evidenza empirica.

bini sarà orientato a un particolare grado di accuratezza nella trascrizione, proprio perché non si conoscono a priori le funzioni attribuite alle diverse forme, inoltre non si sa se le forme prodotte siano legate a un processo sistematico o frutto della variazione casuale dovuta a un non completo controllo del sistema motorio che regola l'articolazione (Vihman 2014: 83-84). Se invece l'obiettivo sarà quello di costruire un corpus sociolinguisticamente variato di lingua parlata, allora le scelte metodologiche da compiere saranno differenti.

Le opzioni metodologiche del ricercatore ruotano a questo punto attorno al livello di precisione nella classificazione del segnale acustico. Il primo pericolo da fronteggiare è rappresentato dall'idea di perseguire la massima accuratezza della trascrizione. Le conseguenze di un eccesso di dettaglio nella classificazione fonetica sono ben esemplificate in Vietti & Spreafico (2008, 2016). In questi studi, la ricerca della massima accuratezza nella classificazione spettrografica porta inizialmente all'individuazione di 15 varianti (poi ridotte a 9) del fonema /r/ nell'italiano di Bolzano, suddivise in due serie per luogo di articolazione (uvulare e alveolare)[12]. La situazione di contatto linguistico tra la comunità italofona e quella tedescofona giustifica solo in parte l'alto grado di variazione nella forma, infatti le due serie di rotiche costituiscono una rappresentazione estremamente accurata delle caratteristiche uditive e spettro-acustiche dei segmenti analizzati. L'elemento problematico è costituito proprio dall'eccesso di descrizione e dalla conseguente scarsa pertinenza linguistico-funzionale e sociolinguistica dell'intero insieme di varianti. L'analisi condotta in Vietti & Spreafico (2016) rivela infatti che le varianti, grazie a un'analisi statistica multivariata, possono essere ricondotte a sottogruppi in ragione del loro comportamento coerente, sia sul piano della distribuzione per contesti fonetici, sia per quanto riguarda il valore sociale (in questo caso l'attribuzione di una identità linguistico-culturale di "italiano" o "tedesco"; cfr. anche Kaland *et al.* 2019).

[12] Le 15 varianti individuate sono: monovibrante alveolare [ɾ] e uvulare (non c'è un simbolo IPA accettato); monovibrante flap alveolare; flap laterale alveolare [ɺ]; polivibrante alveolare [r] e uvulare [ʀ]; retroflessa [ɽ]; approssimante alveolare [ɹ], retroflessa [ɻ] e uvulare [ʁ̞]; fricativa retroflessa [ʐ] e uvulare sonora [ʁ] e uvulare sorda [χ]; approssimante labiodentale [ʋ]; fricativa post-alveolare sorda [ʃ].

La strategia della massima accuratezza può essere percorribile quando si analizza una varietà di lingua sconosciuta, oppure, come osservato in precedenza, quando le finalità della ricerca richiedano un alto grado di precisione nella trascrizione del dato, oppure ancora quando si voglia osservare il comportamento di un'unica variabile. Come ulteriore corollario, una classificazione che preveda una lunga lista di categorie spesso porta con sé un problema di scarsa numerosità di occorrenze per alcune categorie. Questo aspetto rende molto difficile costruire un modello statistico significativo, affidabile e generalizzabile al di là dello specifico *dataset*.

La strategia della massima accuratezza è difficilmente praticabile quando si debba costruire una base di dati su vasta scala, come per esempio un corpus rappresentativo della variazione nella lingua parlata. In questo caso, infatti, è necessario adottare soluzioni differenti per raggiungere il duplice obiettivo di ottenere una rappresentazione dettagliata del dato e una categorizzazione sufficientemente astratta da consentire una facile interrogabilità. Questo obiettivo è duplice, ma non contraddittorio.

Osserviamo per un attimo la strategia opposta a quella della massima accuratezza, cioè quella di una trascrizione astratta, distante dalla realtà fonetica e perciò coincidente con una rappresentazione fonologica standard. Con questa scelta si rinuncia evidentemente alla trasposizione fedele della pronuncia nel dato trascritto, d'altra parte, il beneficio è rappresentato dalla presenza di una classe astratta che permette un efficace confronto tra gli oggetti che ricadono al suo interno. Gli oggetti appartenenti a questa categoria saranno presumibilmente piuttosto eterogenei tra di loro e taluni anche devianti rispetto al prototipo. In assenza di codifica esplicita della diversità interna alla categoria, la base per un confronto può tuttavia essere garantita introducendo una proprietà fondamentale dei corpora di lingua parlata, ossia l'allineamento temporale tra il segmento fonologico e il segnale acustico. Nell'esempio di corpus annotato e allineato[13] in Fig. 3 possiamo osservare come attraverso l'allineamento temporale si riesca ad ottemperare alle esigenze di astrazione e di massima rappresentazione della variabilità. L'astrazione in termini metodologici diventa sinonimo di facilità di interrogazione delle catego-

[13] L'esempio è tratto dal corpus DIA (Mereu & Vietti, 2021).

rie al di sotto delle quali sono sempre recuperabili le informazioni acustiche più dettagliate.

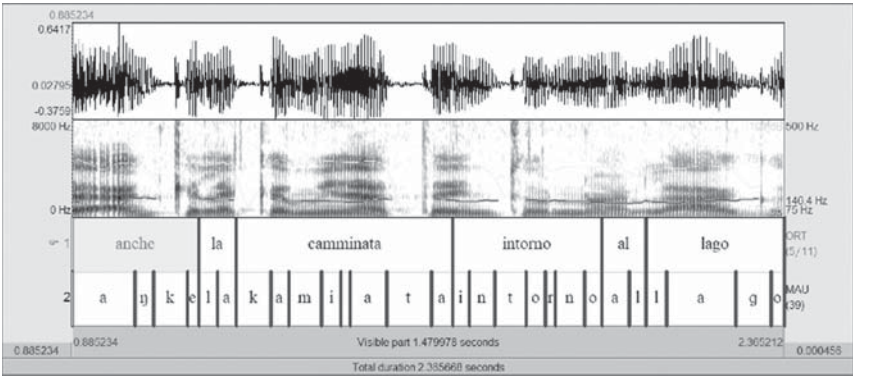


Fig. 3. Esempio di trascrizione ortografica e fonologica allineata temporalmente con forma d'onda e spettrogramma in un corpus di parlato spontaneo di italiano

In questo modo è possibile condurre analisi acustiche su tutte le occorrenze di un fonema in una data varietà e osservare, sulla base della loro distribuzione, se esistano sottogruppi. Grazie anche all'impiego di tecniche statistiche esplorative[14], saranno i dati stessi a mettere in evidenza la presenza di gruppi distinti di forme individuati sulla base di parametri acustici. Al ricercatore spetta quindi il compito di (a) specificare le nuove sottocategorie (al di sotto di quella fonologica) e (b) determinarne le funzioni, ricercando possibili correlazioni con fattori linguistici e sociali.

Un esempio di uso dell'analisi strumentale per esplorare i dati e scoprire nuove sottocategorie è presente nello studio di Mereu (2017) sulla variazione sociofonetica delle sibilanti in sardo cagliaritano. In questo caso la distinzione fatta tra quattro varianti di sibilante (alveolare sorda [s], alveolare sonora [z], alveo-laminale sorda [s̻] e postalveolare sorda [ʃ]) sulla base del parametro acustico denominato *Centre of Gravity*

[14] Come per esempio l'analisi per componenti principali (PCA) o la *cluster analysis*.

(CoG)[15] mette in evidenza una distribuzione anomala per la variante [z] (cfr. Fig. 4).

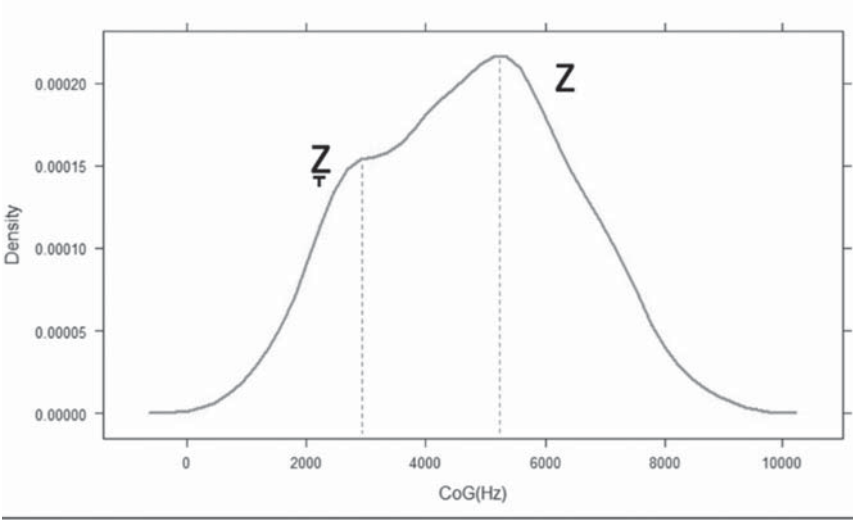


Fig. 4. Distribuzione di densità di probabilità del CoG di /z/ dei parlanti maschili (Mereu 2017)

L'anomalia consiste nel fatto che la distribuzione dei valori di questo parametro per [z] rivela la presenza di due picchi, cioè una distribuzione bimodale con due intervalli di valori considerati modali uno attorno a 2500 Hz e il secondo attorno 5500 Hz. Il primo picco, quello inatteso, ha permesso di rivelare la presenza di una serie di varianti di [z] il cui modo di articolazione era più approssimante ed era collegato a uno stile di elocuzione decisamente più ipoarticolato.

La possibilità di definire con precisione le categorie segmentali o eventualmente di affinarle scoprendone di nuove è consentito dall'allineamento temporale tra categorie e segnale acustico, una sorta di "for-

[15] Il CoG rappresenta una misura che indica in quale gamma di frequenza è concentrata la maggiore energia all'interno dell'inviluppo spettrale del suono prodotto. Questa misura discrimina con buona approssimazione il luogo di articolazione delle fricative, ovvero il rumore di frizione prodotto da questo modo di articolazione.

zatura" nella rappresentazione del dato che mette in relazione il piano continuo con quello discreto. Questa forzatura rende così superflua una rappresentazione del dato troppo ricca in termini di trascrizione, in quanto una precisa definizione può essere sempre demandata all'analisi diretta dei parametri continui. È chiaro che una scelta di questo tipo implica una qualità del dato acustico sufficiente per poter svolgere un'analisi strumentale affidabile e un'accurata trascrizione fonologica. Tuttavia, la possibilità di effettuare registrazioni digitali di alta qualità della voce umana è da tempo alla portata di chiunque e non è più una prerogativa di esperti della registrazione del suono.

La presenza dell'allineamento temporale permette dunque di condurre un'analisi strumentale sulle categorie fonologiche definite a livello di segmento, rendendo possibile specificare in modo dettagliato le proprietà acustiche di un segmento in una fase successiva a quella della trascrizione. Questo non è l'unico vantaggio di avere un segnale acustico allineato e annotato foneticamente. Infatti, in corpora di parlato che rispettano un principio gerarchico[16] nella costituzione delle unità linguistiche, l'allineamento tra segmento e segnale implica anche la possibilità di svolgere analisi della sostanza fonetica anche in relazione a unità superiori al segmento, sia di tipo prosodico, sia di tipo morfologico-lessicale e sintattico-pragmatico[17].

La presenza dell'allineamento temporale tra le nostre etichette e il segnale continuo sotteso dall'intervallo ci permette un ciclico processo di analisi tra i due livelli. Questo permette di orientare le nostre scelte di trascrizione verso delle categorie più generali (fonologiche) che possono essere poi in un secondo momento essere affinate o modificate sulla base dell'analisi acustica.

[16] Un po' artificioso, in quanto presuppone isomorfismo tra le unità e coincidenza dei confini di costituenti a vari livelli, ma non troppo lontano dalla realtà e soprattutto pragmaticamente molto utile sul piano metodologico nella costruzione di corpora di lingua parlata.

[17] Gli esempi in questa direzione sono molti e prevedono sia l'analisi della variabilità dei segmenti in relazione alle unità superiori come predittori, sia lo studio delle caratteristiche fonetiche dei costituenti.

4. *Conclusioni*

Nella prima parte di questo contributo sono stati portati argomenti e illustrati studi che pongono la variabilità fonetica al centro della riflessione fonologica. Le evidenze empiriche accumulate all'interno del paradigma di ricerca della *Laboratory Phonology* mostrano come le informazioni che provengono dalla dimensione acustica non vengano scartate dai parlanti nella rappresentazione delle categorie fonologiche e soprattutto svolgano un ruolo determinante nel processo di produzione e comprensione online della lingua parlata. L'acquisizione, in larga parte ancora da verificare, che emerge da quest'area di ricerca è che le categorie fonologiche mostrano al contempo di essere sia astratte e robuste, sia malleabili e sensibili a nuove esperienze.

Se le cose stanno così, o se si vuole scoprire se le cose stanno realmente così, è necessario creare strumenti e approcci metodologici che permettano ai ricercatori di collegare il piano acustico a quello delle categorie fonologiche. In altri termini per indagare il ruolo della variabilità bisogna disporre di dati linguistici multidimensionali e multimodali che integrino le unità discrete tradizionali con le dimensioni continue che le hanno generate.

Nella seconda parte del contributo vengono perciò proposte alcune riflessioni sulla natura del dato, sul relativo grado di definizione e sulle modalità di integrazione del piano continuo con quello discreto. L'allineamento temporale tra una trascrizione e il segnale acustico costituisce lo strumento metodologico che permette la connessione e lo scambio di informazioni tra due piani altrimenti separati. In questo modo è possibile connettere i segmenti sul piano fonologico e le informazioni acustiche sul piano continuo. Come discusso nei paragrafi 3.1 e 3.2, questa connessione apre nuove possibilità di indagine, ma rivela allo stesso tempo anche i rischi portati dall'enorme ricchezza informativa del dato acustico. La principale insidia è quella di determinare il discrimine tra ciò è misurabile e ciò che è rilevante sul piano linguistico, cognitivo e sociale. In estrema sintesi, non tutto ciò che è misurabile è anche rilevante per la comprensione del linguaggio.

La scoperta e la definizione dei confini di ciò che è rilevante rappresentano gli obiettivi di indagine offerti dall'integrazione di questi domini nel dato linguistico. Le potenzialità in questa direzione sono molte. Su un piano più strettamente fonologico, è possibile ottenere una visio-

ne più completa e una comprensione più profonda del processo di generalizzazione di categorie fonologiche via via più astratte attraverso l'organizzazione sistematica delle diverse fonti di variabilità che abbiamo illustrato, da quella idiosincratica del singolo parlante fino alle dimensioni di variazione condivise collettivamente. Su un piano linguistico più generale, secondo la prospettiva di *usage-based phonology* delineata da Bybee e trasposta nello studio dell'italiano da Voghera, è interessante comprendere meglio le funzioni svolte dalle informazioni contenute nella sostanza fonica nella processazione e rappresentazione delle unità linguistiche superiori, sia sul piano della struttura dinamica del lessico, sia su quello della segnalazione locale di contenuti sintattico-pragmatici.

Sul piano più metodologico questo si traduce alla fine in una raccomandazione (auspicio) molto semplice: costruire corpora di parlato spontaneo dialogico che includano un audio di buona qualità, con allineamento temporale affidabile e una trascrizione almeno di livello fonologico.

Riferimenti bibliografici

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Anttila, Arto. 2002. Variation and Phonological Theory. In Chambers, Jack & Trudgill, Peter & Schilling-Estes, Natalie (eds.), *Handbook of Language Variation and Change*, 206-243. Oxford: Wiley-Blackwell.

Bazzanella, Carla. 1994. *Le facce del parlare*. Scandicci: La Nuova Italia.

Berretta, Monica. 1994. Il parlato italiano contemporaneo. In Serianni, Luca & Trifone, Pietro (a cura di), *Storia della lingua italiana II. Scritto e parlato*, 267–270. Torino: Einaudi.

Berruto, Gaetano. 1985. Per una caratterizzazione del parlato: l'italiano parlato ha un'altra grammatica? In Holtus, Günther & Radtke, Edgar (Hrsg.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 120–153. Tübingen: Gunter Narr.

Berruto, Gaetano. 2012 (1987, 1ª ed.). *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.

Berruto, Gaetano. 2015. Intrecci delle dimensioni di variazione fra variabilità individuale e architettura della lingua. In Jeppesen, Kirsten Kragh & Lindschouw, Jan (éds.), *Les variations diasystématiques et leurs interdépendances dans les langues romanes*, 431–447. Strasbourg: Éditions de Linguistique et de Philologie.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cohn, Abigail C. & Fougeron, Cécile, & Huffman, Marie K. 2012. Introduction. In Cohn, Abigail C. & Fougeron, Cécile & Huffman, Marie K. (eds.), *The Oxford Handbook of Laboratory Phonology,* 3–9. Oxford: Oxford University Press.

Coseriu, Eugenio. 1969. *Einführung in die strukturelle Linguistik*. Tübingen: Universität Tübingen, Romanisches Seminar.

Docherty, Gerard J. & Foulkes, Paul. 2014. An Evaluation of Usage-Based Approaches to the Modelling of Sociophonetic Variability. *Lingua* 142. 42–56.

Ernestus, Mirjam. 2000. *Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface*. Utrecht: LOT.

Ernestus, Mirjam. 2012. Message-Related Variation: Segmental Within-Speaker Variation. In Cohn, Abigail C. & Fougeron, Cécile & Huffman, Marie K. (eds.), *The Oxford Handbook of Laboratory Phonology,* 93–102. Oxford: Oxford University Press.

Ernestus, Mirjam & Baayen, R. Harald. 2011. Corpora and Exemplars in Phonology. In Goldsmith, John & Riggle, Jason & Yu, Alan C. L. (eds.), *The Handbook of Phonological Theory*, 374–400. Oxford: Wiley.

Farnetani, Edda & Recasens, Daniel. 2010. Coarticulation and Connected Speech Processes. In Hardcastle, William J. & Laver, John & Gibbon, Fiona E. (eds.), *The Handbook of Phonetic Sciences*, *Second Edition*, 316–352. Oxford: Wiley.

Foulkes, Paul. 2010. Exploring Social-Indexical Knowledge: A Long Past but a Short History. *Laboratory Phonology* 1(1). 5–39.

Foulkes, Paul & Docherty, Gerard J. 2006. The Social Life of Phonetics and Phonology. *Journal of Phonetics* 34(4). 409–438.

Goldinger, Stephen D. 1996. Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5). 1166–1183.

Goldinger, Stephen D. 1998. Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review* 105(2). 251–279.

Hay, Jennifer & Nolan, Aaron & Drager, Katie. 2006. From Fush to Feesh: Exemplar Priming in Speech Perception. *The Linguistic Review* 23(3). 351–379.

Hawkins, Sarah. 2003. Roles and Representations of Systematic Fine Phonetic Detail in Speech Understanding. *Journal of Phonetics* 31(3). 373–405.

Hinskens, Frans & Hermans, Ben & van Oostendorp, Marc. 2014. Grammar or Lexicon. Or: Grammar and Lexicon? Rule-Based and Usage-Based Approaches to Phonological Variation. *Lingua* 142. 1–26.

Hoole, Philip & Kühnert, Barbara & Pouplier, Marianne. 2012. System Related Variation. In Cohn, Abigail C. & Fougeron, Cécile & Huffman, Marie K. (eds.), *The Oxford Handbook of Laboratory Phonology,* 115–130. Oxford: Oxford University Press.

Hoole, Philip & Pouplier, Marianne. 2015. Interarticulatory Coordination. In Redford, Melissa A. (ed.), *The Handbook of Speech Production*, 131–157. Oxford: Wiley-Blackwell.

Iskarous, Khalil. 2018. The encoding of vowel features in Mel-Frequency Cepstral Coefficients. In Vietti, Alessandro & Spreafico, Lorenzo & Mereu, Daniela & Galatà, Vincenzo (a cura di), *Il parlato nel contesto naturale*, 9–18. Milano: Studi AISV.

Johnson, Keith. 1997. Speech Perception without Speaker Normalization: An Exemplar Model. In Johnson, Keith & Mullennix, John W. (eds.), *Talker Variability in Speech Processing*. 145–165. San Diego: Academic Press.

Kaland, Constantijn & Galatà, Vincenzo & Spreafico, Lorenzo & Vietti, Alessandro. 2019. Which Language R You Speaking? /r/ as a Language Marker in Tyrolean and Italian Bilinguals. *Language and Speech* 62(1). 137–163.

Labov, William. 1963. The Social Motivation of a Sound Change. *WORD* 19(3). 273–309.

Labov, William. 2006. A Sociolinguistic Perspective on Sociophonetic Research. *Journal of Phonetics* 34(4). 500–515.

Levinson, Stephen C. & Holler, Judith. 2014. The Origin of Human Multi-Modal Communication. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*. 369(1651). 20130302.

Lindblom, Björn. 1990. Explaining Phonetic Variation: A Sketch of the H&H Theory. In Hardcastle, William J. & Marchal, Alain (eds.), *Speech Production and Speech Modelling*, 403–439. Dordrecht: Springer.

Mereu, Daniela. 2017. Arretramento di /s/ nel sardo cagliaritano: uno studio sociofonetico. In Bertini, Chiara & Celata, Chiara & Lenoci, Giovanna & Meluzzi, Chiara & Ricci, Irene (a cura di), *Fattori sociali e biologici nella variazione fonetica*. *Social and Biological Factors in Speech Variation*, 45–65. Milano: Officinaventuno.

Mereu, Daniela & Vietti, Alessandro. 2021. Dialogic ItAlian: the creation of a corpus of Italian spontaneous speech. *Speech Communication.* 130. https://doi.org/10.1016/j.specom.2021.03.002

Moore, Brian C.J. 2010. Aspects of Auditory Processing Related to Speech Perception. In Hardcastle, William J. & Laver, John & Gibbon, Fiona E. (eds.), *The Handbook of Phonetic Sciences*, *Second Edition*, 454–488. Oxford, Wiley.

Palmeri, Thomas J. & Goldinger, Stephen D. & Pisoni, David B. 1993. Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(2). 309–28.

Pierrehumbert, Janet. 2001. Exemplar Dynamics: Word Frequency, Lenition and Contrast. In Bybee, Joan & Hopper, Paul (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam: Benjamins.

Pierrehumbert, Janet. 2012. The Dynamic Lexicon. In Cohn, Abigail C. & Fougeron, Cécile & Huffman, Marie K. (eds.), *The Oxford Handbook of Laboratory Phonology,* 173–183. Oxford: Oxford University Press.

Pierrehumbert, Janet & Beckman, Mary & Ladd, D. Robert. 2000. Conceptual Foundations of Phonology as a Laboratory Science. In Burton-Roberts, Noel & Carr, Philip & Docherty, Gerard (eds.), *Phonological Knowledge*, 273–303. Oxford, Oxford University Press.

Recasens, Daniel. 2018. Coarticulation. In *Oxford Research Encyclopedia of Linguistics*, https:&&doi.org&10.1093&acrefore&9780199384655.013.416. Oxford: Oxford University Press.

Sornicola, Rosanna. 1981. *Sul parlato*. Bologna: Il Mulino.

Stevens, Kenneth N. & Hanson, Helen M. 2010. Articulatory-Acoustic Relations as the Basis of Distinctive Contrasts. In Hardcastle, William J. & Laver, John & Gibbon, Fiona E. (eds.), *The Handbook of Phonetic Sciences*, *Second Edition*, 424–453. Oxford: Wiley.

Todd, Simon & Pierrehumbert, Janet B. & Hay, Jennifer. 2019. Word Frequency Effects in Sound Change as a Consequence of Perceptual Asymmetries: An Exemplar-Based Model. *Cognition* 185. 1–20.

Vietti, Alessandro. 2014. Alcune riflessioni sulla teoria degli esemplari e la variazione linguistica. *Rivista italiana di dialettologia* 36. 7–22.

Vietti, Alessandro. 2019. Phonological Variation and Change in Italian. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. 10.1093&acrefore&9780199384655.013.494.

Vietti, Alessandro & Spreafico, Lorenzo. 2008. Phonetic Variation of /r/ in a Language Contact Context: The Case of South Tyrol Italian. In *Laboratory Phonology 11*, 145–146. Wellington: Victoria University of Wellington.

Vietti, Alessandro & Spreafico, Lorenzo. 2016. Lo strano caso di /r/ a Bolzano: problemi di interfaccia. In Iacobini, Claudio & Voghera, Miriam & Savy, Renata (a cura di), *Livelli di analisi e fenomeni di interfaccia*, 263–281. Roma, Bulzoni.

Vihman, Marilyn M. 2014. *Phonological Development: The First Two Years*. Oxford: Wiley.

Voghera, Miriam. 1992. *Sintassi e intonazione nell'italiano parlato*. Bologna: Il Mulino.

Voghera, Miriam. 2017. *Dal parlato alla grammatica: costruzione e forma dei testi spontanei*. Roma: Carocci.

Whalen, Douglas H. 2019. The Motor Theory of Speech Perception. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. https:&&doi.org&10.1093&acrefore&9780199384655.013.404.

Cinzia Avesani
(Consiglio Nazionale delle Ricerche)

Barbara Gili Fivela
(Università del Salento)

Analysing Prosody: Methods, issues, and hints on crosslinguistic comparison and L2 learning[1]

1. *Introduction*

During an act of speaking, the flow of speech is not a simple concatenation of segments, but consonants and vowels are modulated by principled variations of fundamental frequency (F0), duration, intensity. These acoustic modulations are perceived as variations in pitch, length and loudness of speech stretches, but they affect single sound segments to various degrees. They are the acoustic and perceptual reflexes of how the sounds are articulated: segments that have higher fundamental frequency are produced with a higher rate of vocal fold vibration, determined by the configuration of the larynx, the subglottal pressure, and the degree of oral closure; segments that have longer duration are produced with speech gestures that are longer (and have phases which are not truncated, e.g., Byrd & Saltzman 2003); segments that have higher intensity are produced with more articulatory effort and higher subglottal pressure.

These are the parameters that, besides affecting each segment, give rise to a set of phonological phenomena such as stress, rhythm and timing, tone and intonation, usually referred to with the cover term of Prosody.

A broad definition of prosody refers to those non-segmental speech events that participate in the organization of lexicon and syntax and play a decisive role in the semantic and pragmatic interpretation of a given utterance. Non-segmental is here preferred to *suprasegmentals*, a term originally coined by Lehiste (1970) and used – often in the past, but still sometime used nowadays – interchangeably with prosody. With it, she

[1] This work has been designed, discussed and conducted in close collaboration between the two authors. Main responsibility in writing the paper is divided as follows: Avesani: §1, 2, 2.1, 2.3, 3, 4, 5; Gili Fivela: §2.2, 6, 6.1, 7, 7.1; Avesani and Gili Fivela: 6.2, 8.

intended to indicate a set of linguistic phenomena that span over domains larger than a segment, like syllables, phrases and utterances. But the term also evokes that prosodic events stand "above" the segments. Using "suprasegmentals" with such denotation instead of prosody can be misleading, as it overlooks one fundamental aspect of speech: there are no utterances of natural languages in which segments are unaffected by prosody, and because prosody is an intrinsic and unavoidable part of any language, investigating speech events without reference to it misses important aspects of how speech is organized. A conspicuous number of experimental studies has shown that prosody affects all aspects of the speech signal. For example, not only elements found in prosodically prominent positions (i.e. stressed and accented) are longer, higher, louder, and more fully articulated than elements in prosodically weak positions (i.e. unstressed and unaccented), but also elements occurring at the edges of prosodic units are affected in their inner articulation: compared to consonants occurring in unit-internal position, consonants occurring at initial edges are strengthened (Fougeron & Keating 1997; Cho & Keating 2001; Keating *et al*. 2004) and segments and syllables occurring at final edges are regularly lengthened (e.g. Beckman & Edwards 1990; Edwards *et al*. 1991).

A better definition of prosody is due to Beckman (1996) who refers to prosody as the "organizational structure of speech". As a musical score is organized in notes, measures, musical sentences and so on, prosody organizes speech in prosodic constituents, dividing the flow of speech in "chunks of information" that help listeners to parse discourse in meaningful units for further linguistic information (syntactic, semantic, conversational: i.e. turns). In line with other proposals regarding prosody (e.g., rule-based and syntactically related as in Nespor & Vogel 1986; Selkirk 1984), the definition suggests that prosodic units stand in a hierarchical relation, on par with the hierarchy of syntactic constituents that determine the order of morphemes inside words and of words inside sentences. Independently of the specific definition adopted, the subdivision of speech in prosodic units and the organization of such units in a structure is the first task of prosody, referred to as *phrasing*. A second main function of prosody is to mark prominence relations within each prosodic constituent with a language-specific variable combination of acoustic parameters: the abovementioned duration, intensity, F0. This is the second task of prosody: *highlighting* in a principled way some elements within each prosodic unit making them to stand out as prominent

at lexical and postlexical level, and marking the prominence relations that hold among them.

One of the most investigated prosodic phenomenon is intonation, which has been defined in the literature at least in a narrow and in a broad sense. One definition equates intonation with the melody of speech, restricting the use of the term to refer to the variations of pitch in the course of the utterance (t'Hart *et al*. 1990), which are not determined by lexical distinctions as in tone languages (a.o., Gussenhoven 2007) and are used by the speakers "to mark pragmatic force of the information in an utterance" (Pierrehumbert 1999). More generally, intonation can be defined as the "linguistically structured and pragmatically meaningful" modulation of pitch (Arvaniti 2012: 265). In the broader use, which refers to the linguistic structure and pragmatic meaning, the term intonation includes also variations in loudness, length and segmental quality besides pitch, basically equating intonation with prosody. However, even definitions focussing on pitch modulation and its link to linguistic information do not assume that pitch is the only correlate of intonation. Rather, pitch is considered as the main correlate, directly linked with phonological representations in the grammar, while variation in loudness, length and segmental quality co-occur with its modulation.

In spoken language, intonation serves a variety of linguistic and paralinguistic functions, ranging from speech act information (assertions, questions, commands etc.), information structure or information packaging (topic, focus, background), information status (given vs. new information), knowledge state (or epistemic position of the speaker with respect to the information exchange), illocutionary force, affective state, emotions. Since the course of fundamental frequency is the main exponent of intonation through which the speakers convey such an array of communicative functions, it is clear that determining the structure of pitch modulations and unravelling the nature of intonational meaning is a challenging task. The main difficulties reside, first, in defining the primitives that make up the pitch contour, since linguistically related F0 changes are not as easily identifiable as in tone languages; and, second, in determining the meaning associated with those primitives, as in the intonational domain meaning is represented by pragmatic and information structure contrasts, which are notoriously more difficult to determine than stable lexical contrasts in the segmental domain. Therefore, both intonational form and its connection with segmental material and intonational meaning are hard to pinpoint.

One solution adopted in the literature is to treat intonation contours as gestalts, an approach that starting from Bolinger (1951) has been embraced by many scholars until recently (e.g., Cooper & Sorensen 1981; Hirst & Di Cristo 1998; Xu 2005; Grabe *et al*. 2003) and, at least to a certain extent, applied also to Italian in the *Language into Act* model proposed by Cresti and collaborators (e.g. Cresti & Moneglia 2018). In those studies, intonational contours are deemed to be holistic entities that directly reflect certain structural or functional aspect of speech, such as the depth of a syntactic boundary (Cooper & Sorensen 1981), or a speech act (e.g. Cresti 2005; Moneglia 2006; Cresti & Moneglia 2018).

This approach faces important problems, though. The contours in Fig. 1 help illustrating the first point. The figure represents the F0 contours of two utterances differing in length: *È Marina?* 'Is she Marina?' on the left and *È Marina la ragazza che ti piace?* 'Is it Marina the girl that you like?' on the right. Globally, both contours share what at a first view could be considered the "same" rise-fall-rise F0 pattern. Although they show some similarities, they cannot be said to be identical: after the initial rise-

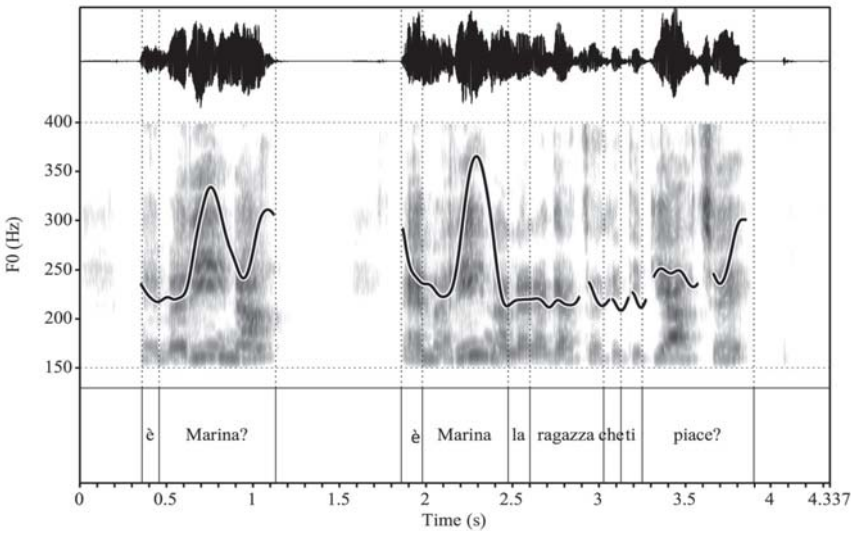


Fig. 1. Intonational contour of the sentences È Marina? *'Is she Marina?' and* È Marina la ragazza che ti piace? *'Is it Marina the girl that you like?'*

fall movement common to both contours, a low plateau follows in the contour of the longer utterance on the right side of the figure. Notwithstanding the difference in their form, both contours are used by the speakers for producing *wh*-questions and are perceived as conveying the same meaning. The same difference in rise-fall-rise contours found in utterances of different length sharing the same communicative meaning are also found in other languages (for example in English, Greek and Polish). As in Italian, in none of those cases it is possible to consider one contour as a "stretched" or "squeezed" version of the other as it would be expected if melodies were undivided wholes (Arvaniti & Ladd 2009; Arvaniti 2012).

This example shows that the shape of intonational contours that share the same pragmatic meaning (asking a question) can vary considerably as a function of the segmental material with which they are coproduced. However, we can make sense of this variation if we factor out the components of the rise-fall-rise pattern and take into consideration the overall prosodic structure of the utterance. That is, if we do not consider the difference of the two contours as the by-product of random variation, but if we take into consideration the main and higher level functions of prosody: highlighting and phrasing. In both utterances the highest prominence is produced on the stressed syllable of word *Marina* 'Marina', which carries in both cases the same rise-fall pattern, and the final rise of the pattern is synchronized with the end of the contour in both cases. However, while in the short utterance the final rise appears at the end of the word that carries the main prominence, and makes the word Marina the docking site of a combined rise-fall-rise pattern, in the longer utterance it appears as a separate pitch event due to the number of segments that separate the last accented syllable from the end of the utterance. Thus, it appears that parts of the melody coordinate independently with parts of the segmental string (Arvaniti 2007; Ladd 2008: chapter 2).

The example illustrates what has been shown in the literature with plenty of evidence, i.e. that when tunes are realized in utterances with different length and metrical structure their form differs substantially. The idea that pitch contours are not non-analyzable gestalts but have an internal structure has been acknowledged also by a configurational approach such as the IPO (Institute for Perception Research, Eindhoven) model of intonation (t'Hart *et al*. 1990). In the IPO model the largest

descriptive unit of intonation is the pitch contour, but contours are decomposed in configurations (Prefix, Root, Suffix) that consist of pitch movements. The pitch movements are specified in terms of features (their direction, timing with regard to syllable boundaries, rate of change, size) and are distinguished based on their function to lend prominence or not to the syllable on which they occur (prominence-lending movements that co-occur with stressed syllables vs. non-prominence lending movements). Importantly, t'Hart and colleagues observed that the same sequence of pitch movements can be distributed differently over an utterance (t'Hart et al. 1990: 98): they could appear either all together on a single syllable or as separated by intervening syllables without affecting the perceptual identity of the contour. By using stylisation techniques, the authors showed that only certain aspects of the contour are important for the listeners while the global shape of the contour is not. Overall, their work showed what will be repeatedly demonstrated in the following years, that parts of the melody appear to coordinate independently with parts of the segmental string.

Beside facing a problem in accounting for intonational form, a holistic approach faces the major difficulty of maintaining a one-to-one relationship between form and meaning. Since the pragmatic functions performed by intonation are manifold, postulating a direct form-function relationship necessarily leads to identifying a specific form for each different meaning across different utterances, even where the diversity between one melody and the other is not justifiable on acoustic and perceptual ground. Typically, different pragmatic functions empirically defined are associated with putatively different types of contours, but their belonging to contrasting categories or to variants of the same category is not always proved on experimental (acoustically, perceptually) ground (e.g. Cresti & Moneglia 2018).

On the contrary, in the past decades many authors have noticed that the mapping between form and function is not a one-to-one, but rather a many-to-many relation: the same melody can be used to convey different meanings, and the same meaning can be expressed by different melodies (a.o., Pike 1945; Lehiste 1970; Ladd 2008; Grice *et al*. 2005; for Italian, Gili Fivela 2008; Gili Fivela *et al*. 2015). Problems like these make it clear that viewing melodies as composed by smaller, phonologically relevant elements is more likely to be successful in accounting for intonational meaning, especially if it is assumed that the mapping between the

superficial form of pitch modulation and linguistic functions/meaning is not direct, but it is mediated by the phonological structure.

2. *Autosegmental-Metrical theory of intonation*

A major breakthrough in our understanding of intonation was achieved with the advent of the *Autosegmental-Metrical theory of intonation* (henceforth, AM). The theory has its origin in Pierrehumbert's dissertation (1980) who incorporates in it the insights of two previous influential theses, Liberman (1975) and Bruce (1977), and it has been further developed into the current model particularly by Beckman and Pierrehumbert (Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988). The basic tenet of the theory is that intonation is part of the grammar and has a phonological structure: it is possible to characterize contours in terms of a string of categorically distinct elements and to provide a mapping from phonological elements to continuous phonetic parameters (Ladd 2008: 43). The term "Autosegmental-Metrical" was coined by Ladd (1996) as it reflects the intellectual heritage and the principles of Intonational Phonology (Bruce 1977; Pierrehumbert 1980; Gussenhoven 1984; Liberman & Pierrehumbert 1984; Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988) and those of Metrical and Prosodic Phonology, with reference to the domains proposed within Prosodic Phonology and the prominence relations holding within them (Liberman 1975; Liberman & Prince 1977; Selkirk 1984, 2004; Nespor & Vogel 1986, 2007).

2.1 Basic elements

According to AM, intonation is represented in terms of a string of static H(igh) and L(ow) tones. H and L tones are the primitives of the abstract phonological representation and are phonetically realized as targets in the F0 contour, typically peaks and dips in the contour. The contrast between H and L tones is paradigmatic, i.e. *ceteris paribus*, a H tone is higher than a L tone in the same context, but the phonetic height of each target is defined in relative terms, with reference to the speaker's range: a L tone is realized as a low target on the hypothetical bottom line of the speaker's range and a H tone as a high target on the topline (Pierrehumbert 1980: 69 and following). Given that the speaker's range shows a natural

downtrend across the utterance and the top and bottom lines tend to converge toward the end of it, a H tone that occurs late in the F0 contour could be as high in F0 as a L tone at the beginning of the contour.

Crucially, tones are represented on an autonomous tier or plane separated from the linguistic material with which they are necessarily co-produced: in line with Autosegmental Phonology, they are auto-segments, connected with units in the skeleton through specific association principles (Leben 1973; Goldsmith 1979)[2]. It is important to notice that the string of tones that represents a melody (or tune) is not intended to describe the whole F0 contour, but to represent only those parts of the melody that are linguistically significant: it is intended not as a mere transcription that describes all the peaks, troughs and turning points in a contour, but rather as an underspecified phonological representation. A direct consequence is that the same tune can be associated with texts of different segmental length and composition, giving rise to intonational contours that can be holistically different in acoustic form, but are the phonetic realization of the same abstract melodic entity.

Tones associate with the string of segments (or *text*) indirectly, through the mediation of the metrical structure of the utterance. With *metrical structure* we refer to a theoretical proposal which considers a given string of language to be organized into a series of hierarchically arranged prosodic constituents (in line with Prosodic Phonology), and that the linguistic units included in those constituents are specified in terms of relative prominence relations (in line with Metrical Phonology; Selkirk 1984, 2004; Nespor & Vogel 1986, 2007; Pierrehumbert & Beckman 1988; Liberman 1975; Libermann & Prince 1977).

In the literature, different theoretical proposals have been made as for the number of prosodic constituents that compose the hierarchy of prosodic domains. Nespor & Vogel (1986) for example propose seven constituents, which are, from the smallest to the larger: *Syllable*, *Foot*, *Prosodic Word*, *Clitic Group*, *Phonological Phrase*, *Intonational Phrase*, and *Utterance*. Others (e.g. Selkirk 1978, 1986), do not posit the existence of a *Clitic Group* but propose a *Minor phrase* and a *Major phrase* between the level

[2] In an autosegmental representation, different characteristics of a sound message - for example tones and phonemes - are represented on different tiers that all converge on a common plane, called a skeleton; this consists of a sequence of temporal units designed to fix the linear order of consonants and vowels (Leben 1973; Goldsmith 1979).

of the *Prosodic Word* and the *Intonation Phrase*. This is the highest level of the hierarchy, i.e., Selkirk does not always posit the *Utterance* as prosodic domain (but see Selkirk 1978), while she does consider the *Mora* as the lowest unit of the hierarchy. Beckman & Pierrehumbert (1988) consider three levels of constituents above the prosodic word: the *Accentual Phrase*, the *Intermediate Phrase* (ip), roughly corresponding to the phonological phrase of Nespor & Vogel (1986) and to the *Major Phrase* of Selkirk (1978, 1986), and the *Intonational Phrase* (IP). As it appears, there is agreement on the higher levels of the hierarchy but not as on the mid-levels (for a discussion of the different prosodic hierarchies and their correspondences: Shattuck-Huffnagel & Turk 1996; Frota 2017).

The constituents proposed in different models are motivated by the theoretical and empirical analyses of specific languages: for example, the proposal of an *Accentual Phrase* as a prosodic domain in Beckman & Pierrehumbert (1986) stems from their prosodic analysis of Japanese and has been shown as pertinent in the analysis of other languages as well, such as French (Verluyten 1982; Jun & Fougeron 1995). So far, most works adopting the AM framework to analyse the intonation of many languages (for an overview, Jun 2005; Frota & Prieto 2015) show that at least two levels of constituents are pertinent for intonation: a minor phrase, be it the *Intermediate*, *Phonological* or *Minor phrase*, and a major one, the *Intonational Phrase*.

The authors of the above proposals agree that all the constituents of a certain hierarchical level are exhaustively included in the constituents of the upper hierarchical level, a constrain known as the Strict Layer Hypothesis; in other words, that the prosodic structure is not recursive, differently from the syntactic structure[3], and that each constituent is endowed with a *head*, a metrical strong element.

Tones can associate with the constituents' heads or with constituents' edges. In the first case, in a language such as Italian, the Tone-Bearing Unit (TBU, the docking site of the tone) is the stressed syllable, which by virtue of this association gets its prominence enhanced. Tones associated with constituent heads are called *pitch accents* (PA) and are marked with a star, e.g. H* or L*. The last pitch accent in the Intermediate Phrase is the *nuclear*

[3] The debate on the recursive vs. non-recursive nature of prosodic constituents is very lively in works developing the original Prosodic Phonology proposal or referring to the prosodic hierarchy in general. For instance, Ladd (2008) opened to a limited recursivity in prosodic domains allowing for compound constituents in a given level of the hierarchy.

accent (Beckman 1996). Tones can also associate with the right boundary of the Intermediate and Intonational phrase: the former is marked by a hyphen and the second by a percent sign, e.g., respectively H- and H%, and are called *phrase accent* and *boundary tone.* Collectively, they are referred to as *edge tones* and their role is to demarcate a phrasal boundary. All languages investigated so far have H and L tones that associate with right boundaries. For several languages it has been postulated also a left edge association for the Intonational Phrase, mostly of a H tone, that in such a case is indicated as %H (e.g. for English: Beckman *et al*. 2005; for Italian: Avesani 1995; Grice *et al*. 2005; Gili Fivela *et al*. 2015). Edge tones associate with the edges of constituents and are phonetically realized on the segments flanking their edges, such as the final vowels or sonorant consonants for H% or L% and phrase initial ones for %H – see Figure 2.

Pitch accents (henceforth, PAs) can be monotonal, i.e. composed by one tone only, as in H* or L*, or bitonal, as L*+H or H*+L, whose phonetic realization gives rise to glissandos, namely rises and falls. The star notation reflects the fact that the starred tone is stronger and is directly associated with the TBU. The weaker tones are called *leading* if they precede the starred tone, *trailing* if they follow it. For a discussion on the nature of the starred tone and for the internal structure of pitch accents the reader is referred to

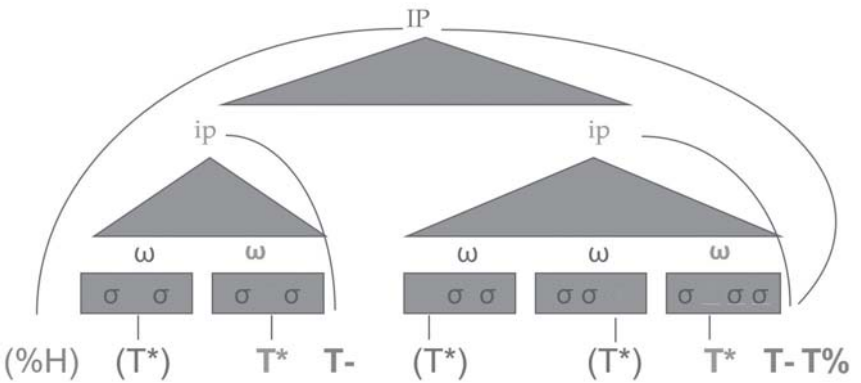


Fig. 2. Example of prosodic tree, including an Intonational Phrase (IP), Intermediate Phrases (ip), prosodic words (ω) and syllables (σ); pitch accents (T, representing both monotonal and bitonal pitch accents) are shown both in prenuclear (bracketed) and nuclear position and are followed by edge tones, either phrase accents (T-) or boundary tones (T%)*

Arvaniti *et al*. (2000) and Grice (1995). Edge tones are monotonal, but in the literature multitonal combination of edge tones have been occasionally proposed in the analysis of specific languages, for example a tritonal LHL% boundary tone has been used for the analysis of Catalan (Prieto 2014).

Since the beginning of the AM approach, a crucial role in the definition of tonal categories and in the coding of the intonative oppositions within the intonational system of a specific language has been recognized to tonal *alignment* and *scaling*. The former corresponds to the synchronization of the F0 peaks and lows to the TBU, and the latter represents the tone height of the H or L tone associated with a structural position (see also § 2.1).

Based on the formal properties of alignment and scaling of PAs and edge tones and on the informational and pragmatic functions they convey, a limited set of contrastive tonal events can be identified as the building blocks of the intonational system of a specific language. Tunes then arise from the phonetic implementation of a linear sequence of pitch accents and edge tones whose targets are assumed to be linearly interpolated. Typically, the melody of an intermediate phrase (ip) is composed by one or more optional prenuclear PAs, one obligatory nuclear PA, and one edge tone. If two ips combine to make up an intonational phrase (IP), then its melodic structure is represented as follows:

(1) $[[(\text{prenuclear PA}) - \text{nuclear PA}]_{ip}\ [(\text{prenuclear PA}) - \text{nuclear PA}]_{ip}]_{IP}$

where at the right boundary of each ip is associated a phrase accent and at the right boundary of the IP is associated a boundary tone. Consequently, the end of the melody is marked by a combination of two edge tones – see Figure 2.

For Mainstream American English (MAE), a consensus analysis has identified two boundary tones and two phrase accents (L%, H%, L-, H-), two monotonal PAs (L*, H*) two bitonal rising PAs, L+H* and L*+H, and one falling PA H*+!H (Beckman *et al*. 2005).

As the inventory of contrastive tonal events is rigorously linguo-specific, not all languages or language varieties share the same tonal inventory. For example, in MAE only one falling PA is deemed to be part of the intonational inventory, while in all the varieties of Italian analyzed so far a H*+!H is not attested, but all of them share a falling H+L* that occurs as the nuclear pitch accent in broad focus declarative sentences. For some of them also a H*+L falling accent is attested (Gili Fivela *et al*. 2015).

Summarizing, in the Autosegmental Metrical framework the two basic tasks of prosody, phrasing and highlighting, are fulfilled respectively by the *placement* of edge tones at the boundaries of prosodic constituents and by the placement of pitch accents within each constituent. The pragmatic meaning of the tune is determined by the linear position and by the selection of the *type* of PAs and edge tones (Pierrehumbert & Hirschberg 1990).

Going back to the example in §1, the similarities and the differences of the two contours represented in Figure 1 can be disentangled by considering how PAs and edge tones are distributed in the contours. In the short utterance shown in Figure 3, the stressed syllable of *Marina* carries a L+H* PA, where H* is associated with the TBU /ri/ and L is the trailing tone. This PA is the only nuclear PA occurring in the contour, as the short utterance is phrased in one intonational phrase only. The PA is then followed by a L- phrase accent and a H% boundary tone realized on the following unstressed syllable[4].

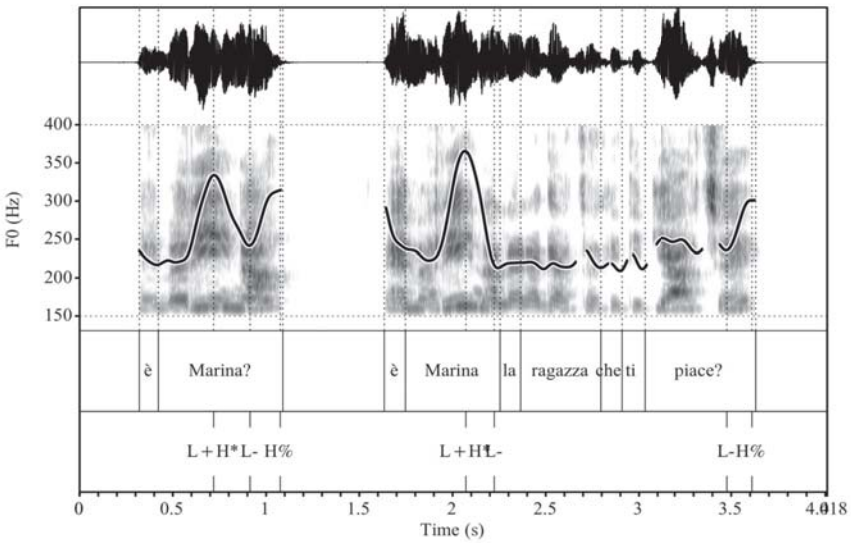


Fig. 3. Intonational contour and tonal transcription of the sentences È Marina? *'is she Marina?' and* È Marina la ragazza che ti piace? *' is it Marina the girl that you like?'*

[4] By default in the first formulations of the AM theory an IP is composed by one ip. Therefore, even a monorematic utterance is right marked by two edge tones.

The long utterance *È Marina la ragazza che ti piace?* 'Is it Marina the girl that you like?' is phrased in two intermediate phrases:

(2) $[[\text{È Marina}]_{ip}\ [\text{la ragazza che ti piace}]_{ip}]_{IP}$

Here *Marina* shares the same PA (L+H*) and the same metrical position (nuclear) of the short utterance and it is followed by the same L- phrase accent that delimits the right boundary of the first intermediate phrase in which it is wrapped. A L- phrase accent delimits the right boundary of the second intermediate phrase and a H% boundary tone marks the right boundary of the intonational phrase[5].

2.2 Phonological representation and phonetic form

As already mentioned, the phonological representation is necessarily language-specific. Notably, such phonological representation is mapped into a phonetic representation through phonetic realization rules, which are again language-specific, and which shape the F0 track in terms of its alignment and scaling characteristics.

The literature on alignment and scaling properties of tones is rich and it offers various points of view on the mapping between phonology and phonetics. For instance, alignment was originally interpreted as somehow defined by phonological association (the association of a tone to a TBU implied its alignment with it in Pierrehumbert's 1980 proposal), while later works did not assume such a strict coordination. The starred tone can also be aligned earlier or later than the tone-bearing unit, pointing to the impossibility to rely on alignment characteristics to identify the starred, associated tone (Arvaniti *et al*. 2000) or a coordination with specific landmarks may be assumed (cfr. Ladd *et al*.'s 1999 Segmental Anchoring Hypothesis). Further, such phonetic features, which were originally discussed with reference to acoustics, have later been investigated as far as their perception is concerned (D'Imperio & House 1997; D'Imperio 2000), and some works also focused on kinematic data (e.g., D'Imperio 2002). However, more recently, some authors argued that alignment and scaling of tonal targets may not be the only properties speakers take care of, as inter-subject differences may be observed in the accuracy speakers

[5] A postfocal compressed L+H* PA occurs on the TBU of the last word *piace* ('you like').

show in aligning targets or preserving pattern shape (Niebuhr 2007; D'Imperio *et al*. 2010).

Further, many works in the literature have basically shown that intra-linguistic alignment differences, e.g., between PAs which are part of a specific linguistic system, convey semantic and pragmatic differences that are (almost) categorically perceived (though results are not consistent in this respect; see Gili Fivela 2012 for an overview). Scaling has long been considered less relevant than alignment in linguistically differentiating linguistic functions, though it is nowadays accepted that scaling too conveys semantic and pragmatic differences and may be categorically perceived (e.g., yes-no questions as opposed to *wh*-questions in Majorcan Catalan: Vanrell 2006, 2007).

One of the key aspects of the AM proposal regards the lack of a "transparent" and direct mapping of phonetic properties onto a phonological representation. Such representation is abstract. However, its units are labelled taking somehow the phonetic properties into account (e.g. a label will include a H+L tone if, in the clearest realization, it is falling). Thus, it is well known that identifying high and low turning points in the phonetic form is not enough to identify phonological targets and to label them. Crucially, a phonetic event may be considered as phonological if its presence/absence implies changes in the linguistic function played by the pattern. Thus, the questions to be answered in analyzing a phonetic continuum are, for instance: Is there a linguistic function played by such F0 event? What is the impact of changing its alignment and scaling? Does the meaning of the utterance change?

Depending on the linguistic system, the inventory of intonational units and their alignment and scaling characteristics changes. Thus, differences are found in systems of different languages and even in the case of varieties of the same language. For instance, a set of nine pitch accents and six edge tones is necessary to analyze 13 Italian varieties (Gili Fivela *et al*. 2015), and each variety shows a specific selection and a specific combination of those units. For instance, the L*+H pitch accent is found in Neapolitan and Turin Italian, but it is not found in Pisa Italian; a HL% edge tone is found in Pisa Italian after a H+L* pitch accent, while it is found after L+H* in Neapolitan (for updates on Italian in this line of research, see Gili Fivela & Iraci 2017; Gili Fivela & Nicora 2018; see the latter, together with Gili Fivela *et al.* submitted, for investigations concerning possible cross-varietal similarieties due to contact situations).

2.3 ToBI

The principles of AM theory are reflected in the ToBI (Tone and Break Index) transcription system (Beckman & Ayers 1997), a common transcription system whose immediate benefit is the possibility to compare the prosody of disparate languages and language varieties. Since 1991, date of the first workshop organised to define a set of common principles for transcribing Mainstream American English, ToBI-like analyses have been proposed for a number of different linguistic systems providing the intonational analysis of 35 languages and almost 30 language varieties (we refer to the following collective volumes: Jun 2005, 2014; Frota & Prieto 2015). In fact, the original ToBI system has been adapted for the description of languages which vary geographically (European, Native American, Asian, Australian aboriginal languages) and typologically, in the type and in the degree of lexical specification of prosody (intonational languages such as English, Italian, French, Spanish, Portuguese; lexical pitch accent languages such as Swedish, some Dutch and German dialects, Chickasaw, Japanese; and tone languages such as Cantonese and Mandarin).

What is important to highlight is that ToBI is not comparable to an International Phonetic Alphabet for intonation, with the choice of adopting a broader or narrower transcription, but it is a phonological representation of intonational contrasts[6]. The aim is not a more or less faithful depiction of the F0 contour; rather, its aim is to define a limited set of categories to represent the intonational contrasts in a sound system. Therefore, the transcription should be driven by system internal considerations, by considering a phonetic detail as part of the representations only if there is evidence it is contrastive (Arvaniti 2016: 8). For example, the decision to transcribe a pitch rise on a stressed syllable as a H* or a L+H* PA must be guided by considering the contrastiveness of such event within the whole system under analysis. The transcriber could then decide to include the rise as part of the phonological specification in one language, with the consequent adoption of the label L+H* because L+H* contrasts with H*,

[6] For a thorough discussion of the phonological assumptions behind current approaches to prosodic transcription, for the choice of discrete units and their granularity and the consequences of considering ToBI as a broad phonetic transcription we refer to the special collection on Advancing Prosodic Transcription that appeared in *Laboratory Phonology* (D'Imperio *et al*. 2016).

but to exclude it from the phonological representation of another language, where such a contrast is not attested (e.g. Arvaniti *et al*. 2000) Importantly, a decision about contrastiveness must be guided by form in combination with meaning: differences in the form of a tonal event should be considered contrastive only after taking into account focus, information structure, and the pragmatic function of utterances in discourse (cf. Pierrehumbert 1980: 59-63).

As a good practice suggested by Arvaniti (2016: 8), the analysis should involve several iterations that lead to both bottom-up and top-down decisions: a first set of data determines the original analysis, which is then used to annotate more data.

3. *On metrical structure*

As mentioned in section 2, Intonational Phonology also refers to the existence of prosodic domains and, consistently, the AM framework refers to the principles of Intonational Phonology as well as Metrical and Prosodic Phonology. As a matter of fact, different proposals regarding prosodic domains emerge from independent research traditions and they differ according to theories of the syntax-phonology mapping and theories of the structural relations between constituents of prosodic structure (Frota & Vigario 2018).

In one framework, the structure of phonological representation at the word level and above is a hierarchy of phonological constituents that results from the interaction of a limited set of (morpho)syntactic information with phonological principles related, among others, to constituent size and weight. The prosodic word (PW), the phonological phrase (PhP), and the intonational phrase (IP) are the domains of application of segmental rules and bear a relation to a specific syntactic constituent type: respectively a word-like (lexical) morphosyntactic unit, a phrase-like syntactic unit, and a clause-like syntactic unit[7] (Nespor & Vogel 1986; Selkirk 2011; Truckenbrodt 1995; for a thorough , on syntax-

[7] Two main branches of such framework correspond to relation-based and end-based approaches, which differ as for the syntactic information used in the computation of prosody: the former makes reference to notions like head-complement, modifier-head relations, and syntactic branching, while the latter refers to syntactic heads and maximal projections.

prosody interface we refer to Frota & Vigario 2018, and for an analysis of syntax-prosody interface in Italian we refer to Bocci 2013).

Parallel to theories positing that prosodic structure is rule-based and related to syntax, a different approach posits that prosodic structure is intonation and prominence defined, by relying on intonational, durational and segmental phenomena that characterize the constituents above the word level (Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988; Beckman 1996).

There is ample empirical evidence that a constellation of cues mark prosodic domains, functional to supporting the prosodic structure and the constituents it comprises. Constituents are marked by lengthened duration of the segments right-flanking the boundary (*final lengthening*), with degrees of lengthening that correlate with the prosodic boundary level (a.o. Beckman & Edwards 1994; Wightman *et al*. 1992; Byrd & Saltzman 2003); and by lengthening of earlier segments within the preboundary word (Price *et al*. 1991; Wightman *et al*. 1992; Turk & Shattuck-Hufnagel 2007). Segments in initial position of a prosodic domain have their articulatory properties enhanced as a function of the constituent level in the prosodic hierarchy (ip or IP) in which they appear (*prosodic strenghthening* of e.g. linguo-palatal contact or nasal flow e.g., Keating *et al*. 2004); pre-boundary as well as post-boundary segments can be glottalized (Pierrehumbert & Talkin 1992; Byrd & Saltzman 2003). Tonal marking of prosodic domains include the presence of edge tones, the scaling of subsequent H peaks within a domain and the total or partial resetting of pitch range after a boundary (Pierrehumbert 1980; Beckman & Pierrehumbert 1986; Truckenbrodt 2002; D'Imperio & Michelas 2014).

Interestingly, the presence of a pause is only one - and notably not even crucial - cue of a prosodic boundary. In a study on brain responses to prosodic boundaries, Steinhauer & Friederici (1999) found that locally ambiguous sentences that contain the same words but differ in the presence vs. absence of a prosodic boundary elicit a neural response at the position of the boundary that is marked by pause, final lengthening and edge tone. When the prosodic boundary is perceived and used by the listener to drive the syntactic parsing of the sentence, event-related brain potentials (ERPs) show a waveform positive shift in the temporal interval that corresponds to the boundary. This new ERP component was termed Closure Positive Shift (CPS) since it took the form of a positive shift at the closure of an intonational phrase. In later studies not only was it confirmed that the CPS is a neural response to the

prosodic boundary as a whole, but that by removing the pause from the boundary while maintaining the other cues (final lengthening and edge tone) the ERP component was still observed (Bögels *et al*. 2011).

The results of the abovementioned neural studies are in agreement with models of online and offline sentence processing which argue for the constituents of prosodic structure acting as processing units in human sentence comprehension (Carroll & Slowiaczek 1987), that prosodic information contributes to the final structuring of an initial syntactically determined parse (Pynte & Prieur 1996), and that prosodic and non-prosodic factors may enter a cue-trading relation in the process by which syntactic and semantic analyses are constructed (e.g. Beach 1991; Stirling & Wales 1996).

The importance of prosodic constituency for the comprehension of syntactic structure is also shown in a study on Italian (Bocci & Avesani 2015), which builds on the result of a previous production experiment (Bocci & Avesani 2011). In languages such as English or Italian, the default distribution of phonological prominences assigns the head to the rightmost element in a prosodic domain (e.g. Nespor & Vogel 1986). In broad focus sentences or in out-of-the-blue sentences produced without context the highest level of prominence is then assigned to the last head of the last intermediate phrase in the (last) intonational phrase. Therefore, the last word in the higher-level prosodic constituent gets the main prominence of an utterance and attracts the nuclear pitch accent. The same distribution of prominences occurs in sentences where the last lexical constituent is the (narrow) *focus*. In such pragmatic condition, the last word is at the same time the one which is most important informationally and the one that attracts the strongest metrical prominence. Phonological and pragmatic conditions concur in marking the last item as the most prominent. If the pragmatic conditions vary and the focus occurs sentence-initially, the focal element attracts the main prominence and the rest of the sentence has the informational status of *background* information which is prosodically subordinated. In English it is said that postfocal material must be de-stressed and de-accented (Selkirk 2008, a.o) with no phrase-level metrical prominence. Usually, the F0 contour is low and flat after the initial PA that is associated with the focus element and no other PAs follow. On the contrary, in some southern varieties of Italian and in Portuguese, post-focal constituents can be pitch accented, with the proviso that the associated PAs are not fully-fledged: only some types of PA can occur post-focally and their pitch span is highly compressed (e.g. Frota 2000; Grice *et al*. 2005).

Bocci & Avesani (2011) show that in Tuscan Italian post-focal material can be accented, even if the post-focal portion of the F0 contour is low and flat and shows no sign of pitch rises or falls. Their conclusion is based on sentences like (4), where a sentence-initial focussed subject "Germanico" is followed by the verb "vorrebbe invitare" and a right-dislocated object "Pierangela" represented in Fig. 4. In such sentences the focused subject is always associated with a rising PA and the following contour is low and flat.

Their production results show that the tone-bearing unit [ˈta] of the post-focal infinitive verb is longer, has higher spectral emphasis and more extreme formant trajectories than the same verb in a broad focus sentence (3). Moreover, that the last vowel and last syllable of the verb [re] are longer than in the equivalent sentence in broad focus (3). All the acoustic cues indexing phrasing and prominence indicate that even in absence of a "visible" PA, the verb acts as the metrical head of the independent ip "vorrebbe invitare", which is inserted between the ip that includes the focused subject and the ip that includes the right-dislocated object (5):

(3) [Germanico vorrebbe *invitare* Pierangela]$_{BF}$
'Germanico would like to invite Pierangela'

(4) [Germanico]$_{F}$ la vorrebbe *invitare* [Pierangela]$_{RD}$
'Germanico her-would like to invite Pierangela'

(5) [[Germanico]$_{ip}$ [la vorrebbe *invitare*]$_{ip}$]$_{IP}$ [[Pierangela]$_{ip}$]$_{IP}$

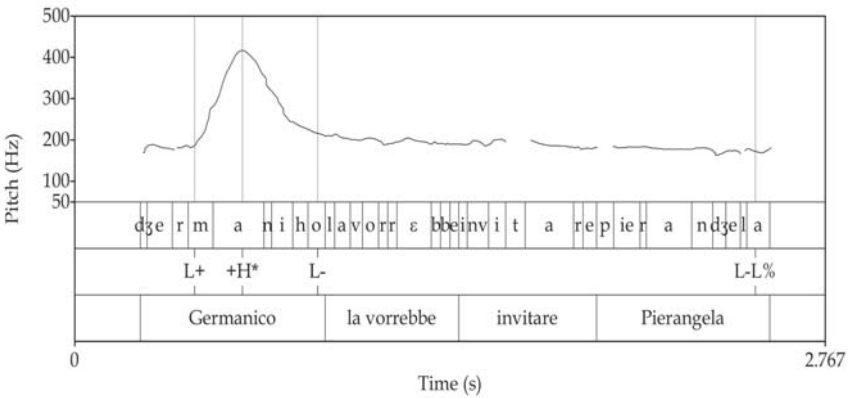


Fig. 4. Intonational contour of the sentence Germanico la vorrebbe invitare Pierangela *'Germanico her-would like to invite Pierangela':* Germanico *is the focus,* la vorrebbe invitare Pierangela *is the background where* Pierangela *is the right-dislocated object*

Building on such results, a comprehension experiment with manipulated stimuli was run in order to determine whether the conclusion of the production experiment had a psychological validity. The rationale of the comprehension experiment is based on two morphosyntactic properties of Italian. First, a clitic cannot double a focus element. Second, Right Dislocated (RDed) objects always involve a resumptive clitic, whereas subjects do not. Starting from sentences like (5) they reasoned that if the sentence is manipulated by deleting the object clitic "la" from the segmental string, in the resulting sentence the phonetic properties of the infinitive's ip-head and of the IP-boundary at its right edge still cue the final proper name "Pierangela" as right dislocated. However, because there is no object clitic, Pierangela cannot be interpreted as a RDed object and the first proper name "Germanico" in focus could be interpreted either as a focused subject or as a fronted focused object. Given the morphosyntactic and prosodic properties of the sentence, they expected the sentence to be interpreted in comprehension as OVS, with "Germanico" being interpreted as a fronted focused object and "Pierangela" being interpreted as a RDed subject. In a second run, they further manipulated the previous sentence with the excised clitic by deleting the phonetic correlates of the ip-head on the infinitive and of the IP-boundary at its right edge. Because no prosodic cue marks "Pierangela" as right dislocated any longer, a SVO order should be restored. Manipulations regarded only segments and specifically the cues of the TBUs [ta] and [re]. The results confirmed that when the infinitive is characterized by durations that correlate with the ip-head and the IP-boundary, the preferred interpretation is OVS. When head and boundary do not occur, the preferred interpretation is SVO. Overall, the results clearly indicate the fundamental role played by metrical structure (constituents and their heads) in sentence comprehension: only small duration differences in relevant positions lead to a specific metrical representation and this, in turn, leads to a specific syntactic representation.

4. *Linguistic functions of prosody and intonation*

As already mentioned in §1, prosody and intonation play a wide range of functions in communication. Some of them are clearly linguistic, such as signalling changes in sentence modality, phrasing, accentuation, and focus (Kohler 2006).

Yes/no questions are often signalled by means of prosodic and intonational changes in comparison to statements, though in some languages morphosyntactic markers are also found (Ladd 1996). Italian, on the other hand, offers a very clear example of the linguistic use of prosody and intonation, as it may indeed express the change from a statement to a yes/no question with no morpho-syntactic means, but rather intonation resources. Further, varieties of Italian may use different patterns to signal yes/no questions, as in Fig. 5 (Grice *et al*. 2005; Savino 2012; Gili Fivela *et al*. 2015).

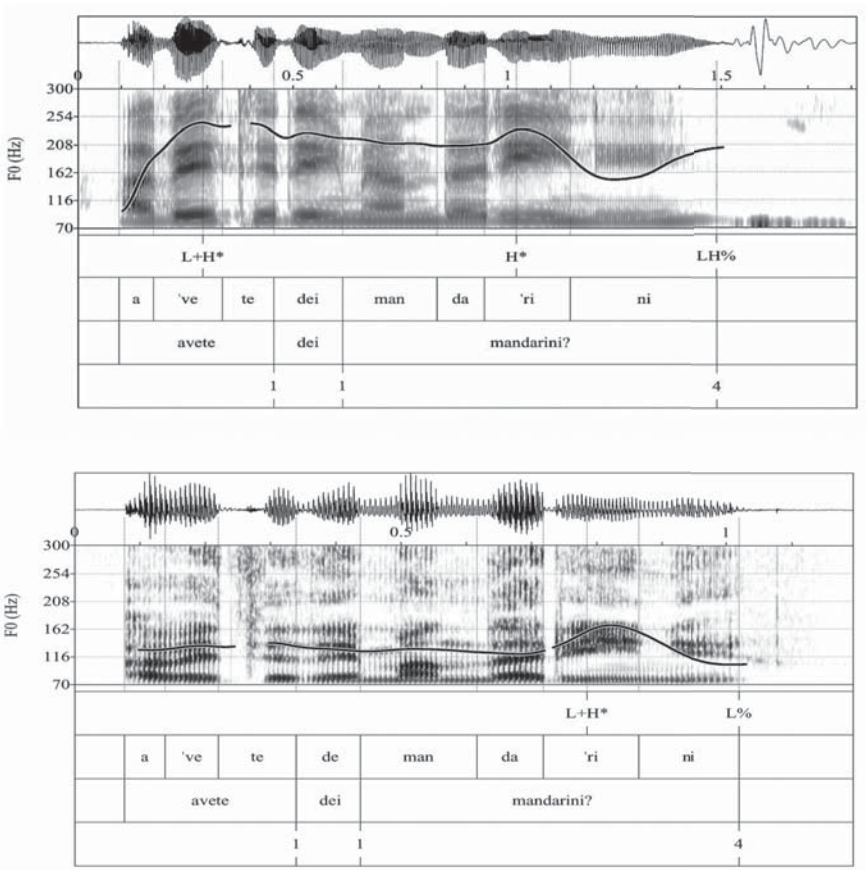


Fig. 5. Information-seeking question Avete dei mandarini? *'do you have mandarins?', speakers from Florence (top) and Salerno (bottom). From Gili Fivela* et al. *(2015)*

Another very important linguistic function, which makes the role of intonation and other prosodic features crucial for the expression of pragmatic meaning, is to highlight the *packaging* of information conveyed by an utterance (Chafe 1976; Krifka 2007). Speakers can choose to structure the information of an utterance in *topic-comment* or *focus-background* or to assign a referent the status of *given* or *new* information. Those choices are reflected in how the utterance is prosodically phrased and how post-lexical prominences are distributed in the utterance through the placement of pitch accents.

Phrasing information is crucial to highlight the information structure of the utterance as well as to solve syntactic ambiguities. Separate intermediate phrases, for instance, may separate the part of utterance that expresses what the speakers are talking about (the sentence *topic*) from the part that is related to what the speaker predicates about the topic (the sentence *comment*).

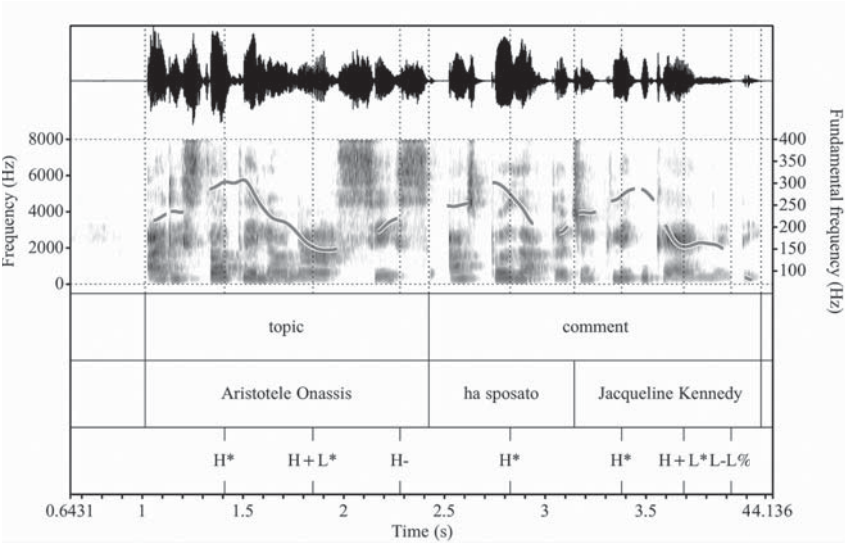


Fig. 6. Intonational contour of the sentence Aristotele Onassis ha sposato Jacqueline Kennedy *'Aristotele Onassis married Jacqueline Kennedy'. The boundary between the two ips that include the topic* (Aristotele*) and the comment* (ha sposato Jacqueline Kennedy*) are marked by a pause, final lengthening in the first ip, resetting of the pitch range and peaks downtrend in the second ip*

As an example, in (6) the *topic* "Aristotele Onassis" and the *comment* "ha sposato Jacqueline Kennedy" are coextensive with two intermediate phrases separated by a boundary which is signalled via a cluster of prosodic cues: the presence of a H- edge tone, a short pause, lengthening of the unstressed syllable before the boundary, resetting of the pitch range after the boundary and H targets downtrend within the *comment* (Fig. 6):

(6) [Aristotele Onassis]$_{\text{Topic}}$ [ha sposato Jacqueline Kennedy]$_{\text{Comment}}$
'Aristotele Onassis married Jacqueline Kennedy'

Phrasing signals also the disambiguation of a constituent's syntactic attachment, particularly of prepositional phrases, relative clauses, and adverbial phrases (a.o. Schafer 1997; Kjelgaard & Speer 1999; Avesani 1999; Hirschberg & Avesani 2000). For instance, in (7a) the absence of a prosodic boundary between "parlato" and "chiaramente" favours the low attachment of the Adverbial Phrase (AdvP) to the Verbal Phrase (VP), while in (7b) the presence of a prosodic boundary favours the high attachment of the Adverbial Phrase to the sentence root. In (7b), the high attachment of AdvP is favoured also by the prominence relation between the nuclear pitch accents of the ips that wrap respectively the VP and the AdvP: the lower height of the nuclear PA on the adverb indexes a prominence subordination of the AdvP with respect to the VP.

(7a) [[Lui] $_{\text{ip}}$ [le aveva parlato chiaramente]$_{\text{ip}}$]$_{\text{IP}}$
'He to-her talked clearly'
(7b) [[Lui] $_{\text{ip}}$ [le aveva parlato] $_{\text{ip}}$ [chiaramente] $_{\text{ip}}$] $_{\text{IP}}$
'It was clear that he talked to her'

The placement of pitch accents within an utterance serves the function to indicate which words or phrases are most salient to the purpose of the discourse, a function that directly relates to the notion of *focus* of information and to the *information status* of referents in the discourse (*given* and *new* information). Focus is a semantic-pragmatic notion (Krifka 2007). A pragmatic use of focus is to highlight the part of an utterance which the speaker presents as being important or assumes to be highly informative for the listener. If the focus is restricted on a constituent as opposed to the whole sentence, the sentence is partitioned in *focus* (the informative part) and *background* (the uninformative part). Focus is usually

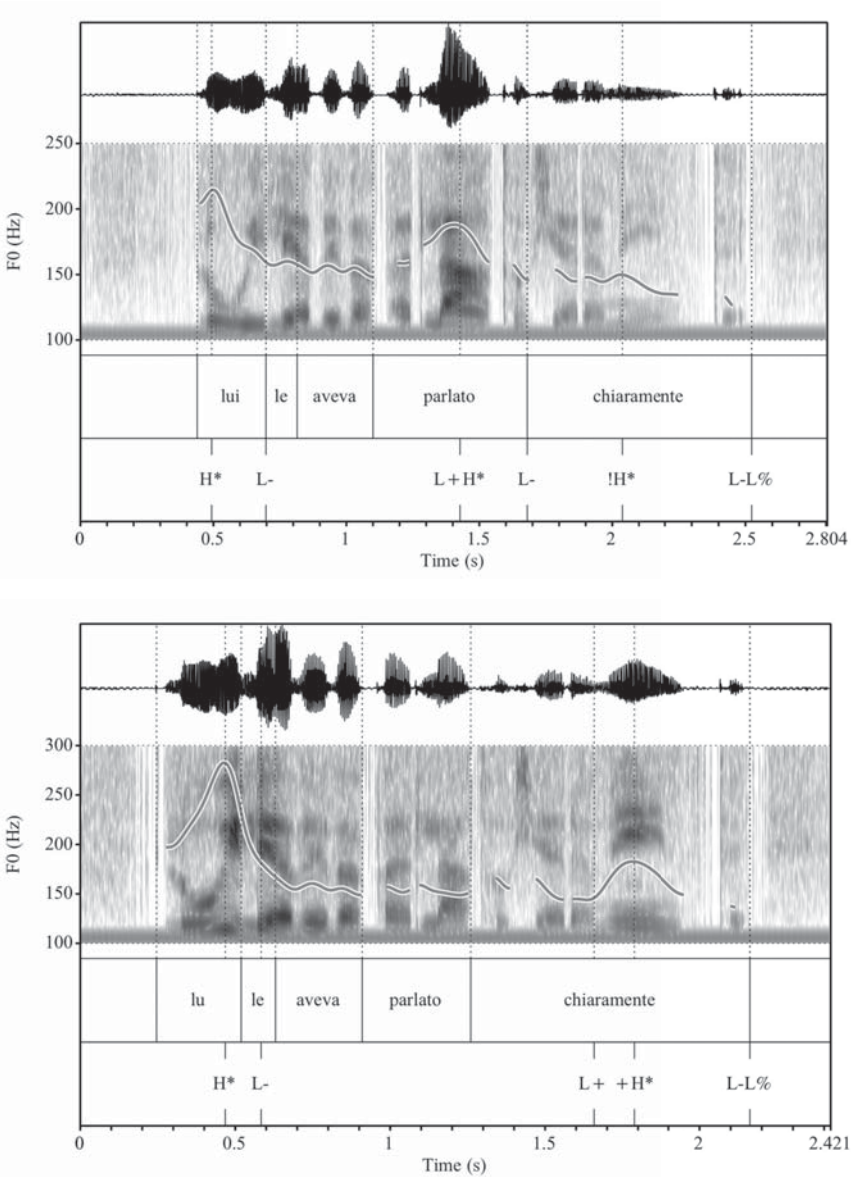


Fig. 7. Intonational contour of the sentence Lui le aveva parlato chiaramente *'he talked to her clearly' with a high attachment (top) and a low attachment (bottom) of the adverb* chiaramente*. In the high attachment the adverb is separated from the rest of the utterance by a prosodic boundary*

determined by taking into consideration the immediately preceding context. That is, according to a classic definition, the focus corresponds to the *wh*-part of a constituent question (Paul 1880, quoted in Krifka 2007) and is defined in terms of Question-Answer Congruence (a.o., Büring 2016). In the exchange in (8), the focused part "Michelangelo" is the answer to the question "a chi hanno presentato Marinella le tue sorelle? (to whom your systers presented Marinella?)".

(8) Q: A chi hanno presentato Marinella le tue sorelle
'to whom your systers presented Marinella?'

A: [Le mie sorelle hanno presentato Marinella]$_{background}$
[a Michelangelo]$_{Focus}$
'my systers presented Marinella to Michelangelo'

According to a more recent formulation, focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions (Rooth 1992). Focus evokes a set of alternative propositions that differ only for the focused element: in the case of (8) the focus on Michelangelo evokes a set of alternative propositions (indicated in braces) such as {le mie sorelle hanno presentato Marinella *a Giacomo* ("my sisters presented Marinella *to Giacomo"*); le mie sorelle hanno presentato Marinella *a Luca* ("my sisters presented Marinella *to Luca"*)*;*….} and from this set the proposition "le mie sorelle hanno presentato Marinella *a Michelangelo* ("my sisters presented Marinella *to Michelangelo"*)" has been selected. The background is defined as the invariant part of the alternative propositions.

The answer in (8A) is a case of *narrow focus,* and in languages like Italian it is marked by a (nuclear) pitch accent on the focus constituent[8]. A special type of focus is *contrastive focus* (CF) that can also be used as a correction of what has been previously said. In Figure 8, (8A) has been produced by a Sienese speaker with a focus of contrastive-corrective import on "a Michelangelo". In the F0 contour, a nuclear L+H* pitch

[8] When the focus is not restricted to a single constituent it is *broad*. In these structures the relation between focus and accent is no longer straightforward and a pitch accent on one word, called the focus exponent, marks the larger focus domain (the phenomenon is called *focus projection*). In Italian the focus exponent in broad focus structures is the last word of the sentence.

accent is associated with the TBU of the focused constituent ("CFocus") and two prenuclear pitch accents are associated with two noun phrases in the background ("background (given)"). If the focus phrase is moved sentence initially as in (9), the element in focus is marked by same L+H* nuclear pitch accent but the prosodic properties of the background are radically different, as no fully-fledged pitch accent occurs post-focally (Fig. 8).

(9) [a Michelangelo]$_{\text{Focus}}$ [le mie sorelle hanno presentato Marinella]$_{\text{background}}$
'to Michelangelo my systers introduced Marinella'

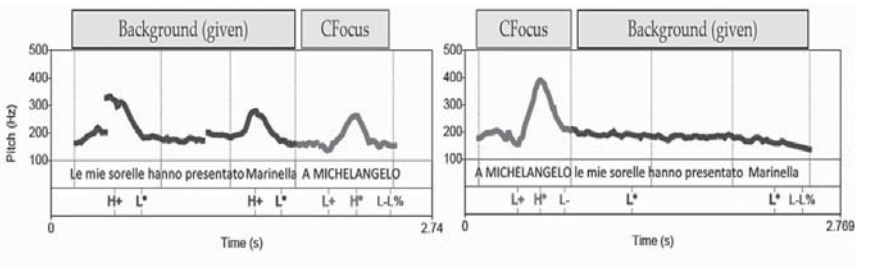


Fig. 8. Intonational contour of the sentence Le mie sorelle hanno presentato Marinella a Michelangelo *'my sisters introduced Marinella to Michelangelo' (left) and* A Michelangelo le mie sorelle hanno presentato Marinella *'to Michelangelo my sisters introduced Marinella' (right). Michelangelo is a contrastive-corrective focus in sentence-final position (left) and in sentence initial position (right). (Courtesy of Giuliano Bocci)*

Besides a pragmatic, there is also a semantic use of *focus*, that leads to change the truth-conditional value of a proposition. This is the case of focus-sensitive operators such as the particles "only", "even" or the negative quantifier "not", in which the linguistic element modified by the logical operator is marked as focused by the association with a pitch accent (for examples in Italian and English see Hirschberg & Avesani 2000).

Further, in Italian different types of focus are distinguished by different types of pitch accents: *information focus* (IF) and *contrastive-corrective focus* (CF) are marked by different types of pitch accents in many varieties

of Italian (Avesani 2003; Avesani & Vayra 2004; Bocci & Avesani 2008; Gili Fivela *et al*. 2015). That is, differently from English in which the same type of pitch accent is used for the two imports of *focus* in the same sentence position, in Italian no ambiguity arises in identifying IF and CF in sentence final position. In Florentine and Sienese Italian, for instance, IF associates with a falling accent, phonologically specified as H+L*, while CF associates with a rising accent, phonologically specified as L+H* (Fig. 9).

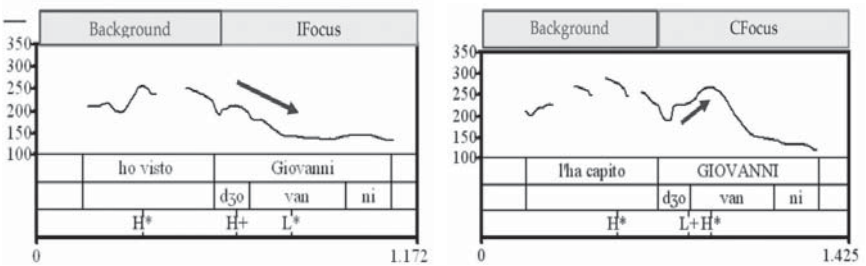


Fig. 9. Information focus (left) in the sentence Ho visto Giovanni 'I have seen John' (as a reply to the question Chi hai visto? *'Who did you see?') and contrastive-corrective focus in the sentence* L'ha capito Giovanni *'it-understood John' (as a reply to the question* Sembra che lo abbia capito Leo *'it looks like it-understood Leo').*

Intonation also encodes discourse-related properties such as the *information status* of referential expressions. With information status of a referent we indicate the specific relation between a linguistic entity and the corresponding non-linguistic entity that holds in the mind of the speaker/hearer or in the discourse model at the moment of communication, which can dynamically change as the discourse evolves (e.g. Chafe 1976). A linguistic expression can be *given* if its representation is already present in the mind of the speaker/hearer because: a) it is part of the encyclopedic knowledge or b) it is part of the shared knowledge of the speaker and the hearer; c) it is visible in the external context, d) it is explicitly present in the immediate linguistic context (i.e. already mentioned), or e) it stands in a hyperonymy relation with its antecedent (Baumann & Riester 2013). The status of *given* is uniquely determined by the knowledge and attention

state of the interlocutor at a specific moment of a conversation (Prince 1981; Lambrecht 1994; Grosz & Sidner 1986). *New* is a referent recently entered in a discourse and not recoverable from the preceding context.

Many studies in the recent and past literature have shown that there is a correlation between how prosody marks a linguistic entity and its information status: a *new* linguistic entity is usually pitch accented and a *given* entity is usually *deaccented*, that is, it lacks a pitch accent that would otherwise be used to mark the same elements if it was occurring in all-new utterances. Listeners are sensitive to the prosodic marking of an entity's information status: accenting of new information and deaccenting of *given* information affects off-line sentence comprehension (Birch & Clifton 1995) as well as on-line processing (e.g. Dahan *et al*. 2002). Moreover, neurolinguistic studies using event-related potentials and investigating the impact of different types of accentuation on the comprehension of referents have shown that appropriate prosodic cues affect the construction of a mental model (e.g. Schumacher & Baumann 2011).

However, not all languages follow the same pattern of accentuation/deaccentuation of a referent according to its status of *new* or *given* information. First of all, the association between deaccenting and information status is not to be conceived as an exceptionless one-to-one relationship, but at most as a strong association (Brown 1983; Terken & Hirschberg 1994; Bard & Aylett 1999), as many intonation patterns that are claimed to convey a certain meaning only represent the most frequent pattern that speakers choose to use in that context (Braun & Chen 2012). Second, *given* can be seen as a scalar notion in which, based on a scale of assumed familiarity, at least three categories are defined: *new*, *given* and *accessible* information (Chafe 1976; Lambrecht 1994; Prince 1981, 1992). Along this line, it was shown that in German *given* and *textually accessible* information are preferably deaccented (respectively: 78% and 63%) while *inferentially accessible* information is preferably accented (64%) with a H* or a downstepped !H* pitch accent (Röhr & Bauman 2010). Even with those provisos, though, it is widely accepted that Germanic languages avoid marking as prosodically prominent referential expressions that strictly convey *given* information.

Contrary to Germanic languages, Romance languages fail to deaccent referents which are informationally *given*. For Italian this was firstly observed by Cruttenden (1993) and Ladd (1996) and later experimentally proved by Avesani (1997), Avesani & Vayra (2005), Swerts et al. (2002).

Pitch accenting *given* referents is reported to occur in different speaking styles such as spontaneous speech (Avesani 1997) or task-directed dialogues (Avesani & Vayra 2005). In the latter study, only 6.5% of coreferential expressions were reported as lacking a pitch accent. Further, in comparing accentuation strategies of typologically different languages such as Dutch and Italian, Swerts *et al*. (2002) showed that Italian speakers always accent *given* items while Dutch speakers always deaccent them; moreover, Italian speakers cannot perceive any difference in prominence between *given* and *new* items (while Dutch listeners can) and they are unable to reconstruct the dialogue history on the basis of the accentuation of an item, while Dutch listeners are able to guess whether a referent was already mentioned in the preceding dialogue.

Avesani & Vayra (2005), however, observed that some cases of coreferential nouns, albeit few, were produced with a pitch accent. All cases related to *given* referents which occurred in longer syntactic constituents, specifically in post-focal position of sentences with fronted foci. Bocci & Avesani (2011) and Bocci (2013) disentangled the question arguing that *given* constituents which occur post-focally are not deaccented, but are assigned phrasal stress, overriding their information status of *given* and part of the background. They are marked by all prosodic cues that identify them as post-lexically stressed, and by a pitch accent that in Tuscan Italian is a L*. By taking into account only the melodic contour though, it could be said that post-focal *given* elements are "deaccented", as superficially no fully-fledged pitch accent (high, rising or falling) is observed. But a more thorough prosodic analysis and distributional considerations argue for the contrary: as can be appreciated in Fig. 8, changing the focus-background partition of the sentence, the same *given* elements that in post-focal position appear as deprived by a fully-fledged pitch accent in pre-focal position clearly bear a H+L*.

5. *The meaning of tunes*

In a seminal paper Pierrehumbert & Hirschberg (1990) address the contribution of the choice of *tune*, or intonational contour, to discourse interpretation. While in the literature the characterization of the meaning of a given tune has been interpreted in terms of speaker attitudes (politeness, surprise, deference etc.), speech acts (statements, requests, contradictions), propositional attitude (belief, uncertainty, etc.), presupposition and focus,

Pierrehumbert & Hirschberg (1990) claimed that neither speech acts nor propositional attitude provided sufficient characterization of available tunes in English. Rather, they claimed that tunes specify a particular relationship between the propositional content of the utterance and the mutual beliefs of discourse participants: speakers choose a specific tune to convey a particular relationship between an utterance, the current beliefs of the hearer(s) and the anticipated contributions of subsequent utterances. They also proposed that these relationships are compositional, composed of pitch accents, phrase accents and boundary tones that make up a tune. Therefore, the main components of intonation offer separate and distinct contributions to discourse interpretation which are related to mutual belief spaces in conversation, capturing the intuition that tunes sharing certain tonal features also share some aspects of meaning (Gili Fivela 2008, Prieto 2015).

Differences in accent type convey differences in meaning when interpreted in conjunction with differences in the discourse context and variation in other acoustic properties of the utterance. For example, in English H* accents are typically found in standard declarative utterances and are commonly used to convey that the accented item should be treated as *new* information in the discourse, and is part of what is being asserted in an utterance. L* accents are broadly characterized as conveying that the accented item should be treated as salient, but not part of what is being asserted. In English, L+H* accents can be used to produce a pronounced "contrastive" effect and H+!H* accents are associated with some implied sense of familiarity with the mentioned item.

As for phrasal tones, phrase accents indicate the presence of an interpretive as well as a phonological boundary (Pierrehumbert & Hirschberg 1990: 302): H- indicates that the current phrase is to be taken as forming part of a larger composite interpretive unit with the following phrase, while a L- emphasizes the separation of the current phrase from a subsequent phrase. The type of boundary tone conveys whether the current intonational phrase is forward-looking or not, that is whether this is to be interpreted with respect to some succeeding phrase or whether the direction of interpretation is unspecified.

Recent proposals have built on Pierrehumbert & Hirschberg's compositional approach to explore the meaning of English (Truckenbrodt 2012) and French pitch contours (Portes & Beyssade 2012). They argue for a systematic relationship between tonal features and their semantic primitives, but they also assume that these meanings are to some extent

context-dependent. Many of these proposals consider that intonation encodes basic meanings from which context-dependent conversational implicatures can be derived.

As an example, Armstrong & Prieto (2015) explored how intonation and context conspire to lead a listener to a given meaning. Their experimental evidence points to the dynamic interaction between context and contour, and also to the fact that individual intonation contours can differ in the type and number of meanings they convey.

6. *Acquisition and crosslinguistic comparisons*

The impact of first language (L1) prosody on second or foreign language (henceforth, L2) is widely discussed in the literature. Think of the grounding work by Mennen (2004) who, analyzing Dutch L1 consecutive bilinguals who learned Greek as L2 in their early adulthood and used it regularly, showed that pitch alignment characteristics of Dutch L1 affect the alignment features of prenuclear rises produced in Greek L2.

However, in our aim to discuss how linguists analyze prosody and intonation, we describe some studies that address issues concerning Italian and L2 learning as well as crosslinguistic comparison. Besides the specific phenomena and languages considered, the first study more generally regards prosodic structure and segmental phenomena in the acquisition of an L2, the others relate to the implementation of prominences in L2.

6.1 Prosody, constituency and vowel insertion in French L2

The prosodic structure has an impact on various aspects of speech production, among which the realization of segments that precede or follow prosodic boundaries. The role of prosodic structure in this respect has been observed in relation to both L1 and L2. As for the latter case, interesting observations stem from an acoustic and articulatory investigation related to the production of consonant clusters in French L2 by Italian speakers (D'Apolito & Gili Fivela 2013, 2018).

Sibilant clusters are common in French (where they can also undergo place assimilation; Niebuhr *et al*. 2008), while they are marked in Italian, where they are not even found across word boundaries, being the word

ending usually a vowel (few exceptions are represented by prepositions, loanwords and contexts in which word final vowel truncation occurs; Muliačić 1973; Farnetani & Busà 2004). Thus, sibilant clusters are phonotactically marked (Eckman 2008) for Italian speakers who, as a general repair strategy to produce such unusual sequences, may insert a vowel between the two consonants. Also depending on prosodic conditions, such vowel insertion may actually correspond to either an epenthetic or an intrusive vowel (Hall 2003, 2006, 2011), that is a vowel with or without an articulatory target. Besides theoretical implications, distinguishing between the two types may show the influence of prosodic factors on segment production and may be useful in order to shed light on their phonetic transcription.

D'Apolito & Gili Fivela (2013, 2018), investigated French heterosyllabic sibilant clusters (alveolar-postalveolar and postalveolar-alveolar sibilant sequences, such as /sʃ/, /ʃs/, /sʒ/, /ʒs/, /zʃ/ e /zʒ/), by creating a speech corpus of acoustic and articulatory (electromagnetic articulography, EMA) data in which consonant sequences were inserted in carrier sentences in which they were realized at a word boundary (/a_/#/_i/). Such word boundary could correspond to either a phonological phrase boundary (e.g., *Il dit tasse chinoise rapidement* 'Dice tazza cinese rapidamente') or an intonation phrase boundary (e.g. *D'abord il a dit tasse. Chinoise l'a dit après* 'Prima ha detto tazza. Cinese l'ha detto dopo'). Three advanced Italian learners of French-L2 (Lecce, Italy) and two native French speakers (Nantes, Paris) produced seven repetitions of the French corpus, both at a fast and at a normal speech rate. The authors performed an auditory evaluation (aimed at verifying the presence of the expected prosodic boundary, the presence of an inserted vowel and the consonant realization) and both an acoustic and an articulatory analysis. The analysis related to the V1C1#C2V2 sequences, including the presence of possible schwas (V0) and/or pauses (P), and it was carried on by performing acoustic measurements of duration (single segments, as well as utterance duration), speech rate (number of syllables/utterance duration), and formant values (F1, F2) for /a/, /i/ and the possible schwa. Acoustic results show that speakers differentiated between the two speech rates, and, almost with no exception, inserted a vowel at normal rate and in the case of a weak boundary; Italians kept inserting a vowel-like segment in the case of a stronger boundary, while French speakers showed a more variable behavior (as for articulatory analysis and results, we refer

to D'Apolito & Gili Fivela 2013, 2018). As far as the fast rate is concerned, Italians inserted vowels, though less than in the normal rate condition, while French inserted very few vowels in the case of strong boundaries and no vowel at all in the case of weak boundaries. Thus, the presence of a prosodic boundary is shown to clearly interact with vowel insertion in both French L1 and French L2 speech. Interestingly, formant values showed that French speakers realized schwa-like vowels, while Italians produced a more closed and anterior vowel, whose quality seemed to be affected by the prosodic context and, in any case, resembled more the following [i], rather than a schwa – see Fig. 10.

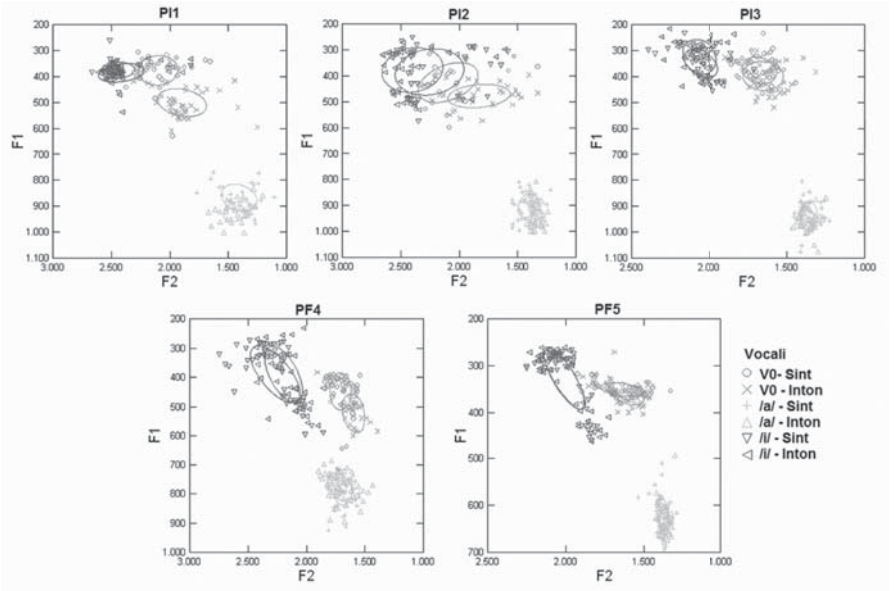


Fig. 10. Formant values of /a/, /i/ and inserted V0 vowels, produced in different prosodic conditions by Italian learners (upper panels) and French natives (lower panels; adapted from D'Apolito & Gili Fivela 2018).

Overall acoustic (and articulatory) results point to the need of taking care of prosodic conditions and differentiating the phonetic transcription, using [ɨ] rather than /ə/ for productions by Italians, who probably realize an intrusive, rather than an epenthetic vowel (with no articulatory target).

6.2 Pitch accent realization, distribution and information status in German L2

As already mentioned, pitch accents play a crucial role in conveying linguistic functions and are indeed language specific as for both their form-function mapping and their implementation, i.e. their phonetic characteristics including the temporal relation between the tonal target(s) and the segmental chain (see §2.1 and §2.2).

Not surprisingly, pitch accent realization in L2 may be as difficult as the production of other phonological events. In investigating the L2 Greek speech by Dutch L1 speakers, for instance, Mennen (2004) showed that non-native productions of L+H* prenuclear rises in L2 Greek declaratives are characterized by an earlier F0 peak alignment in comparison to native Greek productions. Along similar lines, Atterer & Ladd (2004) showed that Northern and Southern speakers of German differed in aligning the L+H* prenuclear rise in English L2 productions, in that Southern speakers produced a later alignment of both L and H targets in comparison not only to English speakers but also to Northern German speakers. Very similarly, results on Italians producing prenuclear pitch accents in German L2 (Stella 2013) showed that production accuracy varies depending on the learner's competence in L2. By means of both acoustic and articulatory data, the author shows that low competence speakers show the same pitch accent observed in their L1 (corresponding to a L+H* transcription). On the other hand, high competence learners show a more stable anchoring of tonal targets to segments and a later alignment of the expected low target, which goes in the direction of the German pitch accent (described as L*+H; Braun 2006).

In learning a foreign language, the strategies of PA assignment specific of an L2 can pose difficulties as well, especially if the native language and the target language differ typologically and what needs to be attained is mastering the prosodic properties at the interface with information structure. That is the case of Italian and German, which differ as for the accenting or deaccenting of referents that are informationally *given*. We have seen in §4 that German tends to deaccent *given* referents while Italian does not. Moreover, German differs from Italian on another respect: in broad-focus verb-final sentences the verb can be accented or deaccented according to the status of argument or adjunct of the element that precedes it (Truckenbrodt 2007), while the last lexical item of the same sentences in Italian are always pitch accented. From a typological point of view,

German can be said to be marked with respect to Italian, because both languages obey the same phonological rules of pitch accent placement (the last metrical head of the final intonational phrase gets the highest prominence of the sentence), but in German deaccenting is driven both by informational principles ('deaccent *given'*) and by syntactic constraints that do not apply in Italian. Based on those structural differences, an asymmetry in the acquisition process can be predicted: structures of an L2 that are marked will be more difficult to be learned than unmarked ones (Eckman, 2008).

A study on the acquisition of prosody of L2-Italian and L2-German by native speakers of German and Italian explores this topic (Avesani *et al*. 2015). By exploiting the same card game methodology used by Swerts *et al*. (2002), a total of five pairs of German-German and Italian-Italian speakers produced in a semi-spontaneous way NPs in which the Adjective (a colour) or the Noun (a fruit name) is *given*, *new* or *contrastive*. The speakers first played the game in their L2 and then played the game in their L1. The experimental set up allows to combine a contrastive analysis of the native languages (L1-German vs. L1-Italian), a contrastive analysis of the speakers' interlanguages (L2-Italian vs. L2-German) and an analysis of speakers' interlanguages with their native languages (L1-Italian vs. L2-German; L1-German vs L2-Italian). Results show that in Italian the final word of the noun phrase is always accented independently from its information status. When it represents *given* information, it is pitch accented in 100% of the cases. When the *given* word is NP-initial, it can be optionally accented, as it occurs in prenuclear position. On the contrary, in German the last word in the NP when it is *given* is deaccented (i.e. it is not associated with a pitch accent) in 87% of the cases and in a lesser percent also when it occurs NP-initially.

The analysis of interlanguages confirms that Italian speakers (who are advanced learners of German) transfer in their interlanguage the distribution of the accentual prominences of their L1 and do accent a German Noun or an Adjective if it is informationally *given*. Differently, when German learners of the same proficiency level speak L2-Italian they show to have acquired the prosodic accentuation of the target language and properly accent *given* information. The authors interpret the results in terms of the different cognitive weight faced by the Italian and the German learners in producing the correct accentuation in the L2. The Germans have only to select one of the strategies of accentuation already

present and active in their L1: the "structural" accentuation, according to which the last word in a broad-focus sentence that is not a verb gets a pitch accent. Conversely, to properly produce the prosody of L2-German, Italians have to master a specific type of pragmatic (de)accentuation that is not present in their L1, as well as its interplay with the phonological structure. For the Germans, the acquisition process is reduced to a suspension of the pragmatic constraints that govern the distribution of the prosodic prominences in their mother tongue; consequently, the default phonological rules do take over, and all NP final words are accented independently of their information status. On the contrary, Italians have a more difficult task: they must realize that prominences' distribution is not only phonologically-based, and that the highest prominence is not necessarily allocated rightmost in a phrase. Then, they have to master a new type of pitch accent association, which is largely ruled by the information status of the lexical items in the NP.

7. *Prosody and gestures*

Since the beginning of the Seventies, a tight relation has been observed between spoken utterances and movements of the hands, head, face and torso. According to Kendon (1972, 1980, 2004), gestures accompanying speech are organized into a hierarchy of constituents which resemble the hierarchy proposed for prosody. Since his work, arguing in favour of a coordination between gestures and Tone Groups (with reference to the proposals of the British School of Intonation, Crystal 1969), various works have been showing that gestures and speech are synchronous. Specifically, the most prominent segment of gestures tends to co-occur with the most prominent segment of speech (e.g., Birdwhistell 1952, 1970; Kendon 1972, 1980; Loehr 2012, among many others), that is, gesturing is timed to prominent syllables. Further, this timing can be influenced by prosodic boundaries.

Co-speech gestures play a crucial role in helping people to comprehend speech, especially in the case of unclear or ambiguous stretches. Further, different types of co-speech gestures have been identified, showing specific relations with the speech message and, therefore, different roles in speech coding and decoding (McNeill 1992; Kelly & Church 1997; Morsella & Krauss 2004; Goldin-Meadow & Beilock 2010; Goldin-Meadow 2013).

The main co-speech gestures, in line with McNeill's (1992) book, are *referential* and *non-referential* gestures. The former visually refer to the content of speech, while the latter offer information on the form rather than the content of the utterance. Among referential gestures, some well-known types are *deictic*, which are typically pointing gestures to specific locations, performed by means of, e.g., fingers, head, nose; *iconic* and *metaphoric*, visually illustrating concrete and abstract aspects respectively of the speech content. On the other hand, non-referential gestures, also known as *beats* (or batons; Efron 1941/1972), are two-phase movements (in/out, up/down) which do not present any discernible meaning, though they co-occur with important parts of the message. For instance, both pointing gestures and beat gestures have been shown to be tightly synchronized to prosodic prominence in spontaneous speech (Loehr 2012; Esteve-Gibert & Prieto 2013; Wagner et al. 2014). In general, *hits*, which are defined as "gestures with sudden sharp end points" (Shattuck-Hufnagel & Ren 2018: 206, but see also Shattuck-Hufnagel *et al.* 2007: 39), occur "towards the end or just after a spoken accented syllable" (Hufnagel & Ren 2018: 214). Hits have been found to be temporally synchronized to prosodic events such as pitch accents and boundary tones in Italian too (Esposito *et* al. 2007).

As far as pointing gestures are concerned, Esteve-Gibert & Prieto (2013) analyzed fifteen Catalan speakers while they pointed at a screen and produced target words with stress in different positions (es. "máma" vs "mamá"). They induced the production of target words in contrastive focus contexts, in which they were pitch accented and followed by a phrase boundary. Their analysis of pointing gestures, pitch accents and boundary tones showed that the apex of deictic gestures is coordinated with the intonation peak, and that the entire pointing gestures are bound by prosodic phrasing. As the authors observed, "the timing of their starting movements and prominence peaks (F0 peak and apex) varies if there is a preceding or an upcoming prosodic phrase boundary" (Esteve-Gibert & Prieto 2013: 863).

Besides being aligned with prosodic prominences, and pitch accents in particular, pointing gestures have indeed been found to be coordinated with prosodic boundaries. The literature on the marking of prosodic boundaries is rich, and regards signals of both incoming breaks (e.g., preboundary lengthening, boundary tone implementation) and post-boundary events (e.g. F0 reset, phrase initial strengthening).

In their articulatory investigation, Krivokapic *et al*. (2015), for instance, found that pointing gestures show phrase initial temporal lengthening which increases with boundary strength. They collected audio and synchronized kinematic data by means of electromagnetic articulometery and a motion capture system while asking their participants to point to the appropriate picture of a doll (named either miMA or MIma) while reading a sentence including the target name. By changing the stress position in the target word (first or second syllable of a disyllable, i.e., MIma and miMA), and the structure of the sentence including the target word, specifically with respect to the strength of the phrase initial boundary (word boundary, ip—intermediate phrase boundary, IP—Intonation Phrase boundary), they could collect data on the alignment of the pointing gesture and oral constriction gestures corresponding to the production of the two target words in the different prosodic conditions. Their results, though preliminary, clearly show that "1) manual and oral gestures are longer phrase-initially than phrase medially and 2) manual and oral gestures lengthen under phrase-level prominence" (2015: 1) (Fig. 11).

Summing up, the prosodic structure (heads and edges) plays a strong role on gesture timing. Gesture movements are bound by both prosodic heads (pitch accents), that is by prosodic prominence - which is related to, e.g., the management of information structure, focus, and discourse marking -, and by prosodic edges (prosodic boundaries), that is by phrasing - which is related to the managements of, e.g., syntactic grouping and turn-taking.

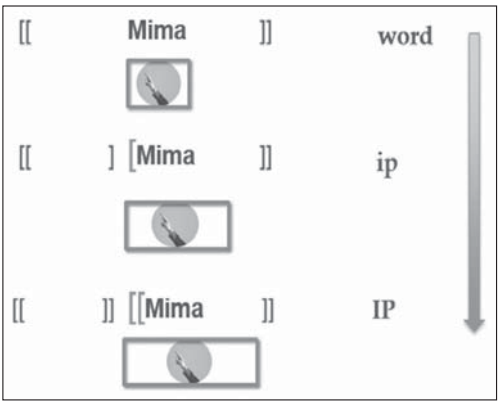


Fig. 11. Schematic representation of Krivokapic et al.*'s (2015) results on the coordination of the pointing gesture with the beginning of different prosodic boundaries.*

However, the relation between prosody and gestures goes beyond the involvement of hand movement, and prosody-related visual information may "spread" over different units. In their work on the interplay of contextual and prosodic information in the coding of politeness, Gili Fivela & Bazzanella (2014) showed the role of visual information in creating two local contexts. Specifically, the local context available to participants in the conversation, who can see each other, and a second, related context, that involves also someone who is not physically present and cannot see the interlocutors. The example the authors describe concerns a conversation that someone may have on the phone, with interlocutors who do not share visual information, while communicating also with someone who is physically present and therefore shares visual information on the context. For instance, in one of their recordings, involving a speaker who expresses on the phone her appreciation with regards to a piece of furniture while denying it with body (hand) gestures and visual expressions, it may be easily seen that both highbrow rising and hand movements participate in conveying the denial of the oral message, and that the eyebrow rising stops before the hand movement does.

The specific contribution of visual and audio information in the coding and decoding of prosody is still a debated issue (Massaro 1989; House 2002; Krahmer & Swerts 2005; Borràs-Comes & Prieto 2011, Crespo Sendra *et al*. 2013, Ambrazaitis & House under review), but the intertwined contribution of both channels has been shown even in relation with the communication of clearly linguistic information. Gili Fivela (2015), for instance, reported on the role of visual information in conveying sentence modality in Italian (variety of Lecce). Specifically, the paper focusses on the way speakers differentiate statements, *wh*-questions and exclamations by means of both prosodic and visual information. In analysing utterances supposing a positive attitude, the author finds that *wh*-questions differ from the other modalities considered as for head movement, while statements differ from other modalities as for eyebrow and lid movements.

7.1 Intonation and visual expressions in Catalan, Dutch and Italian

As already mentioned, there is a tight connection between prosodic events and visual expressions, and, not surprisingly, crosslinguistic differences have been reported with reference to multimodality as well. Crespo Sendra *et al*. (2013), for instance, investigated information seeking

and incredulity yes/no questions in Catalan and Dutch in order to check if differences in the phonology of intonation have an impact on visual information. The authors observe that in Catalan information seeking questions are expressed by means of a L*+H H% pattern, and incredulity questions show the same phonological patter which is realized, though, by reaching a higher final tonal target. On the contrary, in Dutch the two question types are expressed by means of different sequences of phonological events. In particular, information seeking questions are expressed by means of a L* HH%, while incredulity questions show a L+H* LH% pattern. As far as visual information is concerned, the authors observed that incredulity questions show some degree of eyebrow furrowing and eyelid closure in both languages. Results of Crespo Sendra *et al*. (2013) investigation on the perception of audio-visual information by Catalan and Dutch listeners show that the former give more importance to facial cues than the latter, and suggest that this may be due to the more subtle (or ambiguous) information conveyed by the audio channel in Catalan.

A question then arises as for Italian, where the same phonological pattern, H+L* L%, may be used to convey both statements and *wh*-questions, with no disambiguation at the phonological level, similarly to what was observed for Catalan in relation to yes-no information-seeking questions and incredulity questions. Gili Fivela (2015) addresses this issue by means of both production and perception data, regarding neutral statements, *wh*-questions suggesting surprise, and exclamations (for the sake of clarity, here only data on statements and questions are reported; as for exclamations, see Gili Fivela 2015). In the production experiment, five subjects were audio and video recorded while producing dialogues including the target sentences. The AM analysis of intonational patterns confirmed that both statements and *wh*-questions were realized by means of a H+L* L% pattern; however, the analysis of visual expression and head movements (*Facial Action Coding Scheme*, Ekman 1982; Ekman *et al.* 2002) pointed out that *wh*-questions differed as for both head movement and visual expressions. Specifically, statements differed from other utterances as for eyebrow and lid movements, and questions differed from others as for head movements. These results seem to point to a possible integration of audio and visual information, with the latter compensating for the lack of phonological differences.

However, perception results offered a different picture. Subjects had to decide on the modality expressed by audio-video stimuli that were either regular, congruous stimuli corresponding to a specific modality, or

incongruous stimuli in which audio and video did not match. Results showed that subjects were strongly influenced in their judgements when the video corresponded to a question and the audio to a statement. In such situation, statements were recognized only in 55% of cases, while in 40% a question was identified. In the reversed case, in which the audio of a question was imposed over the video of a statement, nothing similar happened, as listeners perceived questions in 93% of cases. Thus, visual information corresponding to a question influences the perception of the audio of a statement, and not the other way around. This is taken to point to a more complex picture than that depicting a balancing of audio and visual information in terms of a negative correlation, that is a balance in which the role of one channel depends on what happens in the other. The balancing of information within the same channel seems also to be important, in that marked facial expressions (in this case, corresponding to surprised *wh*-questions) seem to affect the interpretation of utterances which are not associated to marked information on the same video channel (here, neutral statements)

Thus, a more complex integration could possibly take place and should be considered even in a crosslinguistic perspective. Such integration could possibly play a role in accounting for the variability observed in audio realizations, and could be taken into account in the transcription of intonation categories and contours, in both L1 and L2 studies.

8. *Conclusions*

The paper offered an overview of how linguists analyze prosody and intonation. By discussing the main features of prosody and intonation in terms of both form and function, the added value of a framework that hypothesizes a phonological structure for intonation was shown. The Autosegmental-Metrical framework allows investigators to identify linguistically relevant units, which may be both phonologically labelled and phonetically measured. The analysis may then regard both phrasing and prominences. Further, it may also be performed with reference to a transcription system, the ToBI system, which needs to be developed for each specific (variety of) language, being phonological in nature, but it is based on the very same principles, in that it requires a phonological coding and it also allows for a phonetic analysis of relevant units. Such framework may then be ideal in analyzing language/variety specific phonological inventories.

Specific attention was devoted to the main functions that are crucially played by prosody and intonation, showing that, besides conveying prominence in general, they express sentence modality, focus, phrasing and information structure and, unsurprisingly, they may solve global ambiguities.

Further, issues were addressed concerning acquisition and crosslinguistic comparison, again with the aim of discussing how linguists analyze prosody and intonation. The attention was oriented towards Italian and some works investigating L2 acquisition with respect to prosodic structure and segmental phenomena (specifically, prosody, constituency and vowel insertion in French L2), and prominence patterns and L2 acquisition (specifically, the study of pitch accent realization, distribution and information status in German L2).

Finally, multimodality was also addressed, showing the tight interplay of prosody, co-speech gestures and visual expressions, and pointing out that the analysis of multimodal information requires to refer to prosodic units and may be relevant for linguistic purposes. Some works underlying the interplay of crosslinguistic differences and multimodal information modulation were also discussed in the end (specifically, the role of intonation and visual expressions in Catalan, Dutch and Italian).

Overall these studies show that prosody and intonation have been analyzed by adopting various perspectives and methodologies, but they all refer to a linguistic structure and to linguistic functions, as well as to an abstract representation of intonational events. These aspects allow not to be mislead by phonetic details and variability, offering the reference for analyzing a continuously varying signal, that is the verbal chain. Only adopting a shared framework, including both a phonological and a phonetic level and accounting for both phrasing and prominences, allows to succesfully face investigations on either L1 or L2. Models assuming a direct correspondence between form and function would not be equally successful in this respect.

References

Ambrazaitis, Gilbert & House, David. Under review. Is phonetic prominence underlyingly multimodal? *Laboratory Phonology, Special Collection on Phonological categories and prosodic modulation: importance, representation and implementation*.

Armstrong, Meghan & Prieto, Pilar. 2015. The contribution of context and contour to perceived belief in polar questions. *Journal of Pragmatics* 81. 77–92.

Arvaniti, Amalia & Ladd, Robert D. & Mennen, Ineke. 2000. What is a starred tone: Evidence from Greek. In Broe, Michael & Pierrehumbert, Janet (eds.), *Papers in Laboratory Phonology V*, 119–131. Cambridge: Cambridge University Press.

Arvaniti, Amalia & Ladd, Robert D. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26. 43–74.

Arvaniti, Amalia. 2007. On the Relationship between Phonology and Phonetics (Or Why Phonetics is not Phonology). In Trouvain, Jürgen & Barry, William John (eds.), *Proceedings of the XVI International Congress of Phonetic Sciences*, 19–24. Saarbrücken: Universität des Saarlandes.

Arvaniti, Amalia. 2012. Segment-to-Tone Association. In: Cohn, Abigail & Fougeron, Cécile & Huffman, Marie (eds.), *The Oxford Handbook of Laboratory Phonology*, 256–275. Oxford: Oxford University Press.

Arvaniti, Amalia. 2016. Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani. *Laboratory Phonology* 7(1). 6. http://doi.org/10.5334/labphon.14

Atterer, Michaela & Ladd, Robert D. 2004. On the phonetics and phonology of "segmental anchoring of F0: Evidence from German". *Journal of Phonetics* 32. 177–197.

Avesani, Cinzia & Bocci, Giuliano & Vayra, Mario & Zappoli, Alessandra (2015). Prosody and information status in Italian and German L2 intonation. In Chini, Maria (a cura di). *Il parlato in [italiano] L2: aspetti pragmatici e prosodici /[Italian] L2 Spoken Discourse: Pragmatic and Prosodic Aspects*, 93-116. Milano, Franco Angeli (Materiali Linguistici, 71).

Avesani, Cinzia & Vayra, Mario. 2004. Focus ristretto e focus contrastivo in italiano. In Albano Leoni, Federico & Cutugno, Franco & Pettorino, Massimo & Savy, Renata (a cura di), *Atti del Convegno nazionale Il Parlato Italiano*, F01, 1–20. Napoli: D'Auria Editore.

Avesani, Cinzia & Vayra, Mario. 2005. Accenting deaccenting and information structure in Italian dialogues. *SIGdial6–2005*: 19–24. https://www.isca-speech.org/archive_open/sigdial6/sgd6_019.html.

Avesani, Cinzia. 1995. ToBIt. Un sistema di trascrizione per l'intonazione italiana. In Lazzari, Giovanni (ed.), *Atti delle V Giornate di Studio del Gruppo di Fonetica Sperimentale,* 85–98. Trento: Servizio Editoria ITC.

Avesani, Cinzia. 1999. Quantificatori, negazione e costituenza sintattica. Costruzioni potenzialmente ambigue e il ruolo della prosodia. In Benincà, Paola & Mioni, Alberto & Vanelli, Laura (eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia*, 153–200. Roma: Bulzoni.

Avesani, Cinzia. 2003. La prosodia del focus contrastivo. Un accento particolare? In Marotta, Giovanna & Nocchi, Nadia (a cura di), *La coarticolazione. Atti delle XII Giornate di studio del Gruppo di Fonetica Sperimentale*, 157–167. Pisa: ETS.

Bard, Ellen G. & Aylett, Matthew. 1999. The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In Ohala, John J. & Hasegawa, Yoko & Ohala, Manjari & Granville, Daniel & Bailey, Ashlee C. (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco: University of California.

Baumann, Stefan & Riester, Arndt. 2013. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Elordieta, Gorka & Prieto, Pilar (eds.), *Prosody & Meaning*, 119–162. Berlin – New York: Mouton de Gruyter.

Beach, Ceryl. 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language* 30. 644–663.

Beckman, Mary & Ayers, Gail. 1997. *Guidelines for ToBI labelling*. Columbus: The Ohio State University Research Foundation.

Beckman, Mary & Edwards, Jane. 1990. Lengthening and shortening and the nature of prosodic constituency. In Kingston, John & Beckman, Mary (eds.), *Laboratory Phonology I*, 152–178. Cambridge: Cambridge University Press.

Beckman, Mary & Edwards, Jane. 1994. Articulatory evidence for differentiating stress categories. In Keating, Patricia (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form,* 7–33. Cambridge: Cambridge University Press.

Beckman, Mary & Hirschberg, Julia & Shattuck–Huffnagel, Stefanie. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, Sun-Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 10–54. Oxford: Oxford University Press.

Beckman, Mary & Pierrehumbert, Janet. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3. 255–310.

Beckman, Mary. 1996. The Parsing of Prosody. *Language and Cognitive Processes* 11(1/2). 17–68.

Birch, Stacy & Clifton, Charles. 1995. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 38(4). 365–391.

Birdwhistell, Ray L. 1952. *Introduction to Kinesics: An annotated System for the Analysis of Body Motion and Gesture*. Louisville: University of Louisville.

Birdwhistell, Ray L. 1970. *Kinesics and Context: Essays on Body–Motion Communication*. Philadelphia: University of Pennsylvania Press.

Bocci, Giuliano & Avesani, Cinzia, 2011. Phrasal Prominences Do Not Need Pitch Movements: Postfocal Phrasal Heads in Italian. In Cosi, Piero & De Mori, Renato & Di Fabbrizio, Giuseppe & Pieraccini, Roberto (eds.), *INTERSPEECH*-2011: 1357–1360.

Bocci, Giuliano & Avesani, Cinzia. 2015. Can the metrical structure of Italian motivate focus fronting? In Shlonsky, Ur (ed.), *Beyond Functional Sequence. The cartography of syntactic structures,* 23–41. Oxford: Oxford University Press.

Bocci, Giuliano. 2013. *The Syntax–Prosody Interface*, Amsterdam: John Benjamins.

Bögels, Sara & Schriefers, Herbert & Vonk, Wietske & Chwilla, Dorothee. 2011. Prosodic Breaks in Sentence Processing Investigated by Event–Related Potentials. *Language and Linguistics Compass* 5(7). 424–440.

Bolinger, Dwight. 1951. Intonation: Levels versus configurations. *Word* 7. 199–210.

Borràs–Comes, Juan & Prieto, Pilar. 2011. Seeing tunes. The role of visual gestures in tune interpretation. *Laboratory Phonology* 2(2). 355–380.

Braun, Bettina & Chen, Aoju. 2012. Now for something completely different: Anticipatory effects of intonation. In Niebuhr, Oliver (ed.), *Understanding Prosody: The Role of Context, Function and Communication*, 289–311. Berlin: De Gruyter.

Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech*, 49(4). 451–493.

Brown, Gillian. 1983. Prosodic structure and the given/new distinction. In Cutler, Anne & Ladd, Robert D. (eds.), *Prosody: Models and Measurement*, 67–77. BerlIn Springer.

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Büring, Daniel. 2016. *Intonation and Meaning*. Oxford: Oxford University Press.

Byrd, Dany & Saltzman, Elliot. 2003. The Elastic Phrase: Modelling the Dynamics of Boundary-Adjacent Lengthening. *Journal of Phonetics* 31. 149–180.

Carroll, Patrick & Slowiaczek, Maria. 1987. Modes and Modules: Multiple Pathways to the Language Processor. In Garfield, Jay L. (ed.), *Modularity in Knowledge Representation and Natural–language Understanding*, 221–247. Oxford: The MIT Press.

Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In Li, Charles (ed.), *Subject and Topic*, 27–55. New York, Academic Press.

Cho, Taehong & Keating, Patricia. 2001. Articulatory and Acoustic Studies on Domain–Initial Strengthening In Korean. *Journal of Phonetics* 29(2). 155–190.

Cooper, William & Sorensen, John. 1981. *Fundamental Frequency in Sentence Production*. Heidelberg: Springer.

Crespo–Sendra, Veronica & Kaland, Constantijn & Swerts, Mark & Prieto, Pilar. 2013. Perceiving Incredulity: The Role of Intonation and Facial Gestures. *Journal of Pragmatics* 47. 1–13.

Cresti, Emanuela & Moneglia, Massimo. 2018. The illocutionary basis of Information Structure. Language into Act Theory (L–AcT). In Adamou, Evangelia & Haude, Katharina & Vanhove, Martine (eds.), *Information structure in lesser described languages: Studies in prosody and syntax*, 359–401. Amsterdam: Benjamins.

Cresti, Emanuela. 2005. Per una nuova classificazione dell'illocuzione a partire da un corpus di parlato (LABLITA). In Burr, Elisabeth (ed.), *Tradizione e innovazione: il parlato. Atti del VI Convegno internazionale SILFI*, 233–246. Pisa: Cesati.

Cruttenden, Alan. 1993. The de–accenting and reaccenting of repeated lexical items. *Working Papers 41, Dept. of Linguistics and Phonetics, Lund, Sweden*.16–19.

Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press CUP.

D'Apolito, Sonia & Gili Fivela, Barbara. 2013. Acoustic and articulatory study of French sibilant clusters. In Galatà, Vincenzo (ed.), *Multimodalità e multilingualità: la sfida più avanzata della comunicazione orale*, 135–150. Roma: Bulzoni.

D'Apolito, Sonia & Gili Fivela, Barbara. 2018. Strategie nella produzione del parlato non nativo: l'inserzione vocalica in francese L2. *Studi Italiani di Linguistica Teorica e Applicata* XLVII (2). 269–292.

D'Imperio, Mariapaola & Gili Fivela, Barbara & Niebhur, Oliver. 2010. Alignment perception of high intonational plateaux in Italian and German. In *Proceedings of Speech Prosody 2010. Special Session on Shape, Scaling, and Alignment Effects in the Production and Perception of F0 Event*s. Chicago: ISCA Speech 100186:1–4.

D'Imperio, Mariapaola & House, David. 1997. Perception of questions and statements in Neapolitan Italian. In Botinis, Antonis (ed.), *Intonation: Theory, Models and Applications*, 251–254. Athens: Athanasopoulos.

D'Imperio, Mariapaola & Michelas, Amandine. 2014. Pitch scaling and the internal structuring of the Intonation Phrase in French. *Phonology* 31(1). 95–122.

D'Imperio, Mariapaola. 2000. The role of perception in defining tonal targets and their alignment. Columbus: Ohio State University. (Tesi di dottorato.)

D'Imperio, Mariapaola. 2002. Language–specific and universal constraints on tonal alignment: The nature of targets and anchors. In Bel, Bernard & Marlien, Isabelle (eds.), *Proceedings of the Speech Prosody 2002 Conference*, 101–106. Aix-en-Provence: Laboratoire Parole et Langage.

D'Imperio, Mariapaola & Cangemi, Francesco & Grice, Martine. 2016. *Introducing Advancing Prosodic Transcription. Laboratory Phonology* 7(1). 4. http://doi.org/10.5334/labphon.32.

Dahan, Delphine & Tanenhaus Michael & Chambers, Craig. 2002. Accent and reference resolution in spoken–language comprehension. *Journal of Memory and Language* 47. 292–314.

Eckman, Fred R. 2008. Typological markedness and second language phonology. In Hansen Edwards, Jette G. & Zampini, Mary L. (eds.), *Phonology and Second Language Acquisition*, 95–115. Philadelphia: Benjamins.

Edwards, Jane & Beckman, Mary & Fletcher Jane. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89(1). 369–382.

Efron, David. 1972 (1941). *Gesture, race and culture*. The Hague: Mouton.

Ekman, Paul & Friesen, Wallace V. & Hager, Joseph C. 2002. *The Facial Action Coding System* CD–ROM. Salt Lake City: Research Nexus.

Ekman, Paul. 1982. Methods for Measuring Facial Action. In Scherer, Klaus R. & Ekman, Paul (eds.), *Handbook of Methods in Nonverbal Behavior Research*, 45–90. Cambridge: Cambridge University Press.

Esposito, Anna & Esposito, Daniela & Refice, Mario & Savino, Michelina & Shattuck–Hufnagel, Stefanie. 2007. A Preliminary Investigation of the Relationships between Gestures and Prosody in Italian. In Esposito, Anna & Bratanić, Maja & Keller, Eric & Marinaro, Maria (eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*, 65–75. Amsterdam: IOS Press.

Esteve–Gibert, Nuria & Prieto, Pilar. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research* 56(3). 850–865.

Farnetani, Edda & Busà, Maria Grazia. 2004. Italian clusters in continuous speech. In *Proceedings of the 3rd International Conference on Spoken Language Processing,* 359–362. Yokohama: Japan.

Fougeron, Cécile & Keating, Patricia K. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101(6). 3728–3740.

Frota, Sonia & Prieto, Pilar. 2015. Intonation in Romance: Similarities and differences. In Frota, Sonia & Prieto, Pilar (eds.), *Intonation in Romance*, 140–197. Oxford: Oxford University Press.

Frota, Sonia & Vigario, Marina. 2018. Syntax-Phonology Interface. In Aronoff, Mark (ed.), *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.111.

Frota, Sonia. 2000. *Prosody and focus in European Portuguese: Phonological phrasing and intonation.* Abingdon: Routledge.

Frota, Sonia. 2017. Segment-to-Tone Association. In Cohn, Abigail & Fougeron, Cécile & Huffman, Marie (eds.), *The Oxford Handbook of Laboratory Phonology*, 265–275. Oxford: Oxford University Press.

Gili Fivela, Barbara & Avesani, Cinzia & Bocci, Giuliano & D'Imperio, Mariapaoala & Giordano, Rosa & Marotta, Giovanna & Savino, Michelina & Sorianello, Patrizia. 2015. Intonational phonology of the regional varieties of Italian. In Frota, Sonia & Prieto, Pilar (eds.), *Intonation in Romance*, 140–197. Oxford: Oxford University Press.

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

Gili Fivela, Barbara & Iraci, Massimiliano. 2017. Variation of intonation across Italy: The case of Palermo Italian. In Bertini, Chiara & Celata, Chiara & Lenoci, Giovanna & Meluzzi, Chiara & Riccci, Irene (a cura di), *Fattori sociali e biologici nella variazione fonetica – Social and Biological Factors in Speech Variation*, 169–190. Milano: Officinaventuno.

Gili Fivela, Barbara & Nicora, Francesca. 2018. Intonation in Liguria and Tuscany: checking for similarities across a traditional isogloss boundary. In Vietti, Alessandro & Spreafico, Lorenzo & Mereu, Daniela & Galatà, Vincenzo (a cura di), *Il parlato nel contesto naturale – Speech in the natural context*, 131–156. Milano: Officinaventuno.

Gili Fivela, Barbara & Avesani, Cinzia & Nicora, Francesca. submitted. Variation and contact in Italian intonation across the La Spezia-Rimini line. *Language and Speech*, Special Issue on *Language contact and speaker accommodation.*

Gili Fivela, Barbara. 2008. *The Phonetics and Phonology of Intonation: The case of Pisa Italian*. Alessandria: Edizioni dell'Orso.

Gili Fivela, Barbara. 2012. Meanings, shades of meanings and prototypes of intonational categories. In Gorka, Elordieta & Prieto, Pilar (eds.), *Prosody and meaning*, 197–237. Berlin: Mouton de Gruyter.

Gili Fivela, Barbara. 2015. L'integrazione di informazioni multimodali: prosodia ed espressioni del volto nella percezione del parlato. In Pistolesi, Elena & Pugliese, Rosa & Gili Fivela, Barbara (a cura di), *Parole, gesti, interpretazioni. Studi linguistici per Carla Bazzanella*, 107–127. Roma: Aracne.

Goldin–Meadow, Susan & Beilock, Sian L. 2010. Action's influence on thought: The case of gesture. *Perspectives on Psychological Science* 5(6). 664–674.

Goldin–Meadow, Susan. 2013. How our gestures help us learn. In Müller, Cornelia & Cienki, Alan & Fricke, Ellen & Ladewig, Silva H. & McNeill, David & Teßendorf, Sedinha (eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, 792–803. Berlin: Mouton De Gruyter.

Goldsmith, John. 1979. *Autosegmental phonology*. Garland: New York.

Grabe, Esther & Greg Kochanski & John Coleman. 2003. Quantitative modelling of intonational variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine*: 1–23. http://www.phon.ox.ac.uk/oxigen/Grabe_Kochanski_Coleman_2003.pdf

Grice, Martine & D'Imperio, Mariapaola & Savino, Michelina & Avesani, Cinzia. 2005. Strategies for intonation labelling across varieties of Italian. In Jun, Sun-Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 362–389. Oxford: Oxford University Press.

Grice, Martine. 1995. *The Intonation of Interrogation in Palermo Italian: Implications for Intonational Theory*. Tübingen: Niemeyer.

Grosz, Barbara & Sidner, Candice. 1986. Attentions, intentions and the structure of discourse. *Computational Linguistics* 12. 175–204.

Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrech: Foris.

Gussenhoven, Carlos. 2007. The phonology of intonation. In de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, 253–280. Cambridge, Cambridge University Press.

Hall, Nancy E. 2003. Gestures and segments: Vowel intrusion as overlap. Amherst: University of Massachusetts. (Tesi di dottorato.)

Hall, Nancy E. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429.

Hall, Nancy E. 2011. Vowel epenthesis. In van Oostendorp, Marc & Ewen, Colin J. & Hume, Elisabeth & Rice, Keren (eds.), *The Blackwell Companion to Phonology*, 1576–1596. Malden: Wiley–Blackwell.

Hirschberg, Julia & Avesani, Cinzia. 2000. Prosodic disambiguation in English and Italian. In Botinis, Antonis (ed.), *Intonation. Analysis, modelling and technology*, 87–96. Dordrecht: Kluwer Academic Publishers.

Hirst, Daniel & Di Cristo, Albert. 1998. A survey of intonation systems. In Hirst, Daniel & Di Cristo, Albert (eds.), *Intonation Systems: A Survey of Twenty Languages*, 1–44. Cambridge: Cambridge University Press.

House, David. 2002. Intonation and visual cues in the perception of interrogative mode in Swedish. In Hansen, John & Pellom, Bryan (eds.) *Proceedings of the 7th International Conference on Spoken Language Processing*, 1957–60. http://www.isca–speech.org/archive/icslp02.

Jun, Sun–Ah & Fougeron, Cécile. 1995. The accentual phrase and the prosodic structure of French. In Elenius, Kjell & Branderud, Peter (eds.) *Proceedings of the XIIIth International Congress of Phonetic Sciences*, volume 2, 722–725. Stockholm, Stockholm University.

Jun, Sun-Ah (ed.). 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press.

Keating, Patricia & Cho, Taehong & Fougeron, Cécile & Hsu Chai-shune. 2004. Domain–initial articulatory strengthening in four languages. In Local, ЖJohn & Ogden, Richard & Temple, Rosalind (eds.), *Phonetic interpretation. Papers in laboratory phonology VI*, 145–163. Cambridge: Cambridge University Press.

Kelly, Spencer D. & Church, Breckinridge R. 1997. Can children detect information conveyed through other children's nonverbal behaviors? *Cognition and Instruction* 15. 107–134.

Kendon, Adam. 1972. Some relationships between body motion and speech: an analysis of an example. In Siegman, Aron Wolfe & Pope, Benjamin (eds.), *Studies in Dyadic Communication*, 177–210. New York: Pergamon.

Kendon, Adam. 1980. Gesticulation and speech: two aspects of the process of utterance. In Kay, Mary Ritchey (ed.), *The Role of Nonverbal Communication*, 207–227. Berlin: De Gruyter.

Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge. Cambridge University Press.

Kjelgaard, Margaret M. & Speer, Shari R. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40. 153–194.

Kohler, Klaus. 2006. Paradigms in experimental prosodic analysis: From measurements to function. In Sudhoff, Stefan & Lenertové, Denisa & Meyer, Roland & Pappert, Sandra & Augurzky, Petra & Mleinek, Ina & Richter, Nicole & Schliesser, Johannes (eds.), *Methods in Empirical Prosody Research*, 123–152. Berlin: De Gruyter.

Krahmer, Emiel & Swerts, Mark. 2005. How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech* 48(1). 29–54.

Krifka, Manfred. 2007. Basic notions of information structure. In Fery, Caroline & Fanselow, Gisbert & Krifka, Manfred (eds.), *The Notions of Information Structure. Working Papers of the SFB 632 Volume 6*, 13–55. Potsdam: Universitätsverlag Potsdam.

Krivokapic, Jelena & Tiede, Marc & Tyrone, Martha. 2015. A kinematic analysis of prosodic structure in speech and manual gestures. The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, paper number 658. Glasgow: the University of Glasgow. https://www.internationalphoneticassociation.org/icphs–proceedings/ICPhS2015/Papers/ICPHS0658.pdf. 1-5.

Ladd, Robert & Faulkner, Dan & Faulkner, Hanneke & Schepman, Astrid. 1999. Constant segmental anchoring of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106(3). 1543–1554.

Ladd, Robert. 1996/2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Leben, William. 1973. *Suprasegmental phonology*. Cambridge: Massachusetts Institute of Technology. (Tesi di dottorato.)

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: The MIT Press.

Liberman, Mark, & Pierrehumbert, Janet. 1984. Intonational Invariance under Changes in Pitch Range and Length. In Aronoff, Mark & Oehrle, Richard & Kelley, Frances & Stephens, Bonnie Wilker (eds.), *Language Sound Structure*, 157–234. Cambridge: The MIT Press.

Liberman, Mark. 1975. The intonational system of English. Cambridge: Massachusetts Institute of Technology. (Tesi di dottorato.)

Libermann, Mark & Prince, Alan. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8. 249–336.

Loehr, Daniel. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3. 71–89.

Massaro, David. 1989. Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology* 21(3). 398–421.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32. 543–563.

Moneglia, Massimo. 2006. Units of Analysis of Spontaneous Speech and Speech Variation in a Cross-linguistic Perspective. In Kawaguchi, Yuji & Zaima, Susumu & Takagaki, Toshihiro (eds.), *Spoken Language Corpus and Linguistics Informatics*, 153–179. Amsterdam: Benjamins.

Morsella, Ezequiel & Krauss, Robert. 2004. The role of gestures in spatial working memory and speech. *The American Journal of Psychology* 117(3). 411–424.

Muliačić, Žarco. 1973. *Fonologia della lingua italiana*. Bologna: Il Mulino.

Nespor, Marina & Vogel, Irene. 1986/2007. *Prosodic phonology*. Dordrecht: Foris.

Niebuhr, Oliver & Lancia, Leonardo & Meunier, Christine. 2008. On place assimilation in French sibilant sequences. In Sock, Rudolph & Fuchs, Susanne & Laprie, Yves (eds.), *Proceedings of the VII International Seminar on Speech Production*, 221–224. Le Chesnay-Rocquencourt: INRIA.

Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories – The interplay of synchronization, shape, and height. *Phonetica* 64(2–3). 174–193.

Pierrehumbert, Janet & Beckman, Mary. 1988. *Japanese tone structure*. Cambridge, The MIT Press.

Pierrehumbert, Janet & Hirschberg, Julia. 1990. The meaning of intonational contours in the interpretation of discourse. In Cohen, Phillip R. & Morgan, Jerry & Pollack, Martha E. (eds.), *Intentions in communication*, 271–310. Cambridge: The MIT Press.

Pierrehumbert, Janet & Talkin, David. 1992. Lenition of /h/ and Glottal Stop. In Docherty, Gerard & Ladd, Robert D. (eds.), *Gesture, Segment, Prosody*, 90–117. Cambridge: Cambridge University Press.

Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English intonation*. Cambridge: Massachusetts Institute of Technology. (Tesi di dottorato.)

Pierrehumbert, Janet. 1999. Prosody and intonation. In Wilson, Robert & Keil, Frank (eds.), *The MIT Encyclopedia of the Cognitive Sciences,* 679–682. Cambrige: The MIT Press.

Pike, Kenneth. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

Portes, Cristel & Beyssade, Claire. 2012. Is intonational meaning compositional? *Verbum* XXXIV. 3–27.

Price, Patti Jo & Ostendorf, Mary &, Shuttuck–Hufnagel, Stefanie & Fong, Cynthia. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90(6). 2956–2970.

Prieto, Pilar. 2014. The intonational phonology of Catalan. In Jun, Sun-Ah (ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*, 43–80. Oxford. Oxford University Press.

Prieto, Pilar. 2015. Intonational meaning. *WIREs Cognitive Science* 6. 371–381.

Prince, Ellen. 1981. Toward a taxonomy of given–new information. In Cole, Peter (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.

Prince, Ellen. 1992. The ZPG letter: subject, definiteness and information status. In Thompson, Sandra & Mann, William (eds.), *Discourse description*, 295–325. Amsterdam: Elsevier.

Pynte, Joël & Prieur, Bénédicte. 1996. Prosodic breaks and attachment decisions in sentence processing. *Language and Cognitive Processes* 11. 165–192.

Röhr, Christine & Baumann, Stefan. 2010. Prosodic marking of information status in German. In *Proceedings of the 5th International Conference on Speech Prosody*: 1–4. http://www.isca–speech.org/sp2010/ITRW

Rooth, Maths. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75–116.

Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42(1). 23–48.

Schafer, Ami. 1997. Prosodic parsing: The role of prosody in sentence comprehension. University of Massachusetts. (Tesi di dottorato).

Schumacher, Petra & Baumann, Stephan. 2011. (De-)Accentuation and the Processing of Information Status: Evidence from Event–Related Brain Potentials. *Language and Speech* 55(3). 361–381.

Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In Fretheim, Thorse (ed.), *Nordic Prosody II*, Trondheim, TAPIR. 111-140.

Selkirk, Elisabeth. 1984. *Phonology and Syntax*. Cambridge, The MIT Press.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371–405.

Selkirk, Elisabeth. 1995. Sentence Prosoedy: Intonation, Stress and Phrasing. In Goldsmith, John, (ed.), *The Handbook of Phonological Theory*, 550–569. Cambridge: Blackwell.

Selkirk, Elisabeth. 2005. Comments on Intonational Phrasing in English. In Frota, Sonia & Vigario, Marina & Freitas, Maria Joao (eds.), *Prosodies*, 11-58. Berlin: Mouton de Gruyter.

Selkirk, Elisabeth. 2008. Contrastive Focus, Givenness and the Unmarked Status of Discourse-New. *Acta Linguistica Hungarica* 55(3). 331–346.

Selkirk, Elisabeth. 2011. The Syntax–Phonology Interface. In Goldsmith, John & Riggle, Jason & Yu, Alan (eds.), *The Handbook of Phonological Theory*, Second Edition, 435–484. Malden, Blackwell.

Shattuck-Hufnagel, Stefanie & Ren, Ada. 2018. The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech. *Frontiers in Psychology* 9. 1514. 205-2017.https://www.ncbi.nlm.nih.gov / pmc/articles/PMC6137092/

Shattuck-Hufnagel, Stefanie & Turk, Alice. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Shattuck-Hufnagel, Stefanie & Yasinnik, Yelena & Veilleux, Nanette & Renwick, Margaret. 2007. Method for Studying the Time Alignment of Gestures and Prosody in American English: 'Hits' and Pitch Accents in Academic-Lecture-Style Speech. In Esposito, Anna & Bratanić, Maja & Keller, Eric & Marinaro, Maria (eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*, 34–44. Amsterdam: IOS Press.

Steinhauer, Karsten & Friederici, Angela. 2001. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research* 30(3): 191–196.

Steinhauer, Karsten & Friederici, Angela. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience* 2(2): 191-196

Stella, Antonio. 2013. Coordinazione interarticolatoria nella produzione dell'intonazione del tedesco come lingua straniera. In Galatà, Vincenzo (a cura di), *Multimodalità e multilingualità: la sfida più avanzata della comunicazione orale*, 411–426. Roma: Bulzoni.

Stirling, Lesley & Wales, Roger. 1996. Does prosody support or direct sentence parsing? *Language and Cognitive Processes* 11(1-2). 193–212.

Swerts, Mark & Krahmer, Emiel & Avesani, Cinzia. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics* 30(4). 629–654.

't Hart, Johan & Collier, René & Cohen, Antonie. 1990. *A Perceptual Study of Intonation: An Experimental–Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Terken Jack & Hirschberg, Julia. 1994. Deaccentuation of words representing given information: effects of persistence of grammatical function and surface position. *Language and Speech* 37(2): 125–145.

Truckenbrodt, Hubert. 1995. *Phonological phrases: Their relation to syntax, focus and prominence*. Massachusetts Institute of Technology. (Tesi di dottorato.)

Truckenbrodt, Hubert. 2002. Upstep and embedded register level. *Phonology* 19. 77–120.

Truckenbrodt, Hubert. 2007. The syntax phonology interface. In de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, 435-456. Cambridge: Cambridge University Press

Truckenbrodt, Hubert. 2012. Semantics of intonation. In Maienborn, Claudia & von Heusinger, Klaus & Portner, Paul (eds*.), Semantics: An International Handbook of Natural Language Meaning*, 2039–2969. Berlin: Mouton de Gruyter.

Turk, Alice & Shattuck-Hufnagel, Stefanie. 2007. Phrase–final lengthening in American English. *Journal of Phonetics* 35. 445–472.

Vanrell, Maria del Mar. 2006. A scaling contrast in Majorcan Catalan interrogatives. In Hoffmann & H. Mixdorff (eds), *Proceedings of Speech Prosody 2006*, Dresden, Germany, May 2-5, 2006, 807–810.

Vanrell, Maria del Mar. 2007. A tonal scaling contrast in Majorcan Catalan interrogatives. *Journal of Portuguese Linguistics* 6(1). 147–178.

Verluyten, Sylavin-Paul. 1982. *Investigation on French Prosodics and Metrics*, Université d'Anvers. (Tesi di dottorato.)

Wagner, Petra & Malisz, Zofia & Kopp, Stefan. 2014. Gesture and speech in interaction. *Speech Communication* 57. 209–232.

Wightman, Colin & Shattuck-Hufnagel, Stefanie & Ostendorf, Mary & Price, Patti J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717.

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communicati*on 46(3-4). 220–251.

Sandra Benazzo - Marzena Watorek
(Université Paris 8 & UMR 7023)

Transcription de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques

1. *Introduction*

L'étude de la production langagière orale nous pose face aux problèmes de sa transcription à l'écrit, transcription qui devient nécessaire dès lors qu'on veut observer et revenir sur un objet qui est par définition 'évanescent'. La reproduction à l'écrit de tous les phénomènes typiques de l'oral (volume, débit, intonation, etc.), si jamais elle était possible, rendrait la transcription illisible. La transcription de l'oral implique par conséquent une série de choix, dont une partie est déterminée par l'objet d'étude, à savoir le phénomène de langue à analyser pour lequel le corpus a été conçu, et par la nécessité de trouver un compromis entre lisibilité et fiabilité de la parole transcrite. Ainsi, le degré de finesse de la transcription sera différent si l'on s'intéresse à des phénomènes de nature phonologique (par ex. la liaison) ou bien à des phénomènes lexicaux ou encore morphosyntaxiques (Baude 2006).

Si ces questions méthodologiques concernent la transcription du parler de tout locuteur, ils se posent de manière plus aiguë lorsque l'objet d'étude concerne (a) la production de sujets qui sont en cours d'apprendre la langue, que ce soit la langue maternelle (L1) ou une langue seconde (L2), et (b) lorsque la langue en question présente un grand écart entre oral et écrit, comme c'est le cas en français (cf. Blanche Benveniste 2000). Pour illustrer le type de difficultés rencontrées en français L2, nous pouvons citer l'exemple suivant mentionné par cette auteure :

(1) [mwa vule pa partir] *moi* (*voulais ? vouler ? voulez ?) pas partir*
(Blanche Benveniste 2000 : 29)

La reproduction à l'écrit de certains items de cet énoncé est relativement univoque, pour d'autres elle l'est moins. En l'occurrence, la transcription orthographique d'une forme verbale avec un [e] final – comme [vule] en (1) – oblige à choisir entre différentes désinences qui sont homophones alors qu'elles portent des informations grammaticales (accord, temps, aspect, mode) différentes : imparfait 1ère p.s. *-ais*, infinitif idiosyncrasique *-er*, présent 2e p.pl. *-ez*, etc. L'acte de transcrire implique un premier travail d'interprétation du discours en L2 qui s'avère être particulièrement délicat puisqu'il peut influencer l'analyse ultérieure du chercheur. En effet, le choix d'une forme au détriment des autres comporte le risque de sur-estimer (ou sous-estimer) les connaissances de l'apprenant à ce moment de son apprentissage.

Notre contribution est centrée justement sur les enjeux théoriques liés aux choix de transcription de la morphologie verbale dans les corpus oraux en français L2. En particulier, nous allons reprendre les critères méthodologiques appliqués dans un vaste corpus de productions d'apprenants débutant en L2 en différentes langues (projet ESF), afin d'illustrer pourquoi le choix de transcrire selon l'orthographe standard de la langue cible (LC) peut s'avérer dangereux, du moins en ce qui concerne l'étude de la morphologie verbale en français L2, que ce soit dans le but d'en étudier le développement ou de déterminer à quel stade se trouve un apprenant de cette LC.

Il est utile de souligner que les choix de transcription varient également en fonction de l'orientation théorique adoptée, comme le faisait déjà remarquer Ochs (1979). A ce propos, nous nous situons ici dans l'approche fonctionnaliste : notre contribution sera ainsi limitée à la discussion des enjeux pertinents pour les analyses/études qui adoptent cette orientation théorique.

L'article est structuré en deux parties. La première partie sera consacrée, d'une part, à la présentation du projet ESF (structure, objectifs, approche théorique) et, d'autre part, à la description des spécificités du français, qui expliquent les choix méthodologiques sous-jacents les transcriptions en français L2 appliqués dans ce projet.

Dans la seconde partie, nous allons montrer les enjeux théoriques de ces choix à travers la présentation de deux cas de figure, notamment l'acquisition de la morphologie verbale et l'acquisition de la négation. Toutes les études mentionnées ici se situent dans l'approche fonctionnaliste, dite approche des lectes d'apprenants (Klein & Perdue 1997).

2. *Lectes d'apprenants en L2 et corpus oraux*

L'étude de l'acquisition d'une langue seconde est un domaine de recherche relativement récent, qui a été profondément marqué par l'émergence de la notion d'*interlangue* : en effet, depuis les travaux pionniers de Corder (1967) et Selinker (1972), qui postulaient l'existence d'un système 'idiosyncrasique' propre à l'apprenant, de nombreuses études empiriques ont démontré que l'apprenant d'une L2 développe un système linguistique dont les règles de fonctionnement sont, du moins en partie, indépendantes de celles des langues en contact (la L1, ainsi que d'autres langues connues, et la langue cible).

Le projet ESF (« Adult Second Language Acquisition » cf. Perdue 1993), caractérisé par l'étude comparative de parcours acquisitionnels dans plusieurs L2 (cf. description plus détaillée en 2.1), a joué un rôle crucial dans la description de principes développementaux communs, partagés par des apprenants de langues secondes (LS)/LC différentes, ce qui a conduit à la théorisation de la variété linguistique produite par des apprenants L2 sous la forme de l'approche des lectes d'apprenants (*Learner Variety Approach*, Klein & Perdue 1997; Perdue 1993, etc.).

Les recherches menées dans le cadre du projet ESF, qui ont servi à identifier les stades acquisitionnels présentés en 2.1., sont basées sur des productions orales d'apprenants débutants. Ainsi, la réflexion sur les critères de transcription dans différentes langues cibles a été fondamentale et étroitement liée aux principes de l'approche *Learner Variety*. Selon cette approche, les apprenants adultes d'une L2 possèdent la compétence de communication en tant que locuteurs natifs d'au moins une langue et ont des besoins de communication relativement complexes qui constituent le « moteur de l'acquisition » (*factors 'pushing' L2 acquisition*). L'apprenant adulte exposé à une nouvelle langue (langue cible) élabore des hypothèses sur son fonctionnement à partir des données linguistiques de la L2 auxquelles il a accès. Les facteurs structurels provenant de l'input de la langue cible « façonnent » ainsi l'acquisition (*factors 'shaping' L2 acquisition*) (Watorek & Perdue 2005). Le processus acquisitionnel est donc envisagé comme une interaction entre des facteurs communicatifs (relativement similaires) et des facteurs structurels (dépendant en partie des spécificités de la LC). Le lecte d'apprenant reflète le travail cognitif fourni par l'apprenant dans la construction du nouveau système linguistique. Ce dernier est grammatical dans le sens où, à côté d'éléments

variables, il présente des traits systématiques qui permettent de décrire les règles sous-jacentes à son fonctionnement à un moment donné de l'apprentissage. Les éléments sujets à variation reflètent l'évolution des hypothèses de l'apprenant au fur et à mesure qu'il prend conscience d'autres aspects de la langue à apprendre. Les lectes d'apprenants sont ainsi dotés d'une dynamique que le chercheur doit décrire pour pouvoir comprendre le processus d'acquisition.

Les choix de transcription appliqués dans le projet ESF tiennent compte de cette approche théorique envisageant les lectes des apprenants comme des systèmes à part entière, par définition dépourvus d'erreurs, qui seraient à décrire comme une langue inconnue :

> « Learner varieties are not imperfect imitations of a 'real language' – the target language – but systems in their own right, error-free by definition » (Klein & Perdue 1997: 308)

Pour ce faire, deux dangers sont à éviter. Le premier consiste à juger le lecte de l'apprenant en termes de déviations ou d'erreurs par rapport à la LC (cf. la notion de '*comparative fallacy*', Bley-Vroman 1983), à savoir de considérer sa façon de s'exprimer comme une imitation plus ou moins réussie de la langue à apprendre ; le deuxième concerne la tentation d'interpréter la production en L2 en termes de catégories propres à la LC (cf. la notion de '*proximity fallacy'*, Perdue 1993) et attribuer ainsi le statut et fonctions des noms, verbes, auxiliaires, etc. de la LC aux unités linguistiques produites par l'apprenant sur la base de leur ressemblance formelle.

On est ici face au problème bien connu des chercheurs en acquisition de L2 : comment transcrire les productions des apprenants de manière lisible, sans en donner pour autant une image faussée par les catégories de la LC ?

Même si toute transcription implique inévitablement un travail d'interprétation, l'idée est de fournir une reproduction des segments ambigus en L2 qui serait relativement neutre, afin de ne pas sur-estimer ou sous-estimer le système linguistique que l'apprenant est en train de construire. C'est en effet l'analyse successive de ses productions, établissant des liens entre les formes produites en L2 (niveau phrastique) et leurs fonctions en contexte d'énonciation (niveau discursif), qui permet de mieux interpréter la valeur de ces items chez l'apprenant au moment de l'enregistrement et la dynamique de développement de son lecte.

2.1 Le projet ESF et les stades initiaux de l'acquisition

Le projet européen « Adult Language Acquisition » (Perdue 1993), connu aussi comme projet ESF puisque financé par la Fondation Européenne de la Science (*European Science Foundation*), s'est donné comme objectif de décrire le processus d'acquisition d'une L2 à travers les productions d'apprenants adultes de L2 en milieu naturel. Il s'agit d'une population de migrants, sans guidage systématique lié à l'enseignement, qui ont été observés de manière longitudinale. Le croisement des langues sources et cibles (cf. schéma ci-dessous) a été conçu de manière à pouvoir comparer l'acquisition de deux langues cibles différentes par des apprenants qui ont la même langue maternelle et, inversement, l'acquisition de la même langue cible par des apprenants qui ont des langues maternelles différentes. Ainsi, ce projet se caractérise par une dimension translinguistique forte.

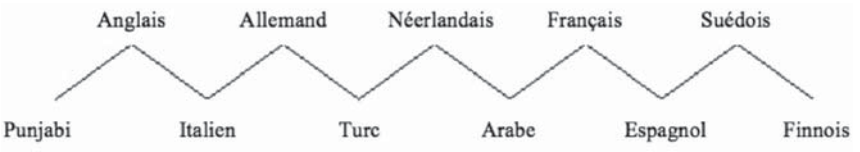


Schéma 1. *Combinaisons de LS (en bas) et de LC (en haut) dans le projet ESF.*

Dans ce projet, les apprenants ont été suivis et enregistrés pendant une période de 30 mois dans différents pays européens. Les résultats des travaux translinguistiques sont comparables car la même méthodologie de recueil des données et d'analyse a été adoptée. Les apprenants ont réalisé des tâches communicatives diverses comme une description d'affiche, un récit de fiction et d'expériences personnelles, des indications d'itinéraires, des indications scéniques, etc[1].

Un des résultats les plus robustes de ce projet a été l'explication du processus d'acquisition de L2 en termes de développement graduel du lecte de l'apprenant indépendamment des spécificités des langues en présence, sources et cibles, ce qui n'exclut pas que ces spécificités interviennent en dehors du développement partagé.

[1] L'ensemble du corpus est disponible à l'adresse suivante: https://archive.mpi.nl/tla/islandora/object/lat%3A1839_00_0000_0000_0004_CCAC_E

En ce qui concerne le développement partagé, le premier stade (variété prébasique) comprend un répertoire restreint, composé surtout d'éléments qui ressemblent aux noms, adverbes, adjectifs (et parfois prépositions) de la langue cible (cf. ex. 2). En réalité, ces items véhiculent essentiellement des valeurs lexicales, alors qu'ils sont dépourvus d'informations grammaticales (genre, nombre, etc.). Par ailleurs, il est parfois difficile de catégoriser précisément leur statut. A titre d'exemple, un énoncé tel que « de man + ausgang » (litt. l'homme + sortie) en allemand L2, bien que compréhensible en contexte, peut se reconstruire de différentes manières (l'homme se dirige vers la sortie ? sort ? est sorti ? veut sortir ?). De manière similaire, l'item « cooking », présent en (2), est employé par le même apprenant dans différents contextes ayant comme référent le lieu 'cuisine', l'action de cuisiner ou encore la nourriture (cf. Perdue 1995). Etant donné l'absence de formes verbales clairement identifiables, ce stade se caractérise par une organisation nominale de l'énoncé. La relation entre les items qui le composent ne peut être comprise que dans des contextes discursifs précis.

Variété prébasique[2]

(2) BE : **y*après [a al + a] la femme *al* camion de la police*
(et après la femme (monte) dans le camion de la police)

AN : *every people happy in the cooking + tea + biscuit*
(tous les gens (sont) contents dans la cuisine/de cuisiner ?
+ thé + biscuit)

Le passage graduel vers le stade suivant se caractérise par l'acquisition de lexèmes verbaux qui explicitent les relations entre les constituants nominaux. Les schémas des énoncés attestés sont très simples et ne contiennent que des verbes sans morphologie fonctionnelle et leurs actants : l'énoncé est ainsi structuré autour d'un verbe non conjugué, comme en (3).

[2] Dans les extraits des apprenants, les conventions suivantes sont adoptées : les deux premières lettres font référence à l'apprenant (en l'occurrence : BE = Berta ; AN = Andrea) ; + indique une pause non remplie ; les astérisques *…* entourent des segments produits dans la langue maternelle de l'apprenant et les crochets [..] la transcription phonétique de segments ambigus. Une glose entre parenthèses signale l'interprétation contextuelle de l'énoncé produit en L2.

Les constituants suivent des ordres déterminés par un petit nombre de principes : un principe pragmatique concernant la structure informationelle de l'énoncé (focus en dernier) et un principe sémantique, selon lequel l'actant le plus agentif se trouve en position pré-verbale (agent en premier). Ce stade d'acquisition, appelé « variété de base » (cf. Klein & Perdue 1997), marque un moment charnière dans l'acquisition de L2 dans la mesure où, à ce stade, l'apprenant devient autonome dans sa production.

Variété de base

(3) (Berta à propos de ses deux filles)
BE: [la du fil] [se] à lycée
(les deux filles sont/vont au lycée)

(récit de fiction *Calling 999*)
AN: *another woman help for ring* (…)
(une autre femme aide pour téléphone)
AN: *after + come back the brigade fire*
(après + arrivent les pompiers)

Le stade suivant (Variété postbasique) se caractérise par la complexification syntaxique du discours en L2 (émergence de la subordination) et, surtout, par le développement progressif de la morphologie verbale : l'énoncé est ainsi structuré autour d'un verbe qui commence à être fonctionnellement fléchi.

Variété postbasique initiale

(4) BE: jamais je [swi perdu] dans le métro
(jamais je ne me suis perdue dans le métro)

AN: he has finished the work
(il a fini le travail)
LA: when I was young + I had a job in a shop
(quand j'étais jeune + j'avais un travail dans un magasin)

Cependant, ce stade n'est pas atteint par tous les apprenants observés tandis que celui de la structuration à verbe non conjugué l'est : un tiers des apprenants analysés dans le projet ESF fossilise même à ce niveau.

Au-delà de cette étape le développement du lecte d'apprenant se diversifie tant au niveau individuel qu'au niveau de l'influence des spécificités des langues sources et cibles.

En résumé, on peut voir que l'évolution des stades initiaux implique de manière cruciale la catégorie du verbe (émergence des formes verbales dans le passage de la variété prébasique à la variété de base) et le développement d'une morphologie verbale fonctionnelle (transition entre la variété de base et la variété postbasique).

Afin de limiter le danger de la surinterprétation, les chercheurs du projet ESF ont opté pour un certain nombre de principes communs de transcription, qui ont été toutefois adaptés par chacune des équipes des cinq pays participant au projet.

« During the early phase of the project, the need was for all-purpose, theory-neutral transcriptions, serving as general basis for different types of more specific analysis in the different research areas of the project (...). Transcription of the verbal material was a compromise between an orthographic and a phonetic transcription reflecting systematic deviations from the standard TL realization of tokens. How exactly this framework was filled differed from team to team » (Feldweg 1993: 109).

La variation à laquelle Feldweg fait référence en termes d'*équipes* s'avère être plutôt une différence en termes de *langue cible*. Pour illustrer l'impact de cette dernière sur la manière de transcrire, nous proposons de comparer deux extraits de transcriptions, dont l'un en anglais L2 (apprenant italophone Santo) et l'autre en français L2 (apprenante hispanophone Berta), qui correspondent à des conversations libres entre un enquêteur du projet ESF (LN dans les extraits) et l'apprenant.

(5) *LN: so you didn't have any holiday ...*
me for holiday er no september
because er ++ er *se* I go in september for holiday
no possible christmas
you understand?
allora I no like london christmas…
er last christmas in london
and next + in my country…
when you when holiday you?
LN I've just had a holiday
and when you going another one?

(6) *LN tu étais partie en vacances + tu peux me raconter un peu ça? qu'est ce que tu as fait là bas?*
moi + [ale] + [a fe] de ski
[se] très très très dur * por por primera * fois
[nEpa] possible [ZE mɔ̃te] *sobre* les [eski]
LN et euh tu avais déjà fait du ski ?
jamais jamais + [sE] la première fois…
LN et alors ? ça s'est bien passé ces vacances ?
oui oui + [solamɑ̃ke] + + [nEpa] de chance * por * gladys
eh + la jambe + la + [se kase] *en* [do] fois
LN comment elle a fait ça ? + comment + c'est arrivé ?
le ski elle + elle [mɔ̃t] + * en * la montagne
y après + *se desliza* seule *y + y* [el tɔ̃b / tɔ̃be]

Les deux extraits présentent les traits typiques de la variété de base, dont l'organisation des énoncés autour d'une forme verbale qui est encore dépourvue de valeur fonctionnelles (soit de marques de temps, mode, aspect et/ou accord avec le sujet). Chez Santo (et tous les apprenants italophones du projet, à ce stade) la plupart des verbes apparaissent sous la forme de la racine verbale (*go, like*) dans tout contexte temporo-aspectuel ; ils alternent parfois avec la forme Ving (comme *go – going*), mais de manière aléatoire, à savoir la variation formelle ne marque pas l'aspect imperfectif. Chez Berta (et les apprenants du français en général) les verbes lexicaux sont produits souvent avec un [e] final, qui alterne parfois avec la racine verbale ([*mɔ̃te*] – [*mɔ̃t*]).

Dans l'extrait on note aussi la forme invariable [*se*], correspondant à la structure *c'est*, mais ne présentant aucune variation formelle de personne, temps, etc. (cf. section sur la négation 3.2). Un autre trait typique de ce stade est l'absence des modaux : à ce propos on peut remarquer que, dans les deux cas, la valeur modale du verbe *pouvoir* est exprimée de manière lexicale grâce à l'adjectif 'possible' et à sa négation : *not possible* en (5) et [*nEpa*] *possible* en (6). Par ailleurs, la négation est clairement préverbale en anglais (*I no like London Christmas*), alors qu'on peut avoir plus de doutes sur son positionnement en français, ou mieux sur la présence d'une forme verbale dans des séquences telles que [nEpa] *de chance* (ce point sera développé dans la section 3.2 suivante).

En revanche, les deux extraits diffèrent de manière remarquable en ce qui concerne le recours à la transcription orthographique vs. phonétique : toutes les formes verbales de la production en français L2 sont transcrites

en symboles phonétiques (bien que simplifiés), alors qu'elles ne le sont pas dans l'extrait homologue en anglais L2. Cette différence est dû à l'ambiguïté phonologique inhérente aux formes verbales produites en français : la transcription orthographique obligerait le transcripteur du français à choisir entre des désinences bien précises, porteuses d'informations grammaticales distinctes. En effet, le français oral présente beaucoup de formes homophones et hétérographes, comme nous le verrons dans la section suivante.

2.2 Français écrit / français parlé

Le français présente un écart considérable entre l'écrit et l'oral, qui concerne non seulement la faible correspondance graphèmes / phonèmes[3], mais aussi le fonctionnement des marques grammaticales ainsi que la structuration de la phrase et du texte (cf. Blanche Benveniste 2000 ; Riegel *et al.* 2009). Aborder l'oral à partir de l'écrit donnerait une image faussée de son fonctionnement, puisqu'une bonne partie de la morphologie écrite n'est pas réalisée à l'oral (cf. la notion de morphologie silencieuse : Agren 2008).

Dans ce qui suit nous nous limitons à attirer l'attention sur la variation dans le nombre de marques grammaticales et leur distribution qui caractérise les deux modalités expressives. Le français est, en effet, une langue qui présente une morphologie 'riche' à l'écrit mais relativement plus 'pauvre' à l'oral.

La phrase reportée en (7) illustre cette différence pour ce qui est des marques du pluriel (cf. Riegel *et al.* 2009: 33). Dans la version écrite, le pluriel est marqué sur 3 items (le déterminant *les*, le nom *enfant* et le verbe

[3] La non correspondance phonème /graphèmes est reconduite à trois cas de figure (cf. Riegel *et al.* 2009: 31-33) : (a) nombre de phonèmes différant du nombre des lettres (par exemple, graphème complexe équivalant à un phonème unique, tel que *eau* = [o], ou graphème unique correspondant à divers phonèmes, tel que x = [ks] dans *taxi* mais [gz] dans *exact*) ; (b) certaines lettres (dites 'muettes') ne correspondent à aucun phonème (par ex. *-t* final dans *doucement*, de *-nt* du pluriel dans *aiment*, etc.) ; (c) des ressemblances dans un système correspondent à des différences dans l'autre : ainsi, le phonème [s] correspond à des graphèmes aussi variés que 's' (*son*), 'ss' (*poisson*), 'c' (*cette*), 'x' (*soixante*), 'ç' (*ça*). En français, il existe de très nombreux homophones se différenciant uniquement par la graphie, tel que [vɛʁ], qui correspond à l'objet *verre* ('un verre à vin') ou la matière dont est composé un objet ('un vase en verre'), l'animal *ver* ('un ver de terre'), l'adjectif *vert* ('un T-shirt vert'), la préposition *vers* ('vers la gare'), etc.

jouer), alors que la version orale est beaucoup plus économique : il n'y a qu'une seule marque perceptible, notamment au niveau de l'article (qui, dans ce cas, donne lieu à la liaison [lez]).

(7) les enfants jouent dans la cour [lezɑ̃fɑ̃ʒudɑ̃lakuʁ]

L'écart est plus étendu en ce qui concerne l'ensemble de la morphologie verbale. Prenons à titre d'exemple la conjugaison au présent de l'indicatif de *parler*, verbe régulier du premier groupe (Tableau 1) : l'écrit présente 5 marques distinctes, alors que l'oral se contente de 3, le radical /paʁl/ pouvant correspondre à la 1ère, 2e et 3e personne du singulier ainsi qu'à la 3e personne du pluriel.

		ECRIT	ORAL
1ère p. s.	je	parl-e	[paʁl]
2e p. s.	tu	parl-es	[paʁl]
3e p. s.	il / elle	parl-e	[paʁl]
1ère p. pl.	nous	parl-ons	[paʁlɔ̃]
2e p. pl.	vous	parl-ez	[paʁle]
3e p. pl.	ils/elles	parl-en	[paʁl]

Tableau 1. *Présent du verbe* parler.

Par ailleurs, étant donné qu'à l'oral 'nous' (1ère p. pl.) est souvent remplacé par 'on', le nombre de désinences à l'oral peut se réduire à 2, [paʁl]-[paʁle] (cf. Riegel *et al.* 2009: 34).

Plus en général, le système verbal est caractérisé par une opposition entre formes simples – présent, imparfait (*je parlais* [ʒəpaʁlɛ]), futur (*je parlerai* [ʒəpaʁl(ə)re], subjonctif (*que je parle* [ʒəpaʁl]), conditionnel (*je parlerais* [ʒəpaʁləʁɛ]) – et formes composées grâce à l'auxiliaire *être* ou *avoir,* telles que le passé composé (*j'ai parlé* [ʒepaʁle] / *je suis venu* [ʒəsɥivəny]) et le plus-que-parfait (*j'avais parlé* [ʒavɛpaʁle] */ j'étais venu* [ʒetɛvəny]), etc. (cf. Riegel *et al.* 2009 pour un tableau plus exhaustif).

La prise en compte de l'ensemble des temps verbaux (simples et composés) conduit Noyau *et al.* (1995) à souligner l'opacité globale qui caractérise le système verbal temporo-aspectuel en français du fait que,

dans la plupart des cas, il n'y a pas de relation directe entre un morphème et une valeur de mode / temps / personne / nombre.

On peut constater cette opacité dans l'aire des suffixes verbaux. En reprenant l'exemple du verbe *parler*, la désinence [e] peut correspondre à l'infinitif, au participe passé, ainsi qu'à la 2e personne au pluriel du présent.

(8) [paʁle] → Infinitif: 'parl-er'
→ Participe passé: 'parl-é'
→ Présent 2e p.s. / impératif : 'parl-ez'

Le suffixe [ɛ] marque à son tour plusieurs personnes de l'imparfait (1e, 2e, 3e p.s. ainsi que 3e p.pl.).

(9) [paʁlɛ] → Imparfait: *je parlais, tu parlais, il/elle parlait, ils/elles parlaient*

Cependant, la distinction [e] vs. [ɛ] est neutralisée chez la plupart des locuteurs natifs (Blanche Benveniste 2000: 31) si bien que plusieurs linguistes préfèrent utiliser l'archiphonème /E/ pour les deux cas, ce qui augmente le nombre de formes associées à ces deux phonèmes.

Quant aux verbes du deuxième groupe, l'homophonie est plus réduite, mais elle implique toutefois la coïncidence entre les trois personnes au singulier du présent ainsi que le participe passé.

(10) [fini] → présent : *je finis, tu finis, il finit*
→ participe passé : *fini*

La zone préverbale est également opaque, puisqu'il s'agit de la position d'un ensemble de marqueurs non accentués dont le pronom sujet (souvent réduit dans le parlé informel : [il]→ [i], [ty] → [t]), le morphème de la négation 'ne', le pronom objet ou oblique et la forme fléchie de l'auxiliaire. Les auxiliaires *être* et *avoir* présentent à leur tour une faible opposition autour des sons [e] / [ɛ] pour certaines personnes du singulier ('j'ai', 'tu es' / 'il est').

Si on raisonne en termes d'input (essentiellement oral) auquel les apprenants sont exposés, il en résulte de nombreuses séquences dont les informations grammaticales sont assez peu perceptibles (cf. 11, exemples tirés de Dietrich *et al.* 1995 : 148) :

(11)	[ʒəpaʁlɛ] vs. [ʒepaʁle]	→	‘je parlais’ vs. ‘j’ai parlé’
	[ʒesɛ] vs. [ʒeesɛje]		‘j’essaie’ vs. ‘j’ai essayé’
	[ilesɛ] vs. [illesɛ]		‘il essaie’ vs. ‘il l’essaie’ ou ‘il les sait’
	[ilaport] vs. [illaport] vs. [illaporte]		‘il apporte’ ou ‘il la porte’ ‘vs. ‘il l’a porté’
	[ilavole] vs. [illavole	→	‘il a volé’ vs. ‘il l’a volé’

Les problèmes de segmentation du parler natif sont également signalés dans le corpus oral du Groupe Aixois de Recherche en Syntaxe (GARS). Blanche Benveniste (2000: 31) souligne ainsi les possibilités multiples de transcription de segments discursifs (indiquées entre barres obliques) lorsque deux transcriptions sont également plausibles en contexte, comme c’est le cas en (12) (cf. faible opposition entre [e] vs. [ɛ], ainsi que présence du *ne* de négation qui se confond avec liaison en *n*-).

(12) un monsieur qui /se fait, s’est fait/ élire
/j’étais, j’ai été / frappé de voir que …
ils disaient qu’en France on (n’) osait pas trop

Compte tenu de l’écart entre français écrit et français parlé, l’orthographe française donnerait une image déviante du fonctionnement de (la morphologie à) l’oral chez les natifs[4]. Cette image serait d’autant plus faussée pour la production d’apprenants débutants qui sont en train d’apprendre la LC, tels que ceux enregistrés dans le projet ESF. Il est bien connu, par ailleurs, que les oppositions phonologiques basées sur le degré d’aperture entre [e] vs. [ɛ] ou sur l’arrondissement des lèvres entre [e, ɛ] vs. [ø, œ], sont développées très tardivement en français L2 et constituent un véritable défi aussi pour les apprenants plus avancés et/ou exposés à un apprentissage en milieu institutionnel (cf. Véronique *et al.* 2009).

Le choix méthodologique appliqué dans le cadre de ce projet pour le français L2 est ainsi motivé, d’une part, par l’ambiguïté inhérente à certains segments verbaux et, d’autre part, par la volonté de retarder leur interprétation afin de l’appuyer sur des critères plus solides.

[4] Par ailleurs, la maîtrise des différentes graphies de /E/ constitue un défi aussi pour les locuteurs natifs, comme le montrent Brissaud *et al.* (2006) dans une étude sur les productions d’enfants entre 8 et 15 ans.

Il est utile de souligner à ce propos que, dans les transcriptions originales ESF, l'emploi du symbole E a été parfois adopté, même si pas de manière systématique, pour éviter les différenciations subtiles entre la voyelle [e] et [ɛ] en français L2. En effet, il est difficile de présupposer d'emblée qu'un apprenant débutant puisse discriminer ces deux sons en langue cible. Nous optons pour ce codage dans les exemples cités dans cet article.

3. *Illustration des enjeux théoriques par deux études de cas*

Les choix de transcription appliqués dans le projet ESF reflètent, comme dit plus haut, l'approche *learner variety* qui envisage les lectes des apprenants comme des systèmes à part entière, dont on essaie de décrire les principes de fonctionnement indépendamment de la LC, à savoir sans attribuer aux formes produites en L2 les catégories des formes correspondantes en LC.

La transcription en phonétique des formes verbales en français L2 s'impose, surtout avec la production d'apprenants débutants, afin de ne pas sur-interpréter d'emblée leurs connaissances de la LC. Ceci dit, ce choix présente également des enjeux théoriques sur l'interprétation du parcours acquisitionnel, que nous allons illustrer à travers deux exemples sur l'acquisition du français L2 : les deux portent sur l'émergence de la morphologie verbale qui sera mise en relation, d'abord, avec l'expression des relations temporelles et, ensuite, avec l'acquisition de la négation.

3.1 Morphologie verbale et relations temporo-aspectuelles

L'acquisition de la temporalité a fait l'objet de nombreuses études. Dans ce qui suit nous allons focaliser l'attention sur l'émergence de la morphologie verbale dans l'expression des relations temporo-aspectuelles, soit sur le développement qui caractérise le passage entre la Variété de Base – permettant d'exprimer un ensemble riche de relations temporo-aspectuelles grâce à des moyens discursifs ou lexicaux, mais limité par l'absence de flexion verbale – et le stade suivant, la Variété postbasique, où la temporalité commence à être encodée par des moyens grammaticaux (émergence de la finitude) (cf. Tableau 2).

Cette transition a donné lieu à un débat concernant la valeur des premières oppositions verbales : s'agit-il d'oppositions encodant une valeur temporelle (*marquage de l'opposition entre présent vs. passé / futur*) ou bien une valeur aspectuelle (*marquage de l'aspect perfectif vs. imperfectif*) ?

Variété de Base		Variété Postbasique
Tps/Asp.: Moyens lexicaux et discursifs		Tps/Asp. : Moyens grammaticaux
(adverbes temporels, principes d'organisation discursive, marqueurs bornes…. mais absence de flexion verbale fonctionnelle)		*Flexion verbale :* *- Temps (prés. / passé / futur) ?* *- Aspect (perf. / imperfectif)?*

Tableau 2. *Expression des relations temporelles.*

Sur ce point on constate des résultats divergents. Dans la synthèse des résultats pour les différentes combinaisons de LS/LC du projet ESF, Dietrich *et al.* (1995) arrivent à la conclusion générale que le marquage grammatical du temps (*présent/passé*) précède le marquage grammatical de l'aspect :

> « **Tense marking precedes aspect marking**. All target languages of this study have grammatical tense marking, only some of them have grammatical aspect marking, but all can mark aspect by various types of periphrastic constructions. In all cases, tense comes first » (Dietrich *et al.* 1995 : 270).

Cette généralisation ne va pas contre le fait que pour certaines langues cible puissent apparaître des formes encodant l'aspect grammatical (par exemple, l'apparition précoce de V-*ing* en anglais), puisque leur emploi n'est pas fonctionnel, ou que certains apprenants puissent développer une opposition indéterminée entre le système aspectuel et temporel, comme c'est le cas des apprenants arabophones du français.

Le projet de Pavia, qui dispose d'un large corpus sur l'apprentissage de l'italien L2 par un public similaire à celui du projet ESF (migrants adultes en immersion)[5], arrive en revanche à la conclusion opposée, à savoir que le marquage grammatical de l'aspect (opposition perfectif/imperfectif)

[5] Le corpus du projet de Pavia comporte des données d'apprenants de l'italien L2 ayant différentes L1, dont entre autres le chinois mandarin, le tigrinya et l'allemand (cf. Andorno & Bernini 2003).

précède l'expression grammaticale du temps. En prenant en compte les trois catégories sémantiques grammaticalisées dans les verbes de la LC, le développement en italien L2 est synthétisé de la manière suivante : aspect > temps > mode (Banfi & Bernini 2003: 95).

Sans remettre en question les conclusions des études mentionnées ci-dessous, dans les sections suivantes nous allons en reprendre les résultats concernant l'italien et le français L2 pour comprendre comment la manière de transcrire les données peut jouer un rôle dans les analyses et leur interprétation.

3.1.1 Morphologie verbale et relations temporelles : italien L2

Les études sur l'acquisition de l'italien L2 (cf. Giacalone Ramat 1992; Banfi & Bernini 2003; Bernini 2010) mettent en évidence la séquence acquisitionnelle suivante concernant le développement de la morphologie verbale : presente / infinito > (ausiliare +) participio passato > imperfetto > futuro > condizionale …(*présent /infinitif > (auxiliaire +) participe passé > imparfait > futur > conditionnel...*).

Avant de revenir sur les débuts de ce parcours acquisitionnel, il est utile de souligner que, tout en étant deux langues romanes, le français et l'italien présentent des différences notables : contrairement au français, l'italien est une langue à sujet nul (*pro drop*) et, surtout, elle ne présente pas de morphologie silencieuse. A titre d'exemple, dans le cas du présent d'un verbe régulier, à l'oral comme à l'écrit, une désinence spécifique encode chaque personne ainsi que le marquage de l'opposition singulier/pluriel.

Italien 'parlare'	ECRIT	ORAL
1ère p. s.	parl-o	[ˈpar.lo]
2e p. s.	parl-i	[ˈpar.li]
3e p. s	parl-a	[ˈpar.la]
1ère p. pl.	parl-iamo	[par.ˈlja.mo]
2e p. pl.	parl-ate	[par.ˈla.te]
3e p. pl.	parl-ano	[ˈpar.la.no]

Tableau 3. *Présent du verbe* parlare *(fr*. parler)

Les premières formes verbales apparaissant en italien L2 ressemblent à des verbes fléchis : elles correspondent le plus souvent à la 3e personne (*parla*) ou la 2e personne du singulier (*parli*) du présent, plus rarement à l'infinitif (*parlare*). La désinence de la forme verbale en question est aisément identifiable du point de vue de la langue cible, mais elle est initialement dépourvue de valeur fonctionnelle en italien L2 : l'apprenant l'utilise comme une forme figée qui peut apparaître dans tout contexte temporo-aspectuel. En effet, comme on peut le voir dans les exemples suivants, il n'y a pas de correspondance entre la désinence verbale utilisée par l'apprenant et ses fonctions (personne, temps) en LC.

(13) io *parla* + mh – tigrigna e am/ amarigna + dell'Itiopia
je parle-3e p.s. + mh – tigrinya et am/ amarinya + de l'Etiopie
(Banfi & Bernini 2003: 92)

(14) appena in Italia + non + non *capisco* idalia tutto
à peine en Italie + NEG + NEG comprends-1e p.s. italien tout

++ poi + *vai* + *va* in libreria *complale* un liblo
++ après + vas-2e p.s. + va-3e p.s. dans librairie acheter un livre

chiama lingua per ai/ + la lingua italiana per i stranieri
(s') appelle langue pour aux/ la langue italienne pour les étrangers

in casa + stasera+ in casa *studiare* ++ questo libro due anni
en maison + ce soir + en maison étudier-INF + ce livre deux ans
(quand je suis arrivé en Italie je ne comprenais pas l'italien, ensuite je suis allé
dans une librairie acheter un livre qui s'appelait la langue italienne pour étrangers.
A la maison le soir, j'ai étudié ce livre pendant deux ans)
(Giacalone Ramat 1992 : 155)

Il s'agit donc de 'formes de base' n'ayant qu'une valeur lexicale. Pour mettre en évidence cet écart par rapport à la LC, dans les travaux sur l'italien L2 elles sont parfois représentées avec le symbole Ø au lieu de la flexion verbale : *parla*-Ø ; *studiare*-Ø (cf. par exemple Banfi & Bernini 2003).

L'étape successive correspond à l'émergence de formes équivalentes au participe passé (forme de base + suffixe *-to*), qui entrent en opposition avec les formes du présent, établissant ainsi une opposition entre perfectif

(et/ou passé) vs. imperfectif. L'extrait (15) illustre bien ces valeurs. L'apprenant décrit des vignettes (dessins) et commente le fait qu'une femme lave des casseroles (*lava pentola*) ; une fois la casserole lavée (*lavato pentola*), elle se regarde dedans comme dans un miroir. Le participe passé encode la fin de l'opération de laver et l'état qui en résulte de propreté (valeur résultative de l'action de laver)[6].

(15) *lava* eh pentola +++
lave (3e p.s.) casserole

la/ eh + lavato eh pentola +++
guarda eh +++ come eh +++ specchio
la/ eh + lavé-part.passé casserole +++
regarde eh +++ comme eh +++ miroir

(Banfi & Bernini 2003 : 94)

En résumant l'évolution des plusieurs apprenants du corpus en question, on constate que tous (à part un) arrivent à disposer de formes du présent (et infinitif) en opposition avec des formes du participe passé (« a parte Hagos (…) tutti gli apprendenti dispongono di forme di presente (e infinito) in opposizione a forme di participio passato », Banfi & Bernini 2003: 90).

Le développement au-delà de ce stade est marqué par l'émergence des auxiliaires *essere* ('être') et *avere* ('avoir') qui contribuent à la composition du *passato prossimo* (Aux+ participe passé), construction équivalente au passé composé en français. Dans cette phase, il n'est pas rare de constater la production de formes verbales auxiliées qui ne sont pas encore conformes à la langue cible (cf. en 16a choix de l'auxiliaire 'avoir' au lieu de 'être' et accord au singulier au lieu du pluriel), ou encore l'émergence de formes analytiques encodant séparément les significations grammaticales (accord, temps de la validité de l'assertion) et lexicales (16b) au lieu d'être fusionnées dans une seule forme (dans ce cas l'imparfait du verbe 'aller'= *andavano*).

[6] Il est à noter que la valeur du participe passé n'est pas toujours aussi claire : dans d'autres occurrences, son emploi présente une certaine ambiguïté entre la lecture perfective et le passé, étant donné leur solidarité sémantique (cf. Giacalone Ramat 1992 ; Banfi & Bernini 2003).

(16a) tanti persone *ha-morto* qui (= tante persone sono morte)
beaucoup de gens a-3e p.s. mort (= sont mortes) ici
(beaucoup de personnes sont mortes ici)
(Giacalone Ramat 1992 : 158)

(16b) erano andare (= andavano)
Étaient-3e p.pl. IMPF. aller-INF (= ils allaient)
(Banfi & Bernini 2003 : 107)

Toujours est-il que l'apparition de l'auxiliaire marque un moment charnière dans le développement des lectes d'apprenants en direction de l'organisation temporelle de l'italien : c'est l'auxiliaire (et/ou la copule) qui encodent l'intervalle temporel de validité de l'assertion.

Le tableau suivant visualise l'évolution de ces premières étapes : l'emploi d'une forme verbale initialement indifférenciée se réduit avec l'encodage d'une première opposition aspectuelle entre imperfectif (présent)/perfectif (participe passé). Le marquage de l'opposition temporelle (présent/passé) est liée à l'émergence de l'auxiliaire qui participe à la construction du *passato prossimo*.

Formes de base	→	presente (*imperfectif*) participio passato (*perfectif*)	→	presente ausiliare + P.P.

Tableau 4 - *Emergence de la morphologie verbale pour les relations temporelles en italien L2*

3.1.2 Morphologie verbale et relations temporelles : français L2

Les premières formes verbales apparaissant en français L2 peuvent ressembler également à des verbes fléchis, mais il est plus compliqué de les faire correspondre à une forme spécifique de la LC, comme le demanderait une transcription orthographique. Il s'agit notamment de formes qui, du point de vue de la LC, semblent soit comporter la présence d'un suffixe (en particulier [e] ou [i]), soit correspondre au radical du verbe, ce qui équivaudrait à différentes personnes du présent pour les verbes réguliers. Ce sont en réalité des formes figées, invariables dans le système de l'apprenant, puisqu'elles apparaissent dans tout contexte temporel ou personnel. A titre d'exemple, l'extrait (17) montre un

contexte, induit par la question de l'intervieweur sur la matinée de l'apprenante, qui demanderait en LC le marquage du passé dans la réponse. Etant donné l'opacité du système verbal français à l'oral (et surtout les nombreux cas d'homophonie), une transcription orthographique des segments verbaux poserait de sérieux problèmes d'interprétation, puisque le même segment pourrait être transcrit de manières différentes. Ainsi, en (17a) pour « [preparer] » il s'agirait de choisir entre l'infinitif (*préparer*) ou le participe passé (*préparé*) – sachant qu'on ne peut exclure non plus des formes d'imparfait (*préparais / préparait /* ...) dont la prononciation est phonétiquement similaire –, en (17b) pour « je [le fe] » entre présent (*je fais* 1ère p.s. présent) et passé composé (*j'ai fait*) ou participe passé (*fait*), et en (17c) pour « [sorti] » entre le participe passé (*sorti*) du verbe *sortir* et une forme potentiellement possible de présent idiosyncrasique (*je *sortis*), construit sur le modèle du verbe *finir* (*je finis*).

(17) *LN: qu'est ce que tu as fait alors depuis ce matin ?*
(a) je [prepare] le *por* [manʒe] → Vinf *préparer* ? P.P. *préparé*?
(b) après je [le fe] un petit peu de ménage… → Présent *fais*? p.p. *fait*?
(c) *y* je [sorti] à [un or trent] → P.P. *sorti*? Présent *sortis ?

A l'inverse, l'exemple (18) montre une forme verbale que l'on serait tenté de transcrire orthographiquement comme *je travaille,* puisque sa forme phonétique pourrait correspondre à la 1ère p.s. présent du verbe *travailler* en LC[7]. Cependant, il est évident que l'apprenante l'utilise pour faire référence au passé, comme l'indiquent clairement les adverbiaux utilisés (*avant, jamais au Chili = jamais quand j'étais au Chili*)

(18) *LN: qu'est-ce que tu fais ici ? tu travailles ?*
BE: avant je [travaj] maintenant non → présent *travaille*?
(avant je travaillais/j'ai travaillé, maintenant non)
LN: tu sais bien faire la cuisine?
BE: oui [solamã] ici
(oui, seulement ici = en France elle travaille dans la cuisine d'un foyer)

[7] NB : il n'est pas à exclure à ce stade que le segment en question puisse être le substantif 'travail', homophone des formes fléchies (1ère p.s., 2e p.s., 3e p.s. et pl.) du verbe *travailler* au présent.

BE: je jamais jamais je [travaj] au chili
(je n'ai jamais travaillé au Chili)

Afin d'éviter toute identification abusive par rapport aux multiples désinences possibles de la LC, les formes verbales en français L2 ont été toutes transcrites en symboles phonétiques. Dans le même but, elles sont également représentées en symboles phonétiques dans les analyses successives ayant conclu qu'à ce stade, celui de la variété de base, les lexèmes verbaux en français L2 apparaissent typiquement sous les trois formes de base suivantes (cf. Noyau *et al.* 1995)[8]:

V-*e* comme [vole]
V-*i* comme [sorti]
V-Ø comme [travaj]

Dans les trois cas, il s'agit de formes dépourvues d'informations grammaticales, à savoir qui ne présentent pas d'opposition fonctionnelle entre elles (comme c'est le cas en anglais pour la variation initiale entre V et Ving), ni de correspondance spécifique au temps ou à la personne demandés par le contexte discursif.

Pour les apprenants dépassant ce stade, le développement est marqué, d'une part, par la spécialisation de V-Ø dans le marquage du présent et, d'autre part, par l'émergence des auxiliaires *être/avoir* + V-*e* / V-*i*, qui permet de discriminer la référence aux contextes présents vs. passés.

Les formes auxiliées peuvent, cependant, coexister longuement avec les formes de base en V-*e*. Ainsi, dans l'extrait (19), qui implique un contexte passé, il est à noter que l'apprenante BE utilise dans le même passage aussi bien de formes correspondantes au passé composé (je [swi arive] = je suis arrivée ; je [swi perdu] = je me suis perdue), appropriés au contexte temporel (récit de son arrivée en France), que des formes en V-*e* ([komense], [reste]). Pour ces dernières, il est difficile de décider s'il s'agit de formes de base persistantes du stade précédent, d'infinitifs, ou bien de premières formes d'imparfait.

[8] L'emploi initial de verbes en Ve/Vi/VØ semble bien être une caractéristique du français L2, puisqu'elle a été attestée chez différents types d'apprenants, aussi bien en milieu naturel qu'institutionnel (cf. Bartning & Schlyter 2004 ; Véronique *et al*. 2009).

(19) *LN vous saviez où aller quand vous êtes arrivée ici?*
je [swi arive] dans un foyer de transit
* y * [ʒe me di] « [se] pas possible [reste] toute la journée ici »
[ilja] de sortir de ici…
après [ʒe komense] à sortir toute seule
y jamais jamais je [swi perdu] dans le métro
(Noyau *et al.* 1995 :195–6)

C'est l'analyse longitudinale de ses productions qui nous permet de voir l'évolution plus complète de cette apprenante. Le tableau 5, tiré de Benazzo & Starren (2007), reporte toutes les formes verbales produites lors d'une tâche de narration (le récit d'un extrait du film *Les Temps modernes*) qui a été répétée trois fois dans le recueil des données ESF.

Le récit du premier cycle montre l'emploi de 6 verbes différents, qui représentent l'actualisation typique des formes de base V-*e*/V-*i*/V-Ø. La production du cycle II permet d'apprécier l'accroissement du répertoire verbal produit pour la même tâche et, en parallèle, l'augmentation de formes en V-Ø . Toutefois, ce n'est qu'au cycle III qu'on peut constater un changement qualitatif dans la morphologie verbale avec des signes fiables de progression, notamment à travers la production de formes auxiliées (*être* ou *avoir* + V-*e* ou V-*i*), qui s'opposent aux formes du présent (par ex. [sɔr] vs. [asorti]), ainsi que la présence de proto-modaux (*vouloir*)[9] et des périphrases aspectuelles (futur proche *aller* + *Vinf* ; progressif *être en train de*).

Malgré cette progression évidente, la proportion des formes de base en V-*e* se réduit mais ne disparait pas jusqu'à la fin du recueil des données. Par ailleurs, une analyse plus fine de leurs contextes d'apparition dans différentes tâches conduit à la conclusion que la valeur temporo-aspectuelle de l'imparfait n'est pas maîtrisée par cette apprenante (en d'autres termes il n'y a pas d'opposition au passé entre PC et IMP) *:*

> « if there is an emergent pattern V-*e* as opposed to V-Ø in the very last data, it does not support an aspectual contrast perfective/imperfective, as in the French Passé Composé/Imparfait (…), but rather the marking of hypothetical (non-referential) situation vs. actual » (Noyau *et al.* 1995: 196)

[9] [vudra] / [vudrE] sont en réalité utilisés dans un contexte demandant le présent *veut.*

APPRENANT	BERTA (esp.> français)		
CYCLE	I	II	III
Formes de base V-*e* – V-*i*	3 [tSerSe], [sorti], [vole]	7 [aSEte], [marSe], [parti], [prepare], [sorti], [entruve], [tombe]	6 [sorti], [prɑ̃de], [vɑ̃dr], [truve], [uvre], [vole]
V-Ø	3 a, va, [di (ke)]	7 [demɑ̃d], [atrap], [pas], [vjen], [rest], [vi]/[vy],[fE], [return]	21 [apel], [searete], [ariv], [demɑ̃d], [di], [dorm], [komɑ̃s], [krwa], [mont], [pas], [prɑ̃d], [profit], [rest], [revej], [sɔr], [suy], [se(n)truven], [turn], [(se)tɔ̃b], [vwa], [(sã)va]/[ve]
Avoir + P.P.	-	-	5 [avole], [avy], [adone], [asorti],[atɔ̃be]
Être + P.P.	-	-	5 [ilesorti], [eparti] / [sonparti], [sontɔ̃bE], [sonperdy], [sefrape]
Proto-modaux + Vinf	-	-	[vudra] / [vudre]
Aller + Vinf	-	-	va + Vinf
Etre en train de + Vinf	-	-	[il entre] de / [sonentre] de V
Imparfait	-	-	-

Tableau 5. *Formes verbales attestées chez BE dans le récit des* Temps Modernes

Etant donné l'opacité du français aussi bien dans la zone des suffixes que des préfixes (cf. Section 2.2), on ne considère comme signes fiables de la présence d'un auxiliaire que les occurrences où une forme auxiliée clairement perceptible s'oppose à une forme simple du même verbe :

(20) [vole] → [avole]
[tɔ̃be] → [sɔ̃tɔ̃be]

En effet, on peut toujours hésiter entre forme de base ou forme auxiliée dans l'interprétation de cas qui impliquent comme préfixe les sons [e] / [ɛ] / [ə] – tels que [sefrape]: *s'est frappé* ou *se frapper* ? [ʒeparle] *: j'ai parlé ou je parlais ?* – oppositions que les hispanophones en particulier (mais aussi les apprenants avec d'autres L1) maîtrisent tardivement en français.

Les premiers stades de développement d'une morphologie verbale fonctionnelle en français L2, du moins en ce qui concerne les relations temporo-aspectuelles, peuvent donc être schématisés de la manière suivante (Tableau 6) :

Formes de base			Présent	V-Ø [ilvol]
V-*e*	[vole]	→	Passé composé	Aux-Vlex [ilavole]
V-*i*	[sorti]			
V-Ø	[travaj]		Futur proche	*va* + Vlex [ilvavole]

Tableau 6. *Emergence de la morphologie verbale pour les relations temporelles en français L2*

La transition entre la variété de base et le stade suivant est ainsi caractérisée par l'apparition d'un ensemble de marqueurs morphologiques préverbaux. La séquence acquisitionnelle comporte ensuite le développement de formes telles que l'imparfait (et plus tard le plus-que-parfait) pour le passé ainsi que du futur simple pour le futur (stades avancés, cf. Véronique *et al*. 2009 ; Bartning & Schlyter 2004).

La comparaison des itinéraires développementaux en italien vs. français L2 met en évidence des tendances communes, telles que l'absence d'informations grammaticales dans les premières formes verbales, mais aussi quelques différences liées à la LC, dont la difficulté à catégoriser la désinence des premières formes verbales en français et l'absence d'une

étape caractérisée par l'opposition entre le présent (inaccompli) vs. participe passé (accompli), qui précéderait l'apparition de l'auxiliaire. A ce propos il est à souligner qu'en fonction de la manière de transcrire, on pourrait arriver à la conclusion inverse, notamment si les formes en V-*e* / V-*i* étaient transcrites comme des participes passés. L'emploi de la transcription phonétique montre en revanche que, même si on voulait supposer un développement en français L2 similaire à celui constaté en italien L2, cette étape ne serait pas détectable à l'oral en français, puisque la plupart des formes de base coïncident phonétiquement avec celles du participe passé ([parle] – [sorti]).

En revanche, ce qu'on remarque pour exprimer la valeur perfective ou passée c'est l'association de la forme verbale non-finie avec le marqueur de bornes [fini] et, en particulier chez les apprenants hispanophones, avec l'adverbe '*déjà*' (cf. Benazzo & Starren 2007). Le recours à cet adverbe est illustré dans l'exemple suivant, tiré d'une conversation libre pendant laquelle BE raconte la fête d'anniversaire d'un de ses enfants.

(21) (contexte : récit de la fête d'anniversaire d'un enfant de BE)
BE: [ʒarive] à la maison à 8 h *y* la fête déjà [fini]
(je suis arrivée chez moi à 8h et la fête était déjà finie)

De même, après un long passage dans lequel un autre apprenant hispanophone (AL) explique ses problèmes dans les démarches administratives à la préfecture et que le LN n'arrive pas à comprendre si la situation décrite s'est déroulée dans le passé ou doit avoir lieu dans le futur (*tu y est allé ou tu iras* ?), c'est la présence de *déjà* (avec une forme verbale non-finie) qui permet de situer la situation dans le passé.

(22) LN: *tu retourneras à la préfecture pour le logement ?...tu y es allé + ou tu iras*?
AL: .. je déjà je [ale] là bas oui

La fonction de *déjà* se prête à plusieurs interprétations qui rappellent la double lecture du participe passé en italien L2 : d'une part, on pourrait le considérer comme un moyen parmi d'autres qui contribue à localiser dans le passé la situation décrite par le verbe non-fini (voir un passé dans le passé en 21) ; d'autre part, il n'est pas à exclure une valeur purement aspectuelle d'accompli. Il n'en reste pas moins que le passage suivant confirme de manière assez convaincante la valeur discriminante de *déjà*

avec des formes verbales de base pour véhiculer la valeur perfective du passé composé.

(23) AL: je [te done] mon numéro téléphone…
LN: y si dices por ejemplo has anotado mi teléfono en un papel ?
AL: tu déjà [ekrive] mon numéro de téléphone

Face au segment verbal ambigü que AL produit en L2 – [te done] pouvant correspondre aussi bien à la forme fléchie avec PC (*je t'ai donné*) qu'à l'énième emploi de la forme de base en V-*e* ([*done*]) –, l'enquêteur demande à l'apprenant de dire en français 'has anotado', soit l'équivalent du passé composé ('tu as pris note du numéro de téléphone'). Dans sa traduction en français, AL semble établir une équivalence entre le passé composé espagnol 'has anotado' et *déjà* + V non fini [ekrive].

Il est à noter que si l'on suivait une transcription orthographique standard, on ne saurait décider comment transcrire la forme [ekrive], qui pourrait être interprétée comme une tentative d'imparfait ('*tu écrivais'*), cependant inappropriée au contexte, ou encore comme un participe passé idiosyncrasique (**écrivé* , au lieu de *écrit*, sur le modèle de *parlé*). L'évolution générale du lecte de l'apprenant nous indique en revanche qu'elle fait partie des formes de base en V-*e* et que l'association avec l'adverbe *déjà* représente l'un des moyens pour la désambiguïser.

Ce survol rapide sur l'émergence de la morphologie verbale en L2 donne lieu à deux remarques : au niveau acquisitionnel, il est tout à fait plausible que les caractéristiques de la LC influent sur les étapes développementales en L2 (la comparaison des trajectoires en l'italien et en français L2 montrent des tendances similaires, modulées par l'exploitation de moyens différents), mais au niveau méthodologique il est évident que, en fonction des choix de transcription des formes verbales, on pourrait arriver à des conclusions assez différentes concernant le degré de morphologie présente dans les productions en français L2 et ses fonctions.

3.2 Acquisition de la négation en français L2

L'acquisition de la négation a fait l'objet de plusieurs travaux se basant sur le corpus du projet ESF (p.ex. Becker 2005, 2012 ; Giuliano & Véronique 2005 ; Stoffel & Véronique 2003 ; Silberstein 2001). Nous l'illustrons ici par l'étude de Giuliano (2005) en reprenant son analyse de

l'acquisition de la négation en français L2 par des apprenants hispanophones du projet ESF. Giuliano (2005) éclaire certains phénomènes liés à l'émergence de la morphologie verbale en français L2 en relation avec le développement de la négation discontinue en français.

La négation en français présente des difficultés d'analyse à la fois du point de vue de l'apprenant et de celui du chercheur qui doit découvrir le système émergent en L2. Ces difficultés sont en grande partie liées à l'ambiguïté phonologique dont il a été question dans la section précédente.

Ainsi, la négation en français est discontinue et construite grâce à deux éléments qui entourent la forme finie du groupe verbal (*ne* + V_{fin} + *pas*), alors qu'en espagnol, langue maternelle des apprenants analysées par Giuliano, la négation est typiquement pré-verbale (*no* + V_{fin}). Pour un apprenant débutant, l'acquisition de la négation discontinue représente un défi important dans la mesure où la maîtrise de celle-ci va de pair avec le processus de grammaticalisation du lecte de l'apprenant et l'émergence de la morphologie flexionnelle productive. Autrement dit, tant que l'apprenant ne parvient pas à décomposer l'énoncé en constituants majeurs (Sujet – Verbe – Complément) et à identifier, grâce à la morphologie, la copule, les auxiliaires et les modaux, les règles de la négation en LC ne pourront pas être opérationnalisées. Si on prend en considération le système phonologique du français particulièrement ambigu (cf. section 2.2. ci-dessus), l'identification des formes verbales à valeur d'auxiliaire et de copule, combinées avec les négateurs, devient très difficile. En effet, certaines distinctions phonologiques sont très subtiles ou nulles (cf. exemple 24 ci-dessous) :

(24) [ilnapa] 'il n'a pas'
[ilnjapa] 'il n'y a pas'
[ilnɛpa] 'il n'est pas'
[ilnɛpa] 'il n'ait pas'

L'apprenant doit faire face à une difficulté de segmentation liée à la faible perception de la copule et à la proximité phonologique avec d'autres formes.

Par rapport aux étapes acquisitionnelles initiales présentées dans la section 2.1., l'acquisition de la négation qui se dégage de l'analyse de Giuliano (2005) montre une intégration du négateur français dans la structure du lecte d'apprenant dans les 3 étapes reportées ci-dessous.

(1) Variété prébasique	**(2) Variété de base**	**(3) Variété postbasique**
Neg + X X + Neg	Neg + Vnonfin	Aux/Mod + Neg + Vnonfini

Tableau 7. *Evolution de la négation chez Berta (apprenante hispanophone, corpus ESF)*

Nous illustrons les difficultés que pose la transcription des séquences contenant la négation en français L2, en nous basant sur des exemples provenant des productions de Berta, une apprenante hispanophone du projet ESF.

Variété prébasique

Les énoncés que l'apprenante construit au niveau du premier palier acquisitionnel, la variété prébasique, ne possèdent pas encore d'élément verbal. Deux négateurs idiosyncrasiques sont produits par l'apprenant, [no/non] et [nEpadE], qui portent sur un des constituants nominaux de l'énoncé. Cependant, on n'atteste que très peu d'occurrences négatives (7 au total) dans les productions de Berta.

En suivant la logique de l'approche des lectes d'apprenants, le négateur [nEpadE] est transcrit phonétiquement et est considéré comme une forme non analysée. Si on optait pour une transcription filtrée par l'orthographe de la LC, une surinterprétation conduirait à présupposer l'existence d'une structure auxiliée complexe (*n'est pas de*). Une telle interprétation va à l'encontre de l'analyse globale du lecte de cette apprenante à ce stade de l'acquisition où elle ne produit pas encore d'énoncés structurés autour d'un lexème verbal. L'emploi de ce négateur ainsi que de [no/non] s'explique par la structure informationnelle de l'énoncé où il apparaît. Ainsi, lorsque Berta produit des énoncés négatifs avec [no/non], l'élément nié est celui qui constitue le topique de l'énonce, comme dans l'exemple ci-dessous :

(25) *LN: il y a des taxis ?*
BE: non + taxis non

En revanche, lorsque l'énoncé négatif contient le négateur [nEpadE], c'est l'expression de focus qui est dans sa portée (cf. exemple 26, ci-dessous).

(26) *LN: c'est un peu une imprimerie alors ?*
BE: non non [nEpadE] imprimerie + *otra forma*

L'analyse des négateurs au niveau de la variété prébasique pose un certain nombre de problèmes liés aux structures souvent ambiguës à cette étape. De plus, comme nous l'avons précisé, les énoncés négatifs sont peu nombreux, ce qui ne permet pas de déduire facilement une systématicité dans leur emploi. Cependant, le domaine d'application du négateur est transparent puisqu'il se trouve en position adjacente par rapport à l'élément nié sans qu'il corresponde à un verbe

Voici donc la distribution des négateurs au niveau de la variété prébasique:

(27) Neg [nEpadE] + X
X + Neg [no/non]

Variété de base

La complexification de l'énoncé au niveau de la variété de base, qui résulte de l'apparition d'un constituant clairement verbal, même s'il correspond à une forme non finie, va avoir des conséquences sur l'acquisition de la négation. Au niveau quantitatif, on observe, à ce stade, un accroissement des énoncés négatifs à la fois avec le négateur *no/non*, à valeur surtout anaphorique, et le négateur à structure [nEpa(dE)].

Nous poursuivons ici notre illustration de la relation entre la production orale des apprenants débutants et les choix de transcription en nous intéressant au négateur [nEpa(dE)].

Dans la période où le lecte de l'apprenante Berta atteint les caractéristiques de la variété de base, le nombre des énoncés négatifs augmente nettement (99 énoncés négatifs). Le passage de la variété prébasique à la variété de base est marqué par l'évolution de « Neg + X » vers « Neg+Vnon fini ». En effet, dans les énoncés construits autour d'un constituant à valeur verbale, les structures comme [nEpa/nEpadE] conduisent à des ambiguïtés phonétiques relevant du système du français, ce qui rend l'analyse des productions en L2 compliquée. Une transcription phonétique du groupe verbal avec ce négateur s'impose afin d'éviter toute interprétation *a priori*. Seule l'analyse de ces éléments dans différents contextes de discours de l'apprenant permettra de trouver les arguments pour préférer une interprétation à l'autre.

Ainsi, une première question qui se pose concerne la décodification de la forme [nE] dans la structure du négateur [nEpa/nEpadE]. Deux interprétations possibles sont proposées par Giuliano (2005). D'une part, cette forme correspondrait à la particule négative [nə] où on atteste un remplacement de [ə] par un son se rapprochant de [ɛ]. Cette interprétation serait fondée sur le fait qu'on atteste ce même phénomène dans la réalisation phonétique par l'apprenante des pronoms personnels, tels que *je* [ʒə] plutôt comme [ʒɛ] et *le* [lə] se rapprochant de [lɛ].

D'autre part, on peut envisager [nE] comme résultat de la contraction de la particule négative et de la forme du verbe *être* : « [nE] → *ne* + *est* (*être*) ».

Les deux interprétations sont importantes pour l'analyse de la nature des deux formes [nEpa/nEpadE], qui peuvent être considérées comme deux formes interchangeables [nEpa(dE)], dans la mesure où elles apparaissent dans les mêmes contextes d'énonciation. On admet donc qu'il s'agit d'une même structure avec deux variantes libres.

Les emplois les plus fréquents de ce négateur correspondent aux fonctions principales de *être* et *avoir*, à savoir celles d'attribution des propriétés, d'existence et de possession. Dans ce cas-là, le négateur précède un constituant à valeur nominale ou adjectivale.

Deux interprétations sont proposées pour rendre compte de la structure interne du négateur. Ainsi, [nEpa(dE)] pourrait correspondre à une négation discontinue où [E] entre [n] et [pa(dE)] aurait la valeur de la forme verbale « est » surgénéralisée à toutes les personnes et tous les temps d'*être, avoir, y avoir*. Il s'agirait donc d'une segmentation [n+E+pa] où [E] constitue un élément à valeur verbale (*est*). Ou bien, le sémantème négatif réalisé par la consonne [n] n'étant pas perçu de façon indépendante de [pas], le [E] serait amalgamée en constituant une négation continue [nEpa(dE)]. Celle-ci serait une forme figée et non analysée.

Afin de trancher entre l'une ou l'autre possibilité d'interprétation, Giuliano (2005) propose d'examiner les emplois de l'apprenante lorsqu'elle construit les énoncés négatifs avec le verbe lexical, en excluant de cette analyse des expressions formulaïques de type « je (ne) sais pas », « je (ne) comprends pas ». La fréquence d'emploi de ces structures dans l'input natif en français conduit à leur apparition précoce en L2 en tant que formes non analysées et, en tant que telles, elles ne sont pas représentatives du développement de la syntaxe négative.

En revanche, l'augmentation des énoncés avec le verbe lexical non fini, typique de la variété de base, permet de comprendre la nature et le

fonctionnement du négateur [nEpa(dE)] dans le lecte basique de Berta. Deux structurations des énoncés négatifs avec le verbe lexical sont attestées : le négateur est ainsi placé soit en position pré-verbale initiale ([nEpa(dE)] + Vlex nonfin), comme en (28), soit entre le SN et le verbe (SN + [nEpa(dE)] + Vlex nonfin + SN), comme en (29).

(28) *LN: est-ce qu'il y a un travail que vraiment vous n'aimeriez pas du tout faire ?*
BE: ah oui + [nEpadE komprende] por français à travail [de kusin]
(ah oui + je ne comprends pas à cause de mon français le travail de cuisine)

(29) BE: mon mari eh [eskri] *y* [kompri] bien le français
(mon mari écrit bien et comprend le français)
BE: mais moi [nEpadEkribir]
(mais moi je n'écris pas)

L'interprétation du négateur dans ces exemples permet d'appuyer l'hypothèse selon laquelle [nEpa(dE)] est une forme non analysée. Etant donné que le contexte d'énonciation exige les verbes au présent et non pas au passé composé, il serait difficile de postuler la décomposition de la part de l'apprenante de type « ne+AUX+pas+Vfini ».

Parallèlement, on atteste également des énoncés négatifs avec un verbe implicite ([nEpa(dE)] + SN/Sprép) comme dans l'exemple ci-dessous.

(30) BE: non + un mois après *y* moi [nEpa] l'école
(non + un mois après et moi je ne (suis) pas (allée) à l'école)

Le sens de l'énoncé produit ici par Berta ne peut être reconstruit que si on admet la présence du verbe lexical n'apparaissant pas en surface. Le négateur marque donc la négation d'un verbe lexical (*aller/faire*) qui n'est pas produit dans l'énoncé.

En résumant le fonctionnement de la négation au stade de la variété de base chez Berta, nous pouvons affirmer que [nEpa(dE)] est une structure figée. Son emploi très systématique dans la position pré-verbale montre qu'il remplit trois fonctions : celle de la négation d'existence, de possession, d'attribution des propriétés, celle de la négation des verbes lexicaux et celle du marquage de l'assertion négative dans les énoncés sans le verbe lexical. La présence du négateur dans des énoncés négatifs

uniquement dans un contexte au présent rend impossible de supposer un éventuel découpage en « Neg+Aux+V ».

Variété postbasique

L'évolution du lecte de Berta de la variété de base vers la variété postbasique est lente et progressive : elle est marquée par une émergence des distinctions morphologiques de plus en plus productives, même si la flexion verbale reste toujours instable et souvent idiosyncrasique.

Pour ce qui est de la négation, on observe un passage de la structure des énoncés négatifs

« SN + [nEpa(dE)] + Vlex nonfin + SN », attestée au stade de la variété de base, vers une intégration progressive du négateur dans le groupe verbal analysé de type « SN + V + neg + SN ». Autrement dit, la négation continue à laisser la place à la mise en œuvre de la négation discontinue. Cependant, pendant un certain temps les anciennes fonctions coexistent avec les nouvelles distinctions qui apparaissent graduellement.

A la même période, nous trouvons des productions de Berta, qui montrent qu'elle emploie les deux types de négation. Le premier exemple (31) illustre l'emploi de la négation continue précédant la copule « être ».

(31) BE: [ʒe nEpa swi] content parce que [medi] mes enfants [kElja] des problèmes
(je ne suis pas contente parce que mes enfants me disent qu'il y a des problèmes)

Dans l'exemple (32) ci-dessous, en revanche, nous attestons une tentative de construire la négation discontinue, même si l'homophonie entre la structure non analysée [nEpa] et les formes *n'ai pas/n'est pas* rend difficile, voire impossible, de décider clairement de la stratégie adoptée par l'apprenante : « [nEpa]+V » ou « ne+AUX+pas + participe passé ».

(32) BE: [ʒe nEpa veny] parce que [ʒEnEpa sy] l'adresse
(je ne suis pas venu parce que je ne savais pas l'adresse)

Et finalement, l'exemple suivant atteste d'un travail cognitif intéressant de la part de Berta qui se manifeste par les auto-reformulations portant sur les structures négatives.

(33) BE: [ilija] de personnes [kE nEpa naʒe + nE save pas naʒe]
(il y a des personnes qui ne savent pas nager)

Dans cet exemple, on voit que la reformulation lui permet de produire la négation discontinue, dans la mesure où le verbe « savoir » se trouve bien entre « ne » et « pas ».

Le choix de la transcription phonétique des structures verbales négatives permet aboutir à une analyse de la cohérence interne du lecte de l'apprenant sans un filtrage préalable par des règles de la langue cible.

4. *Conclusions*

Ce chapitre propose une réflexion concernant la transcription des données orales d'apprenants adultes d'une L2 et, plus particulièrement, les problèmes spécifiques que pose la reproduction à l'écrit du français parlé en L2.

Nous avons vu que l'objectif de reproduire de manière relativement neutre la parole en L2, propre à l'approche des lectes d'apprenants (*learner variety approach*) implique des précautions à prendre afin de ne pas interpréter le lecte de l'apprenant selon les catégories de la LC. Ces précautions sont d'autant plus cruciales pour les langues qui sont marquées à la fois par une faible correspondances graphème-phonème et par une forte présence de formes grammaticales homophones et hétérographes, comme c'est le cas en français (cf. Jaffré & Brissaud 2006). Ainsi, la transcription phonétique peut s'imposer dans certains cas, notamment en fonction des caractéristiques de la langue cible (cf. différence français *vs* italien).

En exposant les études de cas sur le début de la morphologie verbale en français L2, nous avons essayé de montrer que la manière de transcrire peut influencer l'interprétation des formes employées par des apprenants en L2 à un moment donné de l'acquisition : c'est l'évolution du lecte au fil du temps qui permet de mieux comprendre quels éléments morphologiques deviennent fonctionnels. Le risque de sur-interprétation est sensiblement plus fort dans le cas de données transversales, à savoir lorsque le chercheur ne dispose que de peu de données d'un même apprenant.

Par ailleurs, notre réflexion se limite à pointer les problèmes qui se posent dans le cadre d'études issues d'une approche fonctionnaliste, mais les choix de transcription, ainsi que l'interprétation des mêmes données,

varient également en fonction des présupposés théoriques qui guident l'analyse. Sur ce point nous renvoyons à Leclercq (2020), dont l'analyse comparative de différentes études portant sur les formes verbales avec [e] final (V-[e]) en français L2 montre l'influence du type de théorie sur les choix de transcription de cette structure hautement ambiguë à l'oral : elle oppose ainsi les études issues des théories dites « data driven », qui choisissent des solutions permettant d'éviter toute présupposition *a priori* sur les productions des apprenants, à des « theories constrained approaches », où ce choix est davantage guidé par des *a priori* théoriques. Réfléchir sur la manière dont les données sont transcrites et analysées permet de contribuer au débat plus large sur la description et l'interprétation de l'interlangue des apprenants L2, ce que propose par exemple Ortega (2014), en prenant comme exemple l'acquisition de la négation.

Notre contribution se base sur les données d'un des premiers grands corpus oraux en L2, qui a été recueilli dans les années '80, époque à laquelle la transcription était initialement faite à la main[10]. Feldweg (1993) résume les phases et les écueils qui ont jalonné le processus de mise à disposition du corpus correspondant en version électronique (banque de données de l'Institut Max Planck de Nimègue) ainsi que les divergences attestées dans la transcription de chaque LC – compromis entre transcription orthographique / phonétique, contraintes de temps pour homogénéiser les transcriptions initiales –, malgré le fait que les chercheurs impliqués aient suivi le principe général d'une transcription relativement 'neutre'.

Depuis, le nombre de corpus oraux en L2 (et des recherches correspondantes) s'est accru de manière importante. En ce qui concerne le français L2, nous pouvons citer au moins deux grands projets ayant donné lieu à des corpus de taille importante : le projet Interfra[11] (« *Interlangue française – développement, interaction and variation* », cf. Bartning 1997), centré sur les apprenants suédophones, et le projet FLLOC[12] (« *French Learner Language Oral Corpora* », cf. Marsden *et al*. 2002), fondé sur l'observation d'apprenants anglophones.

[10] Comme le fait remarquer Feldweg : « When data transcription started in 1983, the research teams did not even have a personal computer available to them, and transcriptions were handwritten" (1993 : 108).

[11] https://www.su.se/romklass/interfra

[12] http://www.flloc.soton.ac.uk/conventions.html

Les deux ont été conçus et transcrits de manière à être utilisés pour des analyses ultérieures variées. Face aux problèmes de transcription de l'oral en français L2, les auteurs ont élaboré des solutions partiellement différentes, mais les productions en L2 sont en grande partie transcrites suivant l'orthographe standard, avec une sensibilité plus ou moins forte pour l'ambiguïté potentielle des segments en L2. Ceci s'explique par des raisons pratiques telles que la lisibilité, la taille importante du corpus, la nécessité de respecter la période précise couverte par des financements reçus, ou l'objectif de les analyser de façon automatique grâce à des logiciels de traitement des données linguistiques.

Ainsi, par exemple, le corpus Interfra est transcrit pour l'essentiel de manière orthographique, mais s'en écarte lorsque le même segment discursif se prête à deux interprétations également possibles en contexte : on recourt notamment à des parenthèses (par ex. *il(s) prenai(en)t*) et, surtout, au symbole « E » pour indiquer la variation possible dans la réalisation en L2 des sons [ɛ/e] dans les formes verbales ambiguës, telles que « *passé / passait* » (*passE)*[13].

En revanche, le corpus FLLOC opte pour la transcription en format CHAT[14], avec des adaptations nécessaires pour des données en L2[15]. La ligne principale, qui reporte les énoncés des apprenants, suit la transcription orthographique 'normalisée' (en d'autres termes, les erreurs sont corrigées) pour que le programme puisse reconnaître les mots produits. La transcription n'est pas donc fidèle à la forme effectivement produite par l'apprenant, mais la consultation des fichiers sons correspondants, disponibles et reliés à la transcription, permet d'accéder aux productions originales de l'apprenant, et d'adapter, si nécessaire, la transcription aux besoins de l'analyse.

[13] Cf. manuel de transcription disponible sur le site suivant : https://www.su.se/romklass/interfra

[14] Le format CHAT est un code de transcription proposé par la banque de données en acquisition CHILDES, Child Language Data Exchange System (MacWhinney 2000), ce qui permet d'appliquer à des données transcrites dans ce format les programmes du logiciel CLAN pour le traitement automatique.

[15] A titre d'exemple, le format CHAT conçu pour des analyses automatiques par le logiciel CLAN, demandent des ajustements pour la transcription de l'oral en L2 (ajout de lignes de glose, etc.) qui alourdissent de manière considérable la tâche du transcripteur, surtout lorsque le corpus a été recueilli pour l'analyse des phénomènes linguistiques de différente nature.

La base de données *Corpus Inter Langue*, composée des productions d'apprenants du français L2 ayant un niveau de A2 à C1 (Arbach 2015), constitue un autre exemple, plus récent, où un compromis entre la taille du corpus, d'une part, et la lisibilité et les exigences techniques liées à l'analyse grâce aux outils informatiques, d'autre part, a conduit au choix de la transcription orthographique. En effet, les auteurs de ce corpus ont pris cette décision en considérant que la transcription phonétique pose des problèmes de lisibilité et rend très difficile, voire impossible les analyses automatiques.

Le choix du type de transcription au sein de différents grands corpus en L2 conduit donc à des compromis entre fidélité et lisibilité, mais est également conditionné par des contraintes relatives aux moyens matériels (logiciel de traitement des données linguistiques) et au temps à disposition.

La revue de ces différents corpus et leurs critères de transcription des données orales en L2 n'a pas comme but de critiquer les choix adoptés, mais plutôt de réfléchir aux problèmes de la transcription, en distinguant soigneusement entre la tâche de transcrire et celle de l'analyse successive, sachant que l'une peut influencer l'autre. De plus, la transcription et l'analyse ne sont pas toujours accomplies par les mêmes personnes, ce qui amplifie la question de la « neutralité » relative de la transcription.

La discussion proposée dans ce chapitre nous a permis d'aborder la question suivante, selon nous primordiale pour toute recherche en acquisition de L2 : « Les choix de transcription ont-ils un impact sur la description du fonctionnement et de l'évolution des lectes d'apprenants, sachant que cette description est essentielle pour comprendre le processus complexe d'acquisition de L2 ? ».

Références bibliographiques

Agren, Malin. 2008. *A la recherche de la morphologie silencieuse : sur le développement du pluriel en français L2*. Lund: Université de Lund. (Thèse de doctorat.)

Andorno, Cecilia & Bernini, Giuliano. 2003. Premesse teoriche e metodologiche. In Giacalone Ramat, Anna (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 27–36. Milano: Carocci.

Arbach, Najib. 2015. *Constitution d'un corpus oral de FLE : enjeux théoriques et méthodologiques*. Thèse de doctorat, Université Rennes 2.

Banfi, Emanuele & Bernini, Giuliano. 2003. Il verbo. In Giacalone Ramat, Anna (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 70–115. Milano: Carocci.

Bartning, Inge. 1997. L'apprenant dit avancé et son acquisition d'une langue étrangère : Tour d'horizon et esquisse d'une caractérisation de la variété avancée. *Acquisition et Interaction en Langue Etrangère* 9. 9–50.

Bartning, Inge & Schlyter, Suzanne. 2004. Itinéraires acquisitionnels et stades de développement en français L2. *French Language Studies* 14. 281–299.

Baude, Olivier. 2006. *Corpus oraux. Guide de bonnes pratiques*. Paris: CNRS éditions.

Becker, Angelika. 2005. The semantic knowledge base for the acquisition of negation and the acquisition of finiteness. In Hendriks, Henriëtte (ed.), *The Structure of Learner Varieties*, 263–314. Berlin: Mouton de Gruyter.

Becker, Angelika. 2012. Finiteness and the Acquisition of Negation. In Watorek, Marzena & Benazzo, Sandra & Hickmann, Maya (eds.), *Comparative Perspectives on Language Acquisition : A tribute to Clive Perdue*, 54–72. Clevedon: Multilingual Matters.

Benazzo, Sandra & Starren, Marianne. 2007. L'émergence de moyens grammaticaux pour exprimer les relations temporelles en L2. *Acquisition et Interaction en Langue Etrangère* 25. 129–158.

Bernini, Giuliano. 2010. Acquisizione dell'italiano come L2. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, vol. 1, 139–140. Roma: Istituto dell'Enciclopedia Italiana G. Treccani.

Blanche-Benveniste, Claire. 2000. *Approches de la langue parlée en français*. Paris: Ophrys.

Bley-Vroman, Robert W. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(4). 1–17.

Brissaud, Catherine & Chevrot, Jean-Pierre & Lefrançois, Pascale. 2006. Les formes verbales homophones en [E] entre 8 et 15 ans : contraintes et conflits dans la construction des savoirs sur une difficulté orthographique. *Langue française* 151. 74–93.

Corder, S. Pit. 1967. The significance of learner's errors. *IRAL* 5(4). 161–170.

Dietrich, Rainer & Klein, Wolfgang & Noyau, Colette (eds.). 1995. *The acquisition of temporality in a second language*. Amsterdam: John Benjamins.

Feldweg, Helmut. 1993. Transcription, storage and retrieval of data. In Perdue, Clive (ed.), *Adult Language Acquisition: Crosslinguistic perspectives*, vol. I, 108–130. Cambridge: Cambridge University Press.

Giacalone Ramat, Anna. 1992. Sur quelques manifestations de la grammaticalisation dans l'acquisition de l'italien comme deuxième langue. *Acquisition et Interaction en Langue Etrangère* 1. 143–170.

Giuliano, Patrizia. 2005. *La négation linguistique dans l'acquisition d'une langue étrangère*. Berne: Peter Lang.

Giuliano, Patrizia & Véronique, Daniel. 2005. The acquisition of negation in French L2. An analysis of Moroccan Arabic and Spanish 'Learner Varieties'. In Hendriks, Henriëtte (ed.), *The Structure of Learner Varieties*, 355–404. Berlin: Mouton de Gruyter.

Jaffré, Jean-Pierre & Brissaud, Catherine (éds.). 2006. *Morphographie et hétérographie*. Numéro thématique de *Langue française* 151.

Klein, Wolfgang & Perdue, Clive. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13(4). 301–347.

Leclercq, Pascale. 2020. Transcribing interlanguage. The case of verb final [e] in L2 French. In Edmonds, Amanda & Leclercq, Pascale & Gudmestad, Aarnes (eds.), *Interpreting language-learning data*, 169–196. Amsterdam: Eurosla Studies Series.

MacWhinney, Bryan. 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3ª ed. Mahwah: Lawrence Erlbaum Associates.

Marsden, Emma & Mitchell, Rosamond & Myles, Florence & Rule, Sarah. 2002. Oral French Interlanguage Corpora: Tools for Data management and analysis. *Occasional Papers* 158.

Noyau, Colette & Houdaïfa Et-Tayeb & Vasseur, Marie-Thérèse & Véronique, Daniel. 1995. The acquisition of French. In Dietrich, Rainer & Klein, Wolfgang & Noyau, Colette (eds.), *The acquisition of temporality in a second language*, 145–209. Amsterdam: John Benjamins.

Ochs, Elinor. 1979. Transcription as theory. In Ochs, Elinor & Schieffelin, Bambi B. (eds.), *Developmental pragmatics*, 43–72. New York: Academic Press.

Ortega, Lourdes. 2014. Trying out theories on interlanguage : Description and explanation over 40 years of L2 negation research. In Han, ZhaoHong & Tarone, Elaine (eds.), *Interlanguage: Forty years later*, 173–201. Amsterdam: John Benjamins.

Perdue, Clive (ed.). 1993. *Adult Language Acquisition: Crosslinguistic perspectives*, vol. I et vol. II. Cambridge: Cambridge University Press.

Perdue, Clive. 1995. *L'acquisition du français et de l'anglais par des adultes*. Paris : CNRS Editions.

Riegel, Martin & Pellat, Jean-Christophe & Rioul, René. 2009. *Grammaire méthodique du français*. Paris : PUF.

Selinker, Larry. 1972. Interlanguage. *IRAL* 10(3). 209–232.

Silberstein, Dagmar. 2001. Facteurs interlingues et spécifiques dans l'acquisition non guidée de l'anglais L2. *Acquisition et Interaction en Langue Etrangère* 14. 25–58.

Stoffel, Henriette & Véronique, Daniel. 2003. L'acquisition de la négation en français par des adultes arabophones. *Marges linguistiques* 5. 242–252.

Véronique, Daniel & Carlo, Catherine & Kim, Jin-Ok & Granget, Cyrille. 2009. *L'acquisition de la grammaire du français langue étrangère*. Paris : Didier.

Watorek, Marzena & Perdue, Clive. 2005. Psycholinguistic Studies on the Acquisition of French as a Second Language: The 'Learner Variety' Approach. In Dewaele, Jean-Marc (ed.), *Focus on French as a Foreign Language*, 1–16. Clevedon: Multilingual Matters.

Fabian Santiago
(Université de Paris 8)

Transcription et annotation de données orales pour étudier la prosodie en FLE : enjeux méthodologiques

1. *Introduction*

Les études en linguistique, en général, et en phonologie post-lexicale en particulier, sont souvent censées confronter les modèles théoriques d'analyse aux faits observables, que ceux-ci proviennent de données orales ou de productions écrites. De fait, toute étude voulant dégager le fonctionnement intonatif, accentuel ou segmental d'une langue, qu'elle soit L1 ou L2, repose essentiellement sur la description des événements sonores quantifiables extraits des données qui sont ressemblées dans des corpus. Dans le domaine des études de l'acquisition de la phonologie et la phonétique en L2, le chercheur voulant étudier l'interlangue des apprenants doit passer, dans la plupart des cas, par une étape de description fondée sur l'observation des faits. En observant une grande variété de données orales offrant une bonne représentativité de la langue cible, le chercheur peut expliquer la façon dont un système prosodique et phonologique d'une langue, qu'elle soit déjà acquise (L1) ou en cours d'acquisition (L2), fonctionne et se manifeste dans le flux de parole.

De surcroît, comparer des données produites par des locuteurs natifs et non natifs permet de dresser l'inventaire des formes prosodiques et de dire en quoi un système prosodique de la L2 s'écarte/se rapproche de celui de la L1, cela permettant à terme d'évaluer le poids du transfert de la L1 (ou d'autres L2), ainsi que les effets d'un processus universel d'acquisition en L2, entre autres phénomènes. Même si l'étude des corpus oraux permet d'atteindre de nombreux objectifs, toute approche par corpus nécessite de s'interroger sur leur nature, sur la manière de les compiler, de les annoter, etc. Pour garantir une bonne représentativité, plusieurs étapes nous paraissent essentielles (Delais-Roussarie 2003) :

a. effectuer un travail préparatoire préalable à la constitution du corpus ainsi qu'établir une méthodologie appropriée pour enregistrer les données (choix du matériel et conditions d'enregistrements) ;
b. mettre en forme les données pour les rendre exploitables (documentation du corpus, transcription et annotation) ;
c. employer des métriques/mesures appropriées pour analyser des données riches, variées et/ou de grande échelle selon le phénomène étudié.

Dans cette contribution, notre objectif est double. D'une part, nous essayons de discuter des enjeux méthodologiques qu'une telle entreprise impose au chercheur voulant étudier certains phénomènes prosodiques en français langue étrangère (FLE) : Quelles données collecter en L2 : production orale contrôlée ou spontanée ? Comment les transcrire ? Comment annoter les phénomènes segmentaux et prosodiques afin que les données en L2 et en L1 soient comparables ? D'autre part, nous exposons les différentes métriques que les chercheurs peuvent employer pour étudier la prosodie en FLE.

Notre contribution est organisée comme suit. La section deux porte une attention particulière à la méthodologie sur la constitution d'un corpus en français L2 pour l'étude de la prosodie, la phonologie et la phonétique. Dans cette section, nous soulevons les problèmes méthodologiques préalables à l'analyse des données : transcription et annotation. La section trois présente de manière plus détaillée les enjeux de l'annotation prosodique en FLE. Enfin, dans la dernière section, nous dressons un bilan sur les perspectives que ces enjeux méthodologiques soulèvent en phonologie post-lexicale et segmentale des langues secondes.

2. *Corpus oraux en L2 :* collecte *de données, transcription et annotation*

Le premier point auquel tout chercheur est confronté est de savoir en quoi consiste un corpus, d'autant que collecter des données à l'ère d'internet est aisé. Sinclair (1996 : 4) propose la définition suivante : « A corpus is a collection of *pieces of language* that are selected according to explicit linguistic criteria in order to be used as a sample of the language ».

Constituer un corpus (qu'il soit écrit ou oral, en L1 ou en L2) ne se réduit aucunement à une simple collecte arbitraire d'échantillons de langage, ni à un simple dépôt des données orales numérisées. La constitution d'un corpus doit être systématiquement organisée et contrôlée conformément à un protocole élaboré par le chercheur en fonction des hypothèses ou des objectifs fixés.

Construire un corpus d'apprenants nécessite de se poser des questions d'ordre méthodologique et théorique, d'autant que le nombre de variables qui peuvent affecter le processus d'acquisition de la L2 est considérable :

- Quel type de données doit être rassemblé et étudié pour étudier l'acquisition de la prosodie et de la phonologie d'une L2 ?
- Les tâches employées pour collecter ces données doivent-elles être conformes à un protocole très contrôlé ?
- Les tâches réalisées par les participants en L2 représentent-elles des situations de communication authentiques ?
- Comment transcrire et annoter des données en L1 et L2 afin d'étudier les phénomènes prosodiques ?

Nous considérons que répondre à ces questions et justifier les réponses en fonction des présupposés théoriques et méthodologiques retenus est une étape indispensable pour garantir la validité des observations faites lors des études fondées sur une approche par corpus. Un point important pour garantir la validité des résultats concerne la représentativité des données compilées. En effet, quelle que soit la méthodologie employée pour compiler le corpus, tout chercheur veut que les données collectées soient représentatives des situations de communication réelles ; autrement dit, que les analyses qu'il extrait des données soient faites à partir de structures langagières observables. Nous allons présenter quel type de protocole nous avons adopté pour les données que nous examinerons plus tard. Au préalable, nous allons présenter différentes approches méthodologiques utilisées dans des études sur l'acquisition des L2, et plus particulièrement, de la prosodie et de la phonétique.

2.1 Collecte et type de données

La plupart des études en prosodie de L2 a privilégié les données expérimentales (ou construites) et se centrent sur l'étude d'un seul aspect de la prosodie de la L2 (*cf.* la trentaine d'articles cités et examinés par

Colantoni & Steele & Escudero 2015). A cela s'ajoutent les effets que peut avoir le choix des structures linguistiques contrôlées : les énoncés produits par les apprenants dans les interactions orales en salle de classe ne sont pas toujours comparables aux formes langagières canoniques utilisées dans les protocoles expérimentaux très contrôlés. Ce biais affecte, sans aucun doute, les éléments prosodiques produits par les apprenants avec un tel protocole. En outre, plusieurs études ont délaissé l'étude des facteurs extralinguistiques qui peuvent affecter également l'acquisition de la prosodie d'une L2 (l'âge, le style de parole, le type de tâche demandée, etc.). Les approches par corpus peuvent contribuer à l'analyse de tous ces facteurs d'une manière plus intégrale.

En se fondant sur l'observation d'une large base de données, l'approche par corpus permet de faire des analyses d'ordre quantitatif et qualitatif. Le corpus devient une source valide pour évaluer la fréquence et le type d'erreurs chez les apprenants (Gut 2007). Un corpus d'apprenants ayant une bonne annotation linguistique et contenant des informations non linguistiques est en mesure de fournir au chercheur une description exhaustive de plusieurs aspects de l'interlangue des apprenants. Pour garantir tout cela, une réflexion sur la collecte et type de données s'avère indispensable. Dans les travaux de phonologie/phonétique, deux grands types de données orales peuvent être compilées et analysées lors de la constitution de corpus :

- les données construites : obtenues dans des conditions expérimentales très contrôlées par l'expérimentateur afin de faire émerger des structures linguistiques particulières.
- les données authentiques : obtenues dans des conditions expérimentales peu contrôlées par l'expérimentateur où toute sorte de structures langagières peut émerger.

Les données construites sont collectées dans un contexte artificiel moyennant un protocole très précis. Elles peuvent être fabriquées par le chercheur (logatomes, pseudo-mots, phrases, énoncés ou passages de textes) ou pas (emploi de textes journalistiques, histoires, extraits de romans). Ces données sont obtenues avec des tâches de lecture oralisée, mais elles peuvent également être compilées à partir de procédés expérimentaux très précis comme les techniques visant à faire répéter selon un modèle sonore (imitation, répétition). Elles peuvent être obtenues

aussi à partir d'exercices où les participants doivent décrire des images où ils sont amenés à produire des structures linguistiques particulières.

Les données authentiques, contrairement aux données construites, sont collectées à partir de la réalisation de tâches dans des situations de communications non contrôlées. Dans ces tâches, les participants n'ont aucune contrainte sur les structures langagières qu'ils doivent employer. Les techniques employées pour éliciter ce type de données demandent la réalisation de tâches où aucune contrainte n'est demandée : une interview, un échange oral, un débat, raconter une histoire. Enfin, des exercices comme les *Map Tasks* permettent au chercheur de collecter des dialogues spontanés tout en contrôlant certains contextes discursifs ou linguistiques de son intérêt.

Lorsque l'expérimentateur a une hypothèse *a priori* à vérifier, le recours à des données construites s'impose, en tout cas dans un premier temps. Un protocole de collecte plus strict permet de mieux contrôler les structures linguistiques qui seront produites par les locuteurs et ensuite étudiées. Dans la plupart des recherches consacrées à la prosodie en L2, le recours à des données construites se révèle nécessaire puisque cela permet d'évaluer certaines hypothèses de départ (les effets du transfert de la L1 surtout, mais également les hypothèses concernant la validation d'un certain ordre acquisitionnel, entre autres). Ce type de données permet au chercheur de contrôler les contextes linguistiques et discursifs dont il a besoin, et, en conséquence, d'obtenir des données d'une manière plus souple et plus économique (en temps et en investissement). Le chercheur peut opter pour contrôler certaines informations lexicales (type de lexies, taille des mots en termes du nombre de syllabes), métriques (position de l'accent lexical, absence/présence d'une force métrique motivée par une valeur pragmatique comme le focus correctif), syntaxiques (type de construction, taille de constituants syntaxiques), ou sémantiques (contrastes de sens des mots/énoncés) afin de valider certaines hypothèses. En un mot, ce type de données fournit des informations plus claires sur le sens et la manière dont un facteur en particulier peut être corrélé au processus d'acquisition d'un aspect phonologique, phonétique ou prosodique d'une L2.

Bien que nécessaires pour tester certaines hypothèses de départ, les données construites connaissent certaines limites, surtout sur le plan qualitatif. D'une part, ces données sont limitées non seulement du fait des structures prosodiques ou segmentales qu'elles illustrent, mais aussi parce

qu'elles sont déjà orientées par l'expérimentateur lors de la collecte. De fait, les données collectées avec de tels protocoles ne peuvent mettre en évidence que des phénomènes limités et imposés au départ, en l'occurrence, la validation/rejet de l'effet d'un facteur *x* ou *y* sous les conditions *a* ou *b*. En conséquence, ces données ne permettraient pas d'évaluer facilement si d'autres facteurs peuvent motiver les patrons phonétiques observés en L2.

Un autre inconvénient de ce type de données concerne leur généralisation. Les résultats et analyses qui sont extraits de l'observation de données construites en L2 ne fournissent pas une description qualitative et quantitative de leur représentativité dans l'usage effectif de la langue (Vaguer 2007). Ainsi, ce type de données n'est pas nécessairement transposable à d'autres styles de parole : « observations from artificial speech tasks cannot always be extrapolated to natural conditions » (Leather 1999 : 32).

Les données authentiques présentent certains avantages vis-à-vis des données construites. Les exemples collectés sont le résultat d'un usage spontané d'une certaine forme ou structure sans que ces énoncés soient biaisés par une hypothèse établie *a priori*. Les données authentiques connaissent aussi des limites. D'une part, la fréquence et la densité d'une forme/structure qui intéresse le chercheur ne sont pas garanties. D'autre part, même si une forme *x* ou structure *y* apparaissent avec une fréquence et une relative densité dans le corpus, leur distribution par rapport au contexte linguistique dans lequel celles-ci apparaissent ne sont pas nécessairement pertinentes pour les objectifs de la recherche. Enfin, si les données construites ne garantissent pas une description neutre de par l'hypothèse que le chercheur leur a imposée, les données authentiques ne la garantissent pas totalement non plus. En effet, on peut faire le recensement d'une certaine forme/structure dans le corpus, une tâche relativement neutre en elle-même ; en revanche, « Ce qui n'est pas neutre, c'est ce que l'on fait de ce recensement […] on élimine ce qui paraît redondant, du même type ; on garde ce qui semble le plus propre à illustrer ce que l'on veut dire, mais on ne signale pas ce sur quoi on n'a rien de particulier à observer […] » (Vaguer 2003 : 212)

Nous considérons que les données authentiques et construites en L2 ne s'opposent pas mais se complémentent. Passer par l'analyse de deux types de données est nécessaire pour la construction/validation de certaines hypothèses. La compilation des données authentiques ou des

données construites font partie de ce qu'on entend par *corpus*. Il faut noter cependant que, selon Delais-Roussarie (2003), le terme *approche sur corpus* est réservé surtout à des études utilisant des données authentiques : des échantillons de langue non fabriquées élicités selon des protocoles de collecte peu contrôlés. Cependant, une collection de données construites peut également constituer un corpus. Nous considérons que toute *approche sur corpus* repose sur l'analyse de données authentiques à laquelle peuvent s'ajouter ou non des données contrôlées.

Rares sont les corpus destinés à l'étude de l'acquisition de la phonétique et de la phonologie du français L2. De fait, ce n'est que très récemment que ce genre de corpus commence à être constitué. Parmi ces corpus, nous pouvons citer le corpus IPFC (Interphonologie du Français Contemporain) consacré à l'étude de la prononciation du français L2 (Racine *et al*. 2012 ; Durand *et al*. 2009). Ce corpus utilise des tâches similaires à celles du corpus PFC (Phonologie du Français Contemporain), lequel est destiné à l'étude phonologique des variétés du français contemporain (Detay *et al*. 2010 ; Durand & Tarrier 2006). Le corpus IPFC comprend des tâches de répétition d'une liste de mots, des tâches de lecture, un entretien avec un locuteur natif et une interaction semi-contrainte entre deux apprenants. Ce corpus a été conçu pour l'étude de la prononciation au niveau segmental, et n'est, de ce fait, pas toujours adapté pour l'étude de la prosodie.

La constitution du corpus COREIL (Delais-Roussarie & Yoo, 2011) a tenté de répondre au manque de corpus disponibles de nos jours pour l'étude de la prosodie du français L2. Ce corpus repose sur certains principes :

- Représentativité des données : le corpus a été compilé à partir d'extraits de parole comprenant autour de 10 mille mots en français L2 (30 participants), 2,7 mille mots en français L1 (dix locuteurs) et 4 mille mots dans différentes L1.
- Profil des locuteurs : ont été enregistrés des apprenants adultes du français L2 en contexte d'apprentissage au milieu universitaire ayant différentes L1 (espagnol, anglais, grec, coréen, allemand) ainsi que des locuteurs natifs monolingues du français et des locuteurs natifs de toutes les L1 des apprenants.
- Diversification des tâches : le corpus a été constitué à partir de l'application de tâches diverses, à savoir, production orale

monologuée (description d'une image), production orale semi-directive (interview), interaction orale (jeu de rôles), lecture de scripts.
- Niveau de maîtrise de la L2 : dans le but de voir s'il y a un ordre dans l'acquisition des formes prosodiques, deux niveaux de maîtrise dans la langue cible ont été pris en compte pour les apprenants (A2 *vs* B1).

Le protocole du corpus COREIL permet également de traiter une partie des données collectées, même si le processus de collection n'est pas encore terminé. Il est aussi possible d'ajouter des données supplémentaires ou bien des tâches sans perdre l'homogénéité de l'ensemble du corpus. En tout état de cause, nous avons proposé que la collecte du corpus oral d'apprenants englobe nécessairement des données construites et artificielles (enregistrées à partir de la résolution des plusieurs tâches de production orale) et des données authentiques. La compilation d'un corpus suffisamment grand et varié pourrait ainsi permettre de surmonter plusieurs problèmes méthodologiques dans les études d'acquisition de la phonologie d'une L2, en favorisant l'émergence d'une plus grande variété de structures langagières. Dans les sections qui suivent, nous discuterons de l'importance des tâches impliquées dans l'analyse de la prosodie en FLE et son annotation.

2.2 Transcription orthographique

Un corpus oral regroupe des documents qui correspondent à des transcriptions de productions orales alignées ou non avec le signal acoustique. Toute transcription orthographique de données orales en vue d'analyses phonético-phonologiques ou prosodiques doit fournir une image fidèle de ce qui a été dit. Transcrire des données sonores consiste à fournir une représentation symbolique du signal, mais cette représentation n'est pas équivalente au signal, dans la mesure où elle est le résultat d'une analyse, ou plutôt d'une abstraction, des données réelles (Delais-Roussarie 2003). Ainsi, la transcription orthographique peut être vue comme un lien entre les productions orales effectives et ce qui en est compris par l'annotateur.

Transcrire orthographiquement le discours oral nécessite une réflexion théorique et méthodologique à la fois (Bilger 2007). Selon Blanche-Benveniste & Jeanjean (1987), toute représentation écrite de l'oral exige certaines conventions d'écriture. D'autres auteurs consentent à cette idée

mais proposent également que ces conventions doivent refléter le caractère oral du corpus (Durand & Tarrier 2006). Durant ce processus de transcription, de nombreux problèmes apparaissent lorsque l'on veut représenter orthographiquement l'oralité (Blanche-Benveniste & Jeanjean 1987) :

- reconstruction de certains éléments inexistants sur le signal (transcrire *il y a* à la place de *i y a*, *il pleut* à la place de *i pleut*) ;
- choisir une représentation écrite nécessite de lever des ambiguïtés dues à l'oral – la séquence [ilpaʁlfʁɑ̃sɛ] peut correspondre aux suites orthographiques *ils parlent français* ou *il parle français*.

Dans le cas des corpus d'apprenants, s'ajoutent d'autres difficultés : le transcripteur peut se voir influencé par le crible phonologique de sa propre L1 et faire des interprétations erronées lorsqu'il ne connaît pas le système de la L1 de ces derniers (Racine *et al*. 2011). La tâche de transcription n'est pas du tout aisée, surtout lorsqu'il s'agit d'annoter la parole non native contenant « des réalisations déviantes ». Ainsi, par exemple, la suite [jenetɾaβ̪ajpa] en français L2 produit par un hispanophone dans une tâche de production orale spontanée correspondrait, très certainement à *je ne travaille pas* et non à *je né travaille pas,* ou *j'ai n'est travaille pas*.

Pour essayer de surmonter ces problèmes d'ordre méthodologique et théorique, le chercheur peut transcrire orthographiquement les éléments contenus dans le signal à partir des unités existantes en L1 et éviter au maximum de faire des interprétations. Pour cela, le chercheur peut effectuer la transcription de l'ensemble du corpus moyennant une adaptation des recommandations faites par le *Text Encoding Initiative – & Eagles* (TEI Consortium) et des conventions suggérées par le système CHAT de CLAN (MacWhinney 2000). Pour le corpus COREIL, ces conventions peuvent se résumer comme suit :

- Emploi de la ponctuation forte : seuls le point « . », le point d'interrogation « ? » et d'exclamation « ! » ont été retenus, et on a évité l'emploi des virgules, parenthèses et tout autre symbole employé dans le style scriptural conventionnel.
- Emploi de l'orthographe standard, même si le locuteur n'a pas réalisé tous les segments : [ʃsɛpa] est transcrit *je (ne) sais pas* les parenthèses indiquant les segments canoniques non réalisés.

- L'insertion d'une pause a été codée avec le symbole « (.) ».
- Le nombre de répétitions des mots ou des suites de mots a été encodé avec « [x n] », où « n » représente le nombre de répétitions de la suite en question : [ʒəʒəmapɛl] a été transcrit *je [x 2] m'appelle*.
- Les faux départs ont été indiqués avec le symbole « [/ ?] » : la suite [ʒəsɥiaʒəsɥizale] a été transcrite *je suis a [/?] je suis allé*.
- Les fragments inintelligibles ont été transcrits avec « xxx ».

En plus des conventions ci-dessus, la transcription des erreurs des apprenants mérite d'être discutée dans le corpus COREIL. Pour les hispanophones, si la morphologie a été identifiée comme correcte mais sa réalisation phonétique déviante, nous avons utilisé la représentation orthographique standard : [ʒəβ̬eosinema] est transcrit *je vais au cinéma e*t non *je bé au cinéma*. Si la morphologie est identifiée comme incorrecte mais sa réalisation est existante dans la L2, nous avons assumé une certaine connaissance morphologique de l'apprenant dans la L2 et avons fait une adaptation orthographique en fonction de ce qui reflète au mieux son interlangue : [mɔ̃pɛʁaneamɛksiko] est transcrit *mon père a né à Mexico*. On peut supposer que l'apprenant fait une surgénéralisation de l'auxiliaire *avoir* lors de la construction du passé composé. Si l'apprenant emploie sa L1, les mots sont transcrits dans le système orthographique conventionnel de sa L1 : [ʝetudibjoloˈxia] est transcrit *j'etudie biología@s:sp* (l'étiquette *@-s:spa* signale que le mot est produit en espagnol, L1 de l'apprenant). Si l'apprenant a rajouté des phonèmes non spécifiés par le système morphosyntaxique de la L2, nous avons encodé ces éléments avec une adaptation de la norme orthographique de la L2 : [ləɔm] est transcrit *le homme.*

La Figure 1 représente une copie d'écran de l'interface de CLAN montrant un exemple de transcription orthographique dans les lignes principales (*S04 indique le code associé à l'apprenant) et les lignes secondaires recevant un codage linguistique, en l'occurrence ici le type de phrase (*cf*. Santiago & Delais-Roussarie 2015 pour plus de détails). Ces exemples sont tirés de la tâche de lecture. Les surlignés jaunes indiquent que l'extrait est aligné avec la portion du signal sonore (oscillogramme en bas).

L'alignement entre la transcription orthographique et le signal de parole mérite une attention particulière. Chaque ligne dans l'éditeur correspond à ce qu'on appelle ici un « énoncé », c'est-à-dire une séquence

correspondant à une clause (phrase verbale entourée de son sujet et de tous les éléments disjoints à sa gauche/droite qui en dépendent). Cette définition de l'énoncé a permis d'aligner aisément la transcription orthographique avec le signal de parole dans les tâches de lecture oralisée d'autant plus que la ponctuation forte coïncidait avec une frontière syntaxique majeure. En revanche, cette notion d'énoncé se révèle très problématique lorsqu'il s'agit d'aligner la transcription orthographique et le signal de parole dans les tâches de production orale non contrôlées (descriptions d'images, interview, etc.). En effet, découper en séquences le signal sonore afin d'apparier ces extraits à ce qu'on entend par « énoncés » ou « clauses » est extrêmement difficile. Nous avons donc opté d'aligner la transcription orthographique avec la séquence formée par une clause, même si à l'intérieur de celle-ci des pauses y figuraient. Par exemple, dans la séquence « on est allé en Normandie (.) euh donc j' ai essentiellement vécu en Normadie (.) Alançon (.) Falaise (.) Saint+lô +… » produit par un locuteur français natif en parole spontané, nous avons

Fig. 1. *Copie d'écran de l'éditeur de texte sous CLAN montrant l'alignement de la transcription orthographique et le signal de parole en français L2.*

distingué deux clauses, les « (.) » représentant des pauses : *on est allé en Normandie* et *euh donc j'ai essentiellement vécu en Normandie Alençon Falaise Saint Lô*.

Bien que la question de la segmentation en unités fonctionnelles pour la transcription orthographique gagnerait à être approfondie et discutée (*cf*. Benazzo & Wątorek dans ce volume), nous ne nous attarderons pas beaucoup sur ce point. Dans notre cas, la transcription orthographique et son alignement avec le signal sonore ne sont là que pour faciliter l'analyse des données *a posteriori* pour les études prosodiques et phonétiques. Il faut en tout cas remarquer que la transcription orthographique passe nécessairement par une étape d'analyse et d'interprétation du signal sonore. Cette interprétation se révèle moins problématique lorsqu'il s'agit de transcrire les tâches de lecture. Dans ce cas, en effet, les participants sont censés reproduire ce qui est déjà écrit. Dans le cas de la parole spontanée, reconstruire ce qui a été dit dans le signal de parole est une tâche plus compliquée, surtout lorsqu'il s'agit de la parole des apprenants.

2.3 Annotations phonologiques

Même si les données ont été transcrites au niveau orthographique, il peut être utile d'avoir accès à des informations relatives à la forme sonore. Se pose alors la question de savoir quelles informations il faut extraire du signal de parole : les réalisations phonétiques concrètes ou plutôt les formes phonologiques ? Lorsque le chercheur opte pour la transcription phonologique, cela suppose qu'il a déjà découvert le système phonologique de la variété sur laquelle il travaille (Durand & Tarrier 2006 : 141). Si cela n'est pas forcément un problème en L1, ce n'est pas la même chose lorsqu'on analyse la parole des apprenants, le système phonologique qui est sous-entendu est en effet émergeant et méconnu par le chercheur. Puisque nous travaillons sur des données d'apprenants, nous ne pouvons pas à proprement parler de faire une transcription phonologique. Nous avons plutôt opté d'adapter une transcription phonétique de la L1 et avons rejeté une transcription phonétique authentique qui caractérise la prononciation des étudiants.

Il faut noter, cependant, qu'une transcription phonétique authentique de l'ensemble des données procure plusieurs avantages. D'une part, elles fournissent des représentations symboliques de ce qui a effectivement été prononcé par les locuteurs non natifs, et non une abstraction phonologique

qui serait forcément erronée puisque le système des apprenants n'est pas connu. D'autre part, elles rendent mieux compte de ce qui a été dit par les apprenants que la transcription orthographique. En outre, avoir une transcription phonétique authentique permettrait au chercheur de croiser des informations plus fines : les compétences prosodiques en L2 sont corrélées au degré de maitrise de la prononciation des consonnes et des voyelles ?

Cependant, s'engager dans une transcription phonétique des données non natives à partir de l'oreille du transcripteur entraîne une grande subjectivité, d'autant quand le transcripteur ne connaît pas les possibles formes déviantes en L2. Entreprendre la même chose à partir des informations acoustiques enlève la subjectivité, mais cette tâche est assez lourde et coûteuse en temps.

L'annotateur est confronté au choix du degré de finesse qu'il veut atteindre dans la transcription phonétique. Pour s'en convaincre, voilà quelques exemples. Observons les spectrogrammes des suites *m'indiquer* et *si je dois partir* par deux apprenants hispanophones du niveau A2 et B1 respectivement (tâche de lecture oralisée).

Dans la Figure 2, une analyse portant sur la première syllabe de la suite *m'indiquer* nous permet de constater que la production canonique [mɛ̃dike] est remplacée par [maⁿdike]. Cette figure illustre bel et bien l'absence d'anti-formants nasals caractéristiques du [ɛ̃] et montre plutôt

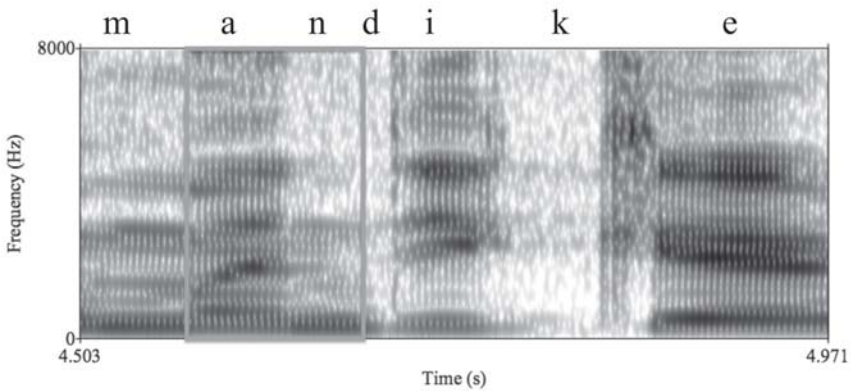


Fig. 2. S*pectrogramme de la suite transcrite comme* m'indiquer *en français L2 produite par une locutrice FL2 niveau A2, le carré enfermant les phones [a] et [n].*

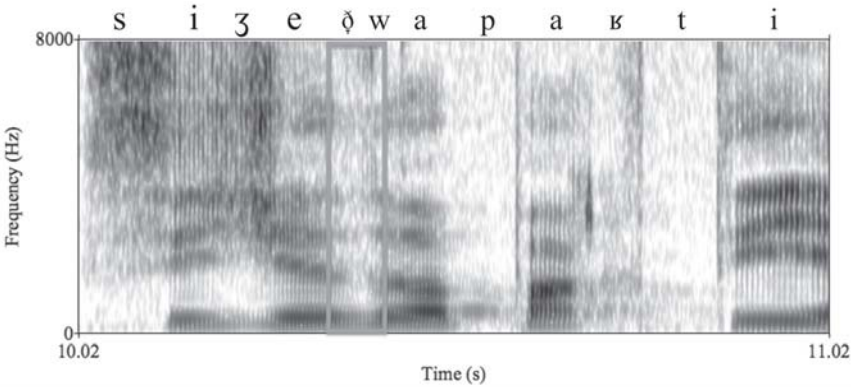


Fig. 3. *Spectrogramme de la suite transcrite comme* si je dois partir *en français L2 produite par un locuteur FL2 niveau B2, le carré enfermant le phone [ð̞] typique de l'espagnol.*

l'articulation de la voyelle [a] qui propage ses transitions formantiques de F1, F2 et F3 vers la nasale [n̪] qui, à son tour, est dentalisée avec le contacte du [d]. Dans la Figure 3, l'articulation du mot *dois* de la suite *si je dois partir* montre l'absence de l'occlusion dentale, celle-ci étant remplacée par une aperture plus ou moins importante des organes articulatoires, en l'occurrence, l'apex de la langue et les dents supérieures. La présence des transitions formantiques de la voyelle [e] précédente nous conduit à affirmer que cette consonne a été articulée comme une approximante dentale [ð̞] et non comme une occlusive [d] (sans doute à cause d'un transfert de l'espagnol).

Transcrire avec fidélité les deux segments ci-dessus devient une tâche extrêmement lourde lorsqu'il s'agit d'un large corpus comme celui qui est traité ici où plus de 100 mille phonèmes sont à encoder manuellement avec ce degré de finesse ! Face à ce problème, les réalisations phonétiques des apprenants peuvent être encodées à partir des informations orthographiques, c'est-à-dire à partir d'une phonétisation reposant sur la phonologie du système. Ainsi, les segments peuvent être transcrits en n'utilisant que les symboles phonétiques du système français. Ainsi, la transcription donnée pour la suite illustrée dans la Figure 2 a été [mɛ̃dike] et non [maɲdike], et pour la Figure 3 [siʒədwapaʁti] et non [siʒeð̞wapaʁti].

Cette méthode présente un avantage lorsque la transcription doit se faire de façon semi-automatique. A partir de la tire contenant la transcription

orthographique sous CLAN, on peut générer des tiers sous le logiciel *Praat* (Boersma & Weenink 2018) afin de générer des transcriptions phonétiques automatiques. Par exemple, une segmentation en mots, syllabes et phones en employant le script *Easyalign* développé par Goldman (2011) peut être menée dans l'ensemble de données en L2. La Figure 4 illustre donc les 5 tires générées sur *Praat* à partir d'*Easyalign.* Nous tenons à faire remarquer que la transcription des mots *parti* et *Marcelle* est ici un exemple des formes morphologiques et prononciations erronées en L2 des mots *partir* et *Marseille* respectivement. Or, selon les conventions sur la transcription de l'orthographe adoptées dans la section 2.2, nous fournissons ici une transcription orthographique la plus proche de ce que l'apprenant a prononcé, mais en utilisant des mots existants en français L1. Les transcriptions phonétiques correspondent donc aux mots en français L1 *parti* et *Marcelle*.

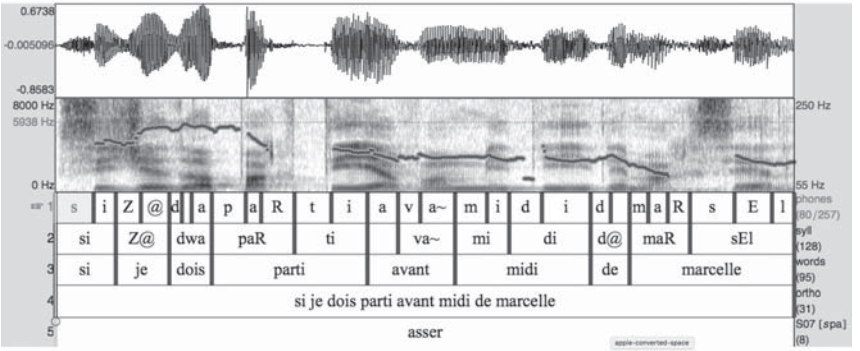


Fig. 4. *Copie d'écran de* Praat *montrant différentes tires encodant les annotations syntaxiques (5), l'orthographe (4), les mots (3), les syllabes (2) et les phones (1) d'un extrait de parole en français L2.*

3. *Transcrire la prosodie en FLE*

Une annotation prosodique consiste à fournir une représentation symbolique et discrète des phénomènes prosodiques observés dans le flux de parole. Lors de la transcription, une grande importance est accordée aux événements qui ont des fonctions linguistiques. Ainsi, l'annotation prosodique doit fournir un encodage symbolique d'une grande variété d'événements tels que les accents, les contours intonatifs, les pauses, etc.

Ces derniers se manifestent généralement à des endroits clefs de la structure prosodique. Sur le plan phonétique, ces événements sont indiqués par des variations de la fréquence fondamentale (f0), dont la force et l'ampleur jouent un rôle, et par des modifications de paramètres temporels comme la durée syllabique.

Pour les données d'apprenants, la transcription envisagée doit permettre d'analyser les systèmes accentuels et intonatifs sous-jacents qui évoluent durant l'apprentissage d'une L2 (Delais-Roussarie & Yoo 2011). Autrement dit, cette annotation doit rendre compte des connaissances de la grammaire de la L2 acquises, des éventuelles règles de la L1 se transposant dans la L2, et pour finir, des propres règles développées dans l'interlangue des apprenants.

Afin d'obtenir une annotation prosodique fournissant un codage symbolique d'une grande variété d'événements prosodiques pour chaque énoncé analysé, le chercheur doit faire face à plusieurs problèmes. Ceux-ci sont dus en grande partie au fait que la transcription des phénomènes prosodiques est très souvent un point épineux largement débattu parmi la communauté scientifique. Cela peut s'expliquer par plusieurs raisons :

- l'étude de l'accentuation ou l'intonation est souvent compliquée par le fait que ces phénomènes sont continus, et donc difficiles à représenter de façon discrète (Post & Delais-Roussarie, & Simon 2006) ;
- les caractéristiques prosodiques varient en fonction d'éléments comme le débit ou l'origine sociogéographique des locuteurs (Post & Delais-Roussarie 2006) ;
- lors d'une transcription prosodique, il faut déterminer quels paramètres acoustiques (f0, intensité et durée) y sont représentés (Delais-Roussarie & Yoo 2011).

Transcrire les événements prosodiques est souvent plus compliqué que coder les segments. Cela résulte sans doute du fait que pour la transcription des segments, qu'il s'agisse d'une analyse phonétique (au niveau des phones) ou phonologique (au niveau des phonèmes), il existe un système de transcription qui semble être accepté par la communauté scientifique : l'API (Alphabet Phonétique International). En revanche, pour les transcriptions prosodiques, aucun système ne fait consensus.

3.1 Quels systèmes de transcription choisir pour la prosodie en L2 ?

L'API comprend un volet qui permet d'encoder différents phénomènes, tels que certains contours mélodiques ou les patrons accentuels. Mais employer le système de transcription prosodique de l'API suppose en général deux choses : (i) le système prosodique à transcrire est connu par l'annotateur ; (ii) les phénomènes prosodiques à annoter sont linguistiquement pertinents. Ces deux caractéristiques font penser que l'emploi de ce système pour l'annotation des événements prosodiques en L2 n'est pas opérationnel : le système prosodique des apprenants est méconnu par l'annotateur et il est difficile de distinguer les phénomènes d'ordre phonologique *vs* phonétique.

D'autres systèmes comme ToBI (*Tones and Break Indices*) ont connu une grande popularité dans la communauté scientifique. Ce système a été développé initialement pour coder l'intonation de l'anglais américain standard (Beckman & Hirschbberg 1994) en distinguant deux types d'informations :

- Le codage des tons selon la théorie métrique et autosegmentale : les contours mélodiques sont représentés comme une séquence linéaire des cibles tonales distinctives représentées avec les symboles L (*Low*)/H (*High*) et leurs combinaisons (HL, LH, etc.).
- Le codage des *Break Index* reflètent la structure prosodique des langues : les unités prosodiques, tels que les groupes accentuels (AP), les syntagmes intermédiaires (ip) ou les syntagmes intonatifs (IP) sont indiqués par des indices numériques apparaissant aux frontières. En outre, ils sont considérés comme se manifestant dans des événements intonatifs qui portent des diacritiques relatifs à leur position dans la hiérarchie prosodique : « * » pour les accents mélodiques ou *pitch accents* (T*), « - » pour les accents de phrases indiquant la frontière des ip (T-), et « % » pour les tons de frontières liés aux IP (T%) (Beckman & Pierrehumbert, 1986).

Ce système est à l'heure actuelle le plus utilisé pour le codage de l'intonation et de la structuration prosodique dans plusieurs langues, dont le français (Delais-Rousarie *et al*. 2015, Jun & Fougeron, 2000), et l'espagnol (de la Mota, Martín Butragueño & Prieto, 2010 principalement) entre d'autres dizaines de langues (*cf*. Jun 2014). Utiliser ce type de

système suppose plusieurs éléments, notamment l'acceptation des concepts centraux de la théorie métrique et autosegmentale que nous n'allons pas détailler ici. En outre, il s'agit d'un système dit phonologique dont l'emploi repose sur quelques conventions clairement établies. Ainsi, les conventions d'annotation doivent témoigner de :

- l'exactitude : quelle que soit la langue annotée, les annotations doivent être fondées sur une analyse rigoureuse de la phonologie intonative, une analyse phonologique de la langue en question étant donc un prérequis.
- l'efficacité : l'annotateur ne devrait transcrire/encoder que les contours mélodiques non distinctifs, qu'ils soient montants ou descendants.

Bien que dans certaines études le système ToBI ait été employé pour coder les événements prosodiques de plusieurs L2 afin de les comparer à la langue cible (Jilka 2000 pour l'allemand L2, Ueyama & Jun 1996 pour l'anglais L2, Gabriel & Kireva 2014 pour l'espagnol L2, Mennen 2004 pour le grec L2.), plusieurs problèmes sont à noter. Le système ToBI repose sur une analyse abstraite. De fait, seuls les événements intonatifs ayant une fonction contrastive sont encodés. En tous cas, cela suppose

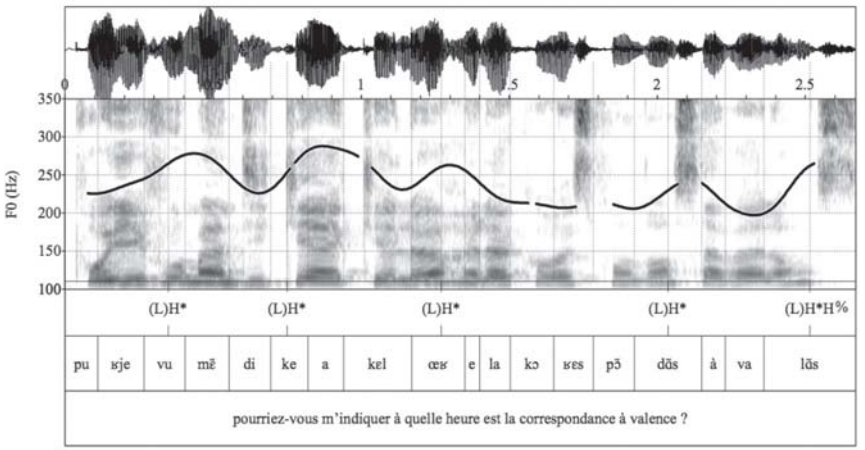


Fig. 5. *Exemple d'annotation de parole en français L1 avec le système ToBI French (Delais-Roussarie* et al*. 2015) d'un énoncé interrogatif.*

que le système intonatif de la langue de l'apprenant soit connu avant que l'annotation prosodique se fasse.

Des exemples de difficultés sont discutés dans les figures qui suivent. Dans la Figure 5, le découpage en groupes accentuels en français L1 est indiqué par la présence des accents mélodiques finals. Les symboles (L)H* représentent un mouvement mélodique dynamique montant à la fin de groupes accentuels (pourriez-vous)$_{GA}$ (m'indiquer)$_{GA}$ (à quelle heure)$_{GA}$ (est la correspondance)$_{GA}$ (à Valence ?)$_{GA}$, et le mouvement H% réalisé à la fin de l'énoncé indique la frontière du syntagme intonatif (en l'occurrence, une question).

Ce codage, et surtout son analyse en unités prosodiques, découle de la phonologie du français et des connaissances connues par l'annotateur : (i) en français, il y a des accents mélodiques (T*) qui marquent la frontière des groupes accentuels , (ii) le phrasé prosodique contraint le choix du contour intonatif (la force du mouvement mélodique dépend de son statut dans la hiérarchie prosodique, selon qu'il s'agisse d'un groupe accentuel ou un syntagme intermédiaire/intonatif, et (iii) en position positon prénucléaire, les mouvements mélodiques sont généralement montants (H*, H- ou H%). Ils sont encodés par les symboles LH ou H (Delais-Roussarie, et al. 2015, Di Cristo 2016).

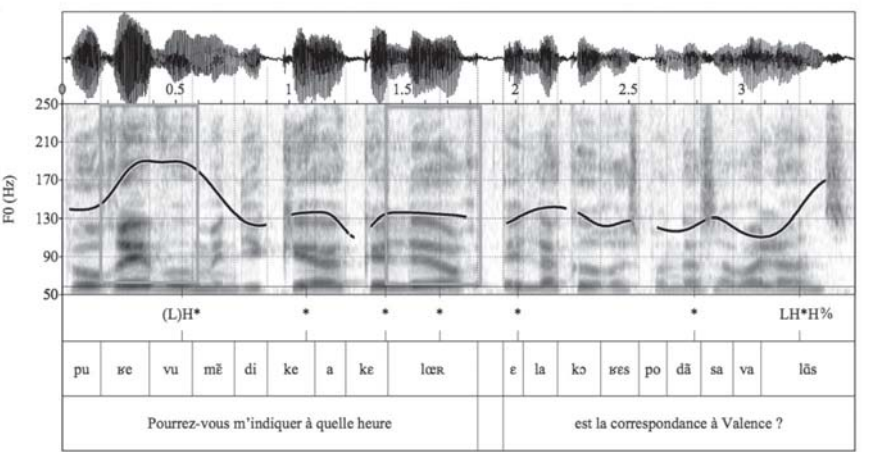


Fig. 6. *Exemple d'annotation de parole en français L2 avec le système ToBI French d'un énoncé interrogatif produit par un apprenant hispanophone.*

Que faire en revanche pour encoder les productions des apprenants qui ne suivent pas nécessairement les règles phonologiques de la langue cible ? Si on compare la réalisation d'un locuteur natif illustré dans la Figure 5 à celle d'un apprenant hispanophone dans la Figure 6, on voit au moins deux différences importantes. Dans la Figure 6, il y a un mouvement mélodique sur la syllabe finale du mot *pourriez*, mais également un ton haut sur la syllabe vous (premier rectangle rouge).

Comment traiter cette réalisation ? Comme une séquence de deux accents mélodiques sur deux syllabes adjacentes (LH* suivi de H*) ou comme un accent initial suivi d'un ton haut ? S'agit-il d'un transfert concernant un positionnement erroné de l'accent de groupe qui est sur la pénultième en espagnol L1 ? Les choix ne sont pas simples, d'autant qu'ils ont des implications sur l'analyse des faits. De la même façon, comment doit-on encoder l'absence de mouvement mélodique montant sur la syllabe finale de « indiquer » et de « heure » (deuxième rectangle dans la Figure 6) : par une absence d'accent mélodique qui va de pair avec une absence de réalisation des groupes accentuels ? Cela refléterait en partie les patrons mélodiques en vigueur dans la L1 de l'apprenant, et pourrait résulter d'un transfert (*cf*. Santiago & Delais-Roussarie 2015 où l'absence d'accents mélodiques en position interne dans l'énoncé interrogatif en espagnol est fréquente). Mais on peut aussi penser que les apprenants réalisent un accent à l'aide de la durée et non pas avec un mouvement mélodique, et que les découpages des groupes accentuels seraient alors satisfaisants.

Cela nous amène au deuxième problème avec le système ToBI. Les étiquettes et événements retenus sont limités surtout aux variations mélodiques de f0. Cela écarte par exemple le codage des phénomènes liés à des variations de durée ou d'intensité, paramètres acoustiques pouvant être importants pour le codage de différents contours mélodiques en français L1 (Martin 2012) ou exploités différemment par les apprenants d'une L2. Sur ce dernier cas, par exemple, à l'écoute de l'exemple illustré dans la Figure 6, l'apprenant semble marquer le phrasé prosodique (deuxième rectangle) du groupe accentuel *à quelle heure* par un allongement de la voyelle de la syllabe finale du mot *heure* à la place de la modulation de la voix, celle-ci étant la plus exploitée en français L1. De fait, l'emploi de l'allongement syllabique pour marquer les groupes accentuels peut être vu comme licite, car l'allongement des syllabes est relié au marquage prosodique des phénomènes d'ordre métriques, et non pas intonatifs.

Le même cas peut être vu dans la Figure 7, où un allongement de la syllabe finale du groupe accentuel *nous avons mangé* est observé. Ces paramètres prosodiques doivent être considérés dans une transcription phonétique en L2 afin d'évaluer dans quelle mesure les locuteurs non natifs exploitent les paramètres acoustiques de la même façon que les natifs. Ces exemples montrent bel et bien que le marquage des unités prosodiques, tels que les groupes accentuels, peut émerger exclusivement avec les variations temporelles en L2. Or, le système ToBI pour le français ne considère pas ces facteurs.

Pour finir, comme dans ce système d'annotation la définition des unités prosodiques repose tantôt sur des informations morphosyntaxiques tantôt sur les formes tonales des mouvements mélodiques, cela peut compliquer les annotations prosodiques de certaines réalisations observées chez les apprenants. En effet, il est généralement admis que la force d'un mouvement mélodique montant associé à la frontière droite d'un groupe accentuel en français a une importance mineure comparée à celle des contours montants bornant les énoncés interrogatifs ou les syntagmes intonatifs en position non finale associés à ladite « continuation majeure ». En d'autres termes, la force de la pente mélodique des contours montants bornant un syntagme intonatif est généralement plus ample que celle d'un accent mélodique bornant un groupe accentuel (Di Cristo 2016). Prendre ces critères pour encoder la prosodie pose un problème lorsqu'il s'agit d'encoder la parole des apprenants.

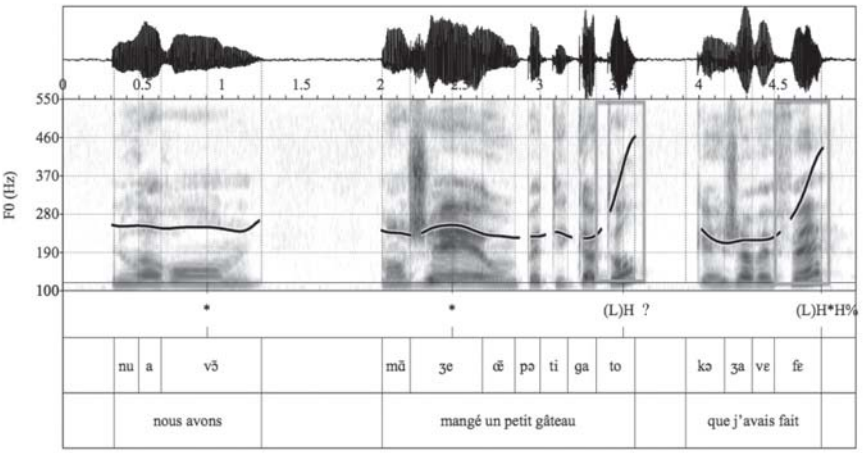


Fig. 7. *Exemple d'annotation de parole en français L2 avec le système ToBI French Delais-Roussarie et al. (2015) d'un énoncé assertif.*

Par exemple, dans la Figure 7, la force de la montée mélodique du groupe accentuel (un petit gâteau)$_{GA}$ est plus ample que celle bornant la frontière droite d'un syntagme intonatif non final [(nous avons mangé)$_{GA}$ (un petit gâteau)$_{GA}$ (que j'avais fait)$_{GA}$]$_{SI}$. Or, il est clair qu'en français une montée exagérée bornant un groupe accentuel dans une position pré-nucléaire n'est pas communément réalisée comme telle dans les données des natifs. L'annotateur se voit donc dans l'incertitude d'encoder cette montée exagérée comme la réalisation d'un groupe accentuel ou celle d'un syntagme intonatif. Cela aurait des implications sur l'analyse : erreur d'implémentation phonétique des contours montants des groupes accentuels ou erreur phonologique relative aux découpages prosodiques ?

Pour toutes ces raisons, nous avons considéré qu'utiliser un système ToBI adapté au français n'était pas convenable pour le codage de nos données en français L2. En revanche, nous considérons qu'une adaptation de ce système doit être envisagée si l'on veut comparer les réalisations natives et non natives.

En plus des deux systèmes de transcription mentionnés ci-dessus, phonologiques dans leur essence, d'autres systèmes de transcription sont utilisés pour encoder les événements prosodiques, dont des systèmes automatiques, notamment le système Momel-INTSINT (*International Transcription System for INTonation*) développé par Hirst (2007). Faute de place, nous renvoyons le lecteur à l'ouvrage de Santiago (2014) afin de réfléchir aux avantages et inconvénients de ce système de transcription.

Les problèmes d'utiliser des systèmes d'annotation tels que l'API ou ToBI en français L2 posent également un problème de par le fait que la précision des techniques d'annotation automatique utilisant ces systèmes d'annotation n'est encore pas satisfaisante, et que le travail d'annotation est donc nécessairement manuel. Bien que certains travaux essaient d'explorer l'automatisation de ces systèmes de transcription (Rosenberg 2010), à notre connaissance, ces systèmes n'ont pas encore été testés dans la parole non native. En revanche, d'autres outils d'annotation prosodique, mais non des systèmes d'annotation à proprement parler, peuvent pallier ces problèmes.

3.2 Vers un outil semi-automatique pour transcrire la prosodie en FLE

Un outil d'annotation prosodique pour des données non-standards a comme objectif de fournir un étiquetage automatique des patrons prosodiques du signal de parole lorsque les systèmes accentuels et intonatifs

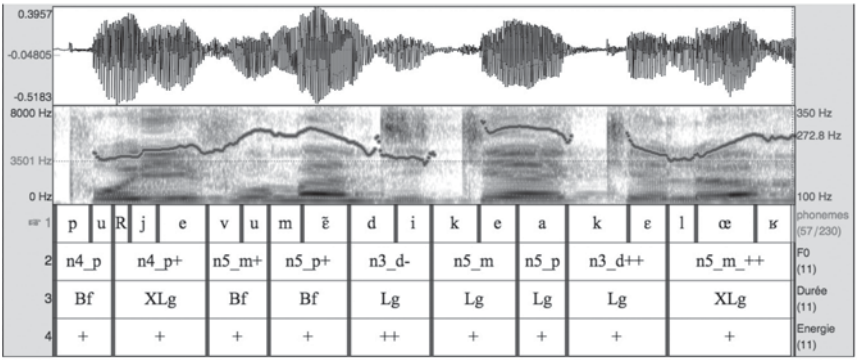


Fig. 8. *Codage fournit par PROSOTRAN de l'énoncé* Pourriez-vous m'indiquer à quelle heure *en français L1*.

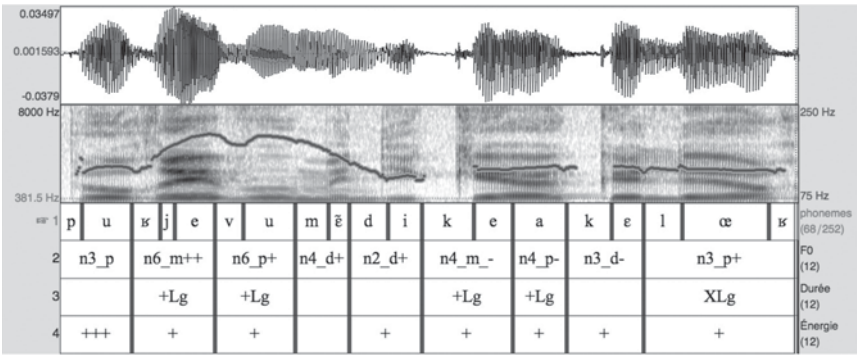


Fig. 9. *Codage fournit par PROSOTRAN de l'énoncé* Pourriez-vous m'indiquer à quelle heure *en français L2.*

des langues à annoter sont méconnus par le transcripteur. L'un de ces exemples et l'outil PROSOTRAN (Bartkova & Delais-Roussarie & Santiago 2012). Il a été conçu pour faire des comparaisons entre le codage prosodique des données en L2 et celui des données en L1. Cet outil est encore en phase de développement. PROSOTRAN fournit un codage de trois paramètres acoustiques : (i) les variations mélodiques de f0 (en termes de hauteur tonale, de direction du mouvement mélodique par rapport à la syllabe précédente et d'ampleur du mouvement selon le seuil du glissando) ;

(ii) les variations temporelles et (iii) les variations d'intensité. Un exemple de l'annotation que fournit PROSOTRAN pour le codage des énoncés est illustré dans les Figures 8 et 9. Dans ces figures, dans la tire « F0 », nous trouvons les informations concernant les variations mélodiques, en l'occurrence, le niveau atteint (codées comme n3, n6…, n= niveau tonal), la direction du mouvement (p = plateau, m=montant, d=descendant) et le seuil du glissando (+, ++…). Dans la tire « Durée », nous trouvons l'étiquetage du rallongement vocalique : bref (Bf), Long (Lg) ou extralong (XLg). Enfin, dans la tire 4, nous trouvons le codage concernant l'énergie (+, ++…).

Le codage fournit par PROSOTRAN permet de faire émerger des événements prosodiques se réalisant par des variations affectant plusieurs paramètres conjointement. Cela permet au chercheur d'analyser les indices acoustiques qui peuvent être exploités différemment par les locuteurs venant de différents profils linguistiques, et ce, de manière indépendante. C'est l'exemple du marquage prosodique du groupe accentuel *à quelle heure* : en français L1, cette syllabe est encodée avec un mouvement mélodique montant (seuil de glissando) et un allongement important, tandis qu'en français L2, la même syllabe ne reçoit aucun étiquetage concernant un mouvement de la direction de f0 (plateau) mais seulement celui de la variation temporelle (syllabe extra-allongée). Comme nous pouvons le constater, un outil de transcription comme celui qui est présenté ici permettrait de pallier certaines insuffisances que nous avons rencontrées dans beaucoup de systèmes de transcription prosodiques. Un avantage du PROSOTRAN est qu'il ne demande pas certaines connaissances linguistiques comme la segmentation et l'annotation de la parole et permet d'encoder plusieurs dimensions prosodiques de manière indépendante. En outre, PROSOTRAN pouvant générer un étiquetage automatique, il serait aisé de comparer une quantité relative des énoncés (en l'occurrence, de la tâche de lecture) afin de pouvoir faire émerger des patrons communs chez les apprenants.

Comme le calcul conduisant au codage se fait à partir des valeurs associées aux paramètres physiques, le système peut être utilisé aussi bien pour annoter des données dans des langues dont le fonctionnement prosodique est connu (L1) que des données dont on ne connaît pas la grammaire sous-jacente, comme c'est le cas de la parole en L2. Enfin, un tel outil fournit une annotation prosodique même si aucune connaissance linguistique de la langue à annoter n'est disponible.

En revanche, le calcul de la durée moyenne de toutes les voyelles ne permet pas de distinguer les différents niveaux de frontières prosodiques en français L1. En outre, l'outil n'a pas été testé pour le codage de la parole spontanée en L2 ou les variations temporelles sont beaucoup plus instables que dans la parole lue.

Un dernier outil qui mérite notre attention est le système Prosogramme (Mertens, 2004) qui fournit une stylisation automatique du tracé de f0 en rapport avec le seuil de perception. Cette stylisation a l'avantage de donner une représentation des événements mélodiques complètement indépendante du système langagier à encoder, et cela que le système intonatif sous-jacent soit méconnu ou non. Le Prosogramme est donc un système de transcription qui se veut neutre. A la différence du système INTSINT, le Prosogramme tient compte des impacts auditifs et perceptifs des paramètres acoustiques analysés aux noyaux syllabiques. En d'autres termes, le Prosogramme offre *une représentation de l'intonation perçue* (Mertens, 2004 : 111).

Les objectifs du Prosogramme sont les suivants (cf. Mertens, 2004 : 112) : (i) fournir une représentation objective et fiable de la prosodie facile à interpréter, et ce, de manière automatique utilisable dans de grands corpus, (ii) considérer l'évolution de la hauteur de f0 en termes de déclinaison, d'attaque, de registre et de changement de registre, et (iii) analyser les facteurs temporels en tenant compte des pauses, hésitations, de déterminer le débit et les aspects rythmiques comme les accélérations ou les ralentissements.

Ce système de transcription est fondé sur une simulation de la perception de la hauteur mélodique tout en considérant le noyau syllabique comme unité de base. Il prend en considération le seuil de Glissando Différentiel ou DG (Alessandro & Mertens 1995). Ce seuil établit que les variations de la fondamentale doivent présenter une ampleur minimale qui varie en fonction de la fréquence de départ et de la durée du stimulus pour qu'elles soient perçues. Le système fournit ainsi une stylisation de la courbe mélodique en fonction des phénomènes fonctionnels ou audibles. Dans ce système, tout changement de pente est comparé au seuil de DG, et s'il est inférieur au seuil, le changement de pente est considéré comme inaudible.

Cette stylisation est très proche de la notation manuelle, au moins pour coder les événements de *glissando*. La validation de la stylisation obtenue par le Prosogramme peut se réaliser avec la resynthèse du signal de parole utilisée. À cette fin, le Prosogramme reprend toutes les caractéristiques

du signal original, sauf la f0, pour laquelle la stylisation calculée est remplacée. Cela permet à l'annotateur de vérifier à l'écoute du signal resynthétisé l'évaluation de la correspondance de la représentation auditive obtenue par le Prosogramme et une transcription manuelle de la part de l'annotateur. Dans la figure 10, nous illustrons une copie d'écran du mode interactif de la stylisation obtenue du Prosogramme de la phrase *Pourriez-vous m'indiquer à quelle heure est la correspondance à Valence ?* en français L2 (donnée préalablement). Cette image illustre la fenêtre interactive montrant la resynthèse du contour mélodique stylisé. Dans le Prosogramme, la ligne continue verte représente l'intensité, le trace fin bleu discontinu la f0 et les traits épais noirs superposés sa stylisation. Les valeurs de f0 sont données en demi-tons (relatifs à 1 Hz). Les lignes horizontales pointillées représentent les changements tonals à 2 demi-tons.

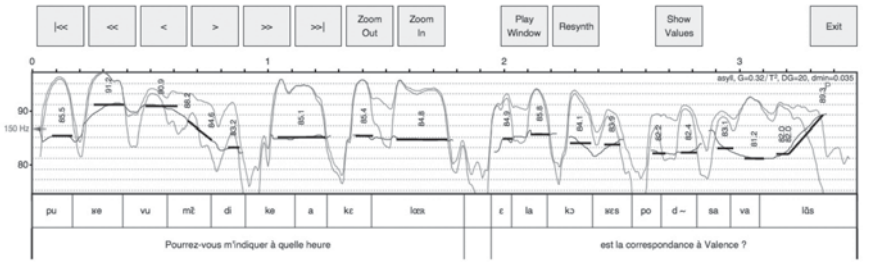


Fig. 10. *Exemple de resynthèse avec le Prosogramme appliqué à un énoncé en français L2.*

Le Prosogramme a en revanche certaines limites. D'une part, il ne fournit que des valeurs de la stylisation de la courbe mélodique et ne fournit aucune représentation symbolique de la nature de ces mouvements. Cependant, ces valeurs peuvent permettre au chercheur de comparer, de façon semi-automatique, où se trouvent les mouvements mélodiques importants dans l'énoncé, leur forme et leur ampleur. C'est en analysant ces patrons de manière récurrente que le chercheur peut ensuite tirer des conclusions sur la nature de ces mouvements. Les études de Santiago & Delais-Roussarie (2015) et Santiago & Mariano (2019) ont montré que cet outil s'avère efficace pour comparer de manière semi-automatique les productions en français par des hispanophones et des

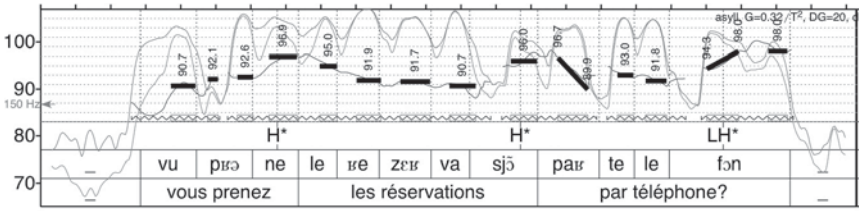


Fig. 11. *Tracée de la stylisation de f0 (ligne noire) et valeurs exprimées en demi-tons obtenues par le Prosogramme d'un énoncé interrogative produit par une locutrice francophone.*

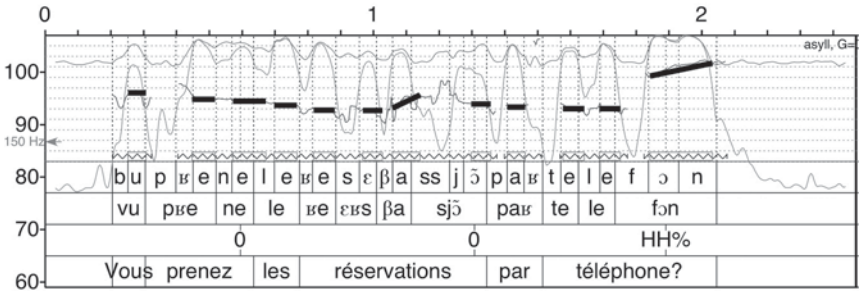


Fig. 12. *Tracée de la stylisation de f0 (ligne noire) et valeurs exprimées en demi-tons obtenues par le Prosogramme d'un énoncé interrogative produit par une étudiante hispanophone de FLE.*

anglophones. De surcroît, cet outil est parfaitement compatible avec des annotations manuelles qui peuvent relever d'autres systèmes comme ToBI.

C'est l'exemple de l'étude de Santiago & Delais-Roussaire (2015) portant sur les groupes accentuels en français L2. L'énoncé *Vous prenez les réservations par téléphone ?* contient trois GA potentiels : [Vous prenez]$_{GA}$ [les réservations]$_{GA}$ [par téléphone]$_{GA}$ (cf. Di Cristo 2016; Jun & Fougeron, 2002). Le codage des GA apparaît dans la Figure 11, selon la stylisation obtenue par le Prosogramme et l'annotation manuelle par les auteurs pour le cas d'un francophone natif. Dans la Figure 12, le Prosogramme et l'annotation ToBI illustrent le même énoncé produit par une étudiante hispanophone. Notons que les étiquettes H* et LH*/HH% sont associés à des contours mélodiques hauts (non dynamiques) et montants (dynamiques)

respectivement. A partir de ces exemples, nous pouvons constater clairement que, dans le cas d'un apprenant hispanophone, les GA ne sont pas marqués prosodiquement : cet énoncé a comme caractéristique générale un contour mélodique final très montant sans avoir aucune variation tonale perçue en position interne (indiqué par 0). Enfin, le Prosogramme fournit également les valeurs temporelles des toutes les syllabes analysées afin de pousser les analyses temporelles.

4. *Bilan et conclusion*

Cette contribution avait pour objectif d'exposer les critères théoriques et méthodologiques pour étudier la prosodie en français L2. Nous avions insisté sur le fait que la constitution d'un corpus ne se réduit aucunement à un simple recueil de données numérisées. En effet, plusieurs étapes d'ordre méthodologique ont été menées afin de documenter et d'annoter le corpus. Ces étapes comprennent : le type de données collectées, la transcription orthographique du corpus et les annotations phonologiques et prosodiques. La réalisation de ces étapes a soulevé plusieurs questions d'ordre théorique et méthodologique.

Plus particulièrement, la transcription orthographique et les annotations phonologiques/phonétiques ont mérité un point de réflexion important. Les codages orthographique et phono-phonétique obéissent à des contraintes d'ordre méthodologique pour l'exploitation des données *a posteriori*. Sans trop s'attarder sur ce point, le chercheur doit être conscient que cette approche pragmatique permet de se focaliser sur la prosodie en L2. En revanche, les choix pris ne permettent pas faire des analyses plus fines entre les compétences phonologiques (consonnes et voyelles) et prosodiques.

Pour étudier la prosodie d'une L2, une annotation prosodique doit être effectuée lors de l'étude du corpus. Afin que l'annotation des patrons prosodiques des apprenants puisse être comparée à celles des productions des natifs, un système de codage prosodique neutre ne tenant pas compte du statut phonologique/phonétique des événements prosodiques doit être conçu. Il ressort de l'analyse portant sur plusieurs systèmes de transcription prosodique actuels, tels que l'API ou ToBI, qu'aucun de ces systèmes ne sont pas adaptés pour coder les phénomènes prosodiques en L2. Plusieurs inconvénients émergent lorsqu'on veut adopter ces systèmes pour le codage de la prosodie en L2. Des systèmes comme l'API ou ToBI

ont été essentiellement conçus pour annoter des événements contrastifs qui impliquent, nécessairement, la connaissance du système langagier à annoter. Or, l'interlangue des apprenants est un système méconnu par le chercheur. En outre, des systèmes comme ToBI ont été conçus pour la modélisation de la variation mélodique sans tenir compte des variations temporelles. Lors de l'analyse de nos données, nous avons souligné que certains paramètres acoustiques, comme la durée, peuvent être utilisés chez les apprenants comme des stratégies compensatoires pour marquer la structure prosodique dans la L2. Il a donc été évident que fonder une analyse exclusivement sur l'observation des mouvements mélodiques pourrait réduire l'analyse de la compétence prosodique en L2.

Au vu des contraintes observées dans les systèmes de transcription existants, nous avons discuté des outils d'annotation prosodique automatiques basés sur des critères perceptifs ou acoustiques : Prosotran et le Prosogramme. Ce dernier a l'avantage de fournir une stylisation de la courbe mélodique à partir du seuil de *glissando* en offrant une image acoustique de la perception des mouvements mélodiques du signal sans tenir compte des connaissances phonologiques de la langue à annoter. Nous avons expliqué comment ce système permet d'évaluer les productions des apprenants et des natifs sans pour autant faire référence à une théorie linguistique ou phonologique en particulier.

Le codage exemplifié ici s'est inspiré des symboles employés dans l'approche métrique et autosegmentale, et plus particulièrement dans les systèmes ToBI appliqués au français et à l'espagnol. En nous appuyant sur les étiquettes du système ToBI et en les adaptant, nous avons présenté les avantages d'un système semi-automatique pour représenter les phénomènes phonologiques et leurs manifestations phonétiques observés dans les données des apprenants, notre but étant de faire émerger les catégories phonologiques de l'interlangue. Lors de cette adaptation, nous avons clarifié ses limites. Nous avons ainsi mentionné que l'adaptation de tels systèmes ne reflète pas non plus les variations temporelles. Bien que les valeurs temporelles fournies par le Prosogramme puissent compléter cette analyse, les annotations prosodiques sont pour l'instant semi-automatiques. Enfin, des outils comme Prosotran doivent être pilotés et améliorés afin de comparer les représentations des deux logiciels. Enfin, il est clair que, pour rendre compte des règles prosodiques des apprenants, il faut que ces analyses soient réalisées dans différents styles de parole et avec des données construites et authentiques.

Bibliographie

(d')Alessandro, Christophe & Mertens, Piet. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language* 9 (3). 257-288.

Bartkova, Katarina & Delais-Roussarie, Élisabeth & Santiago, Fabian. 2012. PROSOTRAN: a tool to annotate prosodically non-standard data. In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel J. (eds.), *Proceedings of 6th International Conference on Speech Prosody,* 55-58. Shanghai : Tongji University Press.

Beckman, Mary & Hirschberg, Julia. 1994. The ToBI Annotation Conventions. Disponible sur http://www.ling.ohiostate.edu/~tobi/ame_tobi/annotation_conventions.html.

Bilger, Mireille. 2007. Réflexions sur un obscur objet de désir : le corpus. *Cahiers de l'Association for French Language Studies (e-journal)*. 13 (1). 2-17.

Blanche-Benveniste, Claire & Jeanjean, Colette. 1987. *Le français parlé*. Paris : Didier Erudition.

Boersma, Paul & Weenink, David. 2018. *Praat: doing phonetics by computer* [Computer program], Version 6.1.09, retrieved 01 January 2018 from http://www.praat.org/

Colantoni, Laura & Steele, Jeffrey & Escudero, Paola. 2015. *Second Language Speech. Theory and Practice*. Cambridge: Cambridge University Press.

De la Mota, Carme & Martín Butragueño, Pedro & Prieto Pilar. 2010. Mexican Spanish Intonation. In Prieto, Pilar & Roseano, Paolo (eds.), *Transcription of Intonation of the Spanish Language*, 319-350. München: Lincom Europa.

Delais-Roussarie, Élisabeth. 2003. Constitution et annotation de corpus : méthodes et recommandations. In Delais-Roussarie, Élisabeth & J. Durand, Jacques (éds), *Corpus et variation en phonologie du français : méthodes et analyse*, 127-157. Toulouse : Presses Universitaires du Mirail.

Delais-Roussarie, Élisabeth & Yoo, Hi-Yon. 2011. Learner corpora and prosody : from the COREIL corpus to principles on data collection and corpus design. *Poznań Studies in Contemporary Linguistics* 41 (1). 26-39.

Delais-Roussarie, Élisabeth *et al*. 2015. Intonational phonology of French: Developing a ToBI system for French. In Frota, Sonia & Prieto, Pilar (eds.), *Intonation in Romance*, 63-100. Oxford : Oxford University Press.

Detay, Sylvain & Durand, Jacques & Laks, Bernard & Lyche, Chantal (éds.). 2010. *Les variétés du français parlé dans l'espace francophone*. Paris: Ophrys.

Di Cristo, Albert. 2016. *Les musiques du français parlé : Essais sur l'accentuation, la métrique, le rythme, le phrasé prosodique et l'intonation du français contemporain*. Berlin : De Gruyter.

Durand, Jacques & Laks, Bernard & Lych, Chantal. 2009. Le projet PFC : une source de données primaires structurées. In Durand, Jacques & Laks, Bernard & Lyche, Chantal (éds), *Phonologie, variation et accents du français*, 19-61. Paris : Hermès.

Durand, Jacques & Tarrier, Jean-Michel. 2006. PFC, corpus et systèmes de transcription. *Cahiers de Grammaire*. 30. 139-158.

Gabriel, Chritoph. & Kireva, Elena. 2014. Prosodic transfer in learner and contact varieties. *Studies in Second Language Acquisition.* 36. 257-281.

Goldman, Jean-Philippe. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*. Florence, Italy, August 27-31, 2011.

Gut, Ulrike. 2007, Learner corpora in second language prosody research and teaching. In Trouvain, Jürgen & Gut, Ulrike (eds), *Non-Native Prosody. Phonetic Description and Teaching Practice*, 145-167. Berlin: Mouton de Gruyte.

Hirst, Daniel. 2007. A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. In Barry, William G. & Trouvain, Jürgen (eds.). *Proceedings of the XVIth International Conference in Phonetic Sciences*, Saarbrücken, Germany, 6-10 August 2007, 1233-1236. Saarbrücken : Universität des Saarlandes.

Jilka, Matthias. 2000. *The contribution of intonation to the perception of foreign accent.* Thèse de doctorat, Universität Stuttgart.

Jun, Sun-Ah (ed.). 2014. *Prosodic Typology II. The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press.

Jun, Sun-Ah & Fougeron, Cécile. 2002. Realizations of accentual phrase in French intonation. *Probus* 14. 147-172.

Leather, Jonathan. 1999. Second-Language Speech Research: An Introduction. *Language Learning* 49. 1-56.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition, Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, Philippe. 2012. The Autosegmental-Metrical Prosodic Structure: not fit for French? In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel J. (eds.), *Proceedings of 6th International Conference on Speech Prosody*, 131-134. Shanghai : Tongji University Press.

Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32. 543-563.

Mertens, Piet. 2004. The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In Bel, Bernard & Marlien, Isabelle (ed.), *Proceedings of the International Conference on Speech Prosody 2004*, *Nara, Japan, March 23-26*.

Post, Brechtje. & Delais-Roussarie, Élisabeth. 2006. Transcribing intonational variation at different levels of analysis. In Hoffmann, Rüdiger & H. Mixdorff, Hansjörg (ed.), *Proceedings of the International Conference on Speech Prosody, Dresden, May 2-5 2006*, Dresden : TUDpress.

Post, Brechtje & Delais-Roussarie, Élisabeth & Simon, Anne-Catherine. 2006. Développer un système de transcription des phénomènes prosodiques. In Simon, Anne-Catherine & Caelen-Haumont, Geneviève & Pagliano, Claudine (éds), *Bulletin PFC*. Vol. 6. 51-68.

Racine, Isabelle & Detey, Sylvain & Zay, Françoise & Kawaguchi, Yuji. 2012. Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet « Interphonologie du français contemporain » (IPFC). In Kamber, Alain & Skupien Dekens, Carine (éds.), *Recherches récentes en FLE*, 1-19. Bern : Peter Lang.

Racine, Isabelle & Zay, Françoise & Detey, Sylvain & Kawaguchi, Yuji. 2011. De la transcription de corpus à l'analyse interphonologique: enjeux méthodologiques en FLE. In Col, Gilles & Osu, Sylvester N. (éds.), *Transcrire, écrire, formaliser*, 13-30. Rennes : PUR. (Travaux Linguistiques du CerLiCO, 1).

Rosenberg, Andrew. 2010. AuToBI - a tool for automatic ToBi Annotation, *INTERSPEECH 2010*, Makuhari, Chiba, Japan, September 26-30, 2010.

Santiago, Fabian. 2014. *Systèmes prosodiques et acquisition d'une L2 : production et perception des mouvements mélodiques en français et en espagnol.* PhD Thesis. Université de Paris (Sorbonne Paris Cité).

Santiago, Fabian & Delais-Roussarie, Élisabeth. 2015. The acquisition of Question Intonation by Mexican Spanish Learners of French. In Delais-Roussarie, Élisabeth & Avanzi, Mathieu & Herment, Sophie (eds), *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*, 243–270. Heidelberg: Springer.

Santiago, Fabian & Mairano, Paolo. 2019. Prosodic effects on L2 French vowels: a corpus-based investigation. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019), Melbourne, Australia, 5-9 August 2019,* 1084-1088.

Sinclair, John. 1996. *Preliminary recommendations on Corpus Typology*. Rapport Technique. EAGLES (Expert Advisory Groups on Language Engineering Standards). Pisa : Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

TEI Consortium, eds. *Guidelines for Electronic Text Encoding and Interchange.* [24.06.2020]. http://www.tei-c.org/P5/.

Ueyama, Motoko & Jun, Sun-Ah. 1996. Focus realization of Japanese English and Korean English intonation. *UCLA Working Papers in Phonetics* 94. 110-125.

Vaguer, Céline. 2003. Corpus, vous avez dit corpus ! De la notion de corpus à la création d'un "corpus informatisé. In Williams, Geoffrey Clive (éd.), *Corpus, Langues et Linguistique, Actes des 3es Journées de la linguistique de corpus*, 207-223.

Luciano Romito
(Università della Calabria)

La trascrizione in ambito forense

1. *Introduzione*

Il processo di trascrizione, normalmente, attiene agli studi sul linguaggio e soprattutto sulla lingua parlata. In questa sede ci occuperemo della trascrizione in ambito forense e quindi indirettamente della disciplina di riferimento. La linguistica forense è una disciplina recente che attiene alla linguistica generale, in particolare alla linguistica applicata e alle scienze forensi in genere. Oggi è una disciplina con una propria autonomia sia metodologica che procedurale e si occupa di ogni testo scritto, registrato o anche solo prodotto oralmente, che sia in qualche modo coinvolto in un procedimento legale, penale o in un contesto criminale come sostiene Olsson (2004: 1): "literally if any text is somehow implicated in a legal or criminal context then it is a forensic text".

Pertanto, le competenze richieste ad un trascrittore forense sono interdisciplinari e spaziano dalla linguistica generale alla fonetica sperimentale e dall'acustica alle scienze forensi.

2. *La trascrizione nella prassi giudiziaria*

Il sistema giuridico italiano non fornisce una precisa definizione di *trascrizione di intercettazioni*[1] o di *trascrizione forense*. Essa può essere dedotta attraverso la lettura degli articoli del Codice di Procedura Penale (c.c.p.) dedicati all'*Esecuzione della perizia*, oppure delle sentenze di Appello e di Cassazione prodotte dai giudici.

[1] Per una corretta definizione del termine *intercettazione*, si veda il § 3 *Le intercettazioni*.

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 220[2]]. Le trascrizioni o le stampe sono inserite nel fascicolo per il dibattimento[3].

Per poter giungere ad una definizione di *trascrizione*, è necessario interpretare il concetto di *integrale* e di *in forma intelligibile* presente nel comma 7 art. 268 citato. Per un linguista la comunicazione non è costituita esclusivamente da parole, ma da tutto un complesso di canali paralleli verbali e non verbali[4], quindi *integrale* comprende tutto il canale della comunicazione. Secondo la prassi giuridica, invece, per *integrale* si deve intendere un *intero* progressivo[5] intercettato (senza selezionare

---

[2] L'art. 220 del c.c.p. - *Oggetto della perizia,* riporta al comma nr. 1 quanto segue: "la perizia è ammessa quando occorre svolgere indagini o acquisire dati o valutazioni che richiedono **specifiche competenze tecniche, scientifiche o artistiche**"; se ne deduce quindi che chiunque sia incaricato di trascrivere una intercettazione debba farlo osservando le forme della perizia e quindi possedere specifiche *competenze tecniche, scientifiche o artistiche*.

[3] Il grassetto presente negli articoli di legge o nelle sentenze di cassazione (in tutte le pagine successive) è sempre inserito dall'autore del contributo.

[4] Romito (2013: 181): "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. Normalmente si sta più vicini quando vi è maggiore confidenza, e ciò influenza diafasicamente il parlato. Una minore confidenza produrrà una maggiore relazione verbale di tipo istituzionale".

[5] Per *progressivo* si intende un numero assegnato ad una registrazione o intercettazione in senso cronologico.

una parte di esso) oppure *tutte le parole presenti* e quindi una trascrizione di tipo lessicale (questa volta anche di una piccola parte di un progressivo intero[6]).

L'interpretazione del termine *integrale* inteso come l'intero progressivo intercettato, ha da sempre sortito un'intensa discussione soprattutto in riferimento a materiale intercettato di conversazioni fra presenti, dunque di tipo ambientale[7]. Il numero di progressivo di una intercettazione ambientale, normalmente, ha una scansione temporale (ogni 60 minuti). Ciò vuol dire che in un progressivo potremmo avere più discussioni affrontate oppure che un unico argomento non venga esaurito. Sull'identificazione della porzione da trascrivere in relazione al concetto di *integrale*, la giurisprudenza non ha assunto una linea univoca, quindi mentre in alcuni processi importanti di 'ndrangheta il giudice ha richiesto al perito la trascrizione di 22 mesi di intercettazione ambientale continua[8], o oltre 600 ore di intercettazione ambientale (tra colloqui in carcere[9], appartamenti ed uffici)[10], altri giudici richiedono solo alcune porzioni di segnale ritenuto importante al fine della decisione[11]. La Cassazione d'altronde riporta che:

---

[6] Di prassi, la trascrizione effettuata dagli operatori di polizia giudiziaria non è integrale, perché riporta solo alcuni passaggi dell'intercettazione, presentando nei verbali molti *omissis* o *porzione di segnale non utile*.

[7] Cfr. il § 3 *Le intercettazioni*.

[8] Come ad esempio nel p.p. nr. 942/06 RGNR (Registro Generale Notizie di Reato) e 1025/06 RGGIP (Registro Generale Giudice per le Indagini Preliminari) del Tribunale di Catanzaro denominato *Anaconda*, dove il quesito posto dal giudice recita: "proceda il perito alla trascrizione delle conversazioni indicate […] p.p. 942/06 RGNR mod. 21 DDA (Direzione Distrettuale Antimafia), nr RIT (Registro Intercettazioni Telefoniche. L'acronimo nato quando le intercettazioni erano solo effettuate su linee telefoniche oggi viene utilizzato per qualunque tipo di intercettazione) 333/06 Intercettazione ambientale presso magazzino sito in via […], tutto il periodo compreso dalle ore 9.29 del 20.07.2006 alle ore 9.26 del 14.05.2008 […]".

[9] Sappiamo, ad esempio, che in un colloquio in carcere (la cui durata oscilla da una a quattro ore) la maggior parte della conversazione verte su argomenti privati che niente hanno a che vedere con l'obiettivo dell'intercettazione o dell'indagine.

[10] Come ad esempio nel p.p. nr. 655/17 RGT (Registro Generale Tribunale) del Tribunale di Reggio Calabria denominato *Gotha*.

[11] Come ad esempio nel p.p. 1674/2017 RGT del Tribunale di Catanzaro denominato *Bad Company*, dove il quesito posto dal giudice recita: "il Tribunale rappresenta la necessità […] che le parti individuino il solo minutaggio utile delle conversazioni in carcere oggetto di trascrizione al fine di contenere i tempi di durata dell'incarico peritale".

> l'incarico peritale che limiti la **trascrizione** ad una parte soltanto del contenuto delle intercettazioni telefoniche non è affetto da alcuna nullità, sia perché la nullità non è prevista né può farsi discendere dalla previsione di cui all'art. 268, comma 7, c.p.p., sia perché ciò che rileva ai fini del diritto di difesa è che, nell'espletamento della **trascrizione**, siano osservate le forme di garanzia previste per la **perizia**, atteso che in caso di **perizia** disposta in dibattimento la facoltà di nomina di propri consulenti, con le spedite modalità di cui all'art. 152 disp. att. c.p.p., consente all'imputato di svolgere osservazioni circa la rilevanza delle registrazioni non trascritte e di provvedere, attraverso il proprio consulente, a far trascrivere quanto altro possa interessargli, potendo comunque estrarre copia delle trascrizioni e far eseguire la trasposizione delle registrazioni su nastro magnetico. (Fattispecie antecedente all'entrata in vigore della riforma di cui al d.lg. 29 dicembre 2017 n. 216)[12].

Dalla lettura di questa sentenza, così come nei quesiti riportati negli incarichi precedenti (cfr. note 8-11), l'interpretazione di *integrale* risulta essere la meno corretta dal punto di vista linguistico poiché riduce la comunicazione esclusivamente alla trascrizione di *tutte* le parole presenti nella registrazione, non considerando i molti canali di cui è composta. La trascrizione delle sole parole a volte non è sufficiente per comprendere il senso di una conversazione[13]. Sarebbe quindi più corretto richie-

[12] Cassazione penale, sez. II, 30/01/2019, n. 15814.

[13] È infatti necessario, in alcuni casi, integrare la trascrizione con la segnalazione delle pause, dell'intonazione, del tono, di alcuni rumori di fondo ecc. si veda ad esempio il caso di uno spostamento semantico segnalato da una pausa es.: *Antonio: ciao è arrivata la (pausa) pizza?*, dove la pausa tra l'articolo e il nome è la chiave di lettura che indica all'ascoltatore lo spostamento semantico dal lessema *pizza* a *quello che sappiamo solo io e te*; oppure nell'esempio seguente dove il parlante utilizza il cambio dell'intonazione per modificare il significato di un aggettivo: *A: Come è il lavoro di Gianni? B: Buono*. In questo caso l'interlocutore B può produrre l'aggettivo *buono* con una intonazione discendente, alterando il normale modello intontivo, per indicare che il lavoro di Gianni non è buono ma sufficiente, oppure con una intonazione ascendente sottolineando la bontà dal lavoro di Gianni. Nel primo esempio la valutazione del lavoro di Gianni da parte di B è negativa mentre nel secondo esempio è positiva. Certo B avrebbe potuto sfruttare la ricchezza degli aggettivi in italiano utilizzando ad esempio: insufficiente o scarso, invece, come normalmente si fa nel parlato, conferisce una gradualità a *buono* attraverso un preciso modello intonativo. Questa gradualità viene utilizzata dai parlanti in molti casi e non solo con gli aggettivi ma anche con i nomi e con i verbi (cfr. Romito 2013: 258–262).

dere al perito trascrittore di concentrare la propria attenzione su pochi minuti di intercettazione utili allo scopo, non soffermandosi solo sulle parole, ma trascrivendo anche tutte le altre informazioni presenti nel flusso comunicativo[14].

La sentenza seguente da una parte ritiene la trascrizione un mero fatto tecnico, quindi espletabile da chiunque (al contrario di quanto affermato in precedenza), dall'altra invece riconosce nella bobina (difatti nel suo contenuto originale, cioè il segnale sonoro) l'unico elemento di prova. Leggendo le sentenze seguenti si potrebbe pensare che il segnale intercettato sia di ottima qualità e che sia sufficiente poter ascoltare (senza nessuna particolare competenza o strumentazione particolare) per percepire l'informazione necessaria:

> 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

[14] Riteniamo che la ragione per cui la maggior parte delle trascrizioni forensi consegnate siano lessicali e superficiali sia da imputare a fattori economici innanzitutto e temporali in secondo luogo. Una trascrizione integrale a scopo di ricerca effettuata da un esperto linguista (effettuata su registrazioni di ottima qualità al contrario di quanto avviene nelle intercettazioni a scopo forense), come ad esempio nel progetto CLIPS (CLIPS 2020), presuppone che per 1 minuto di registrazione vengano stimate 10 ore di lavoro, con una retribuzione per il trascrittore che oscilla da 800€ a 1000€. Il compenso di un trascrittore forense in base all'art. 4 della legge n. 319/1980 è di 4,7€ per ogni ora di lavoro e normalmente il giudice riconosce congruo che un'ora di intercettazione si trascriva in 4/8 ore al massimo che rapportato al minuto della trascrizione a scopo di ricerca, equivale ad una retribuzione per il trascrittore forense che oscilla da 0.31€ a 0.62€. Certo riconosciamo che il livello di approfondimento dei due esempi citati sia molto differente, ma parimenti riteniamo esageratamente sproporzionata la relazione tra i due compensi. La seconda ragione risiede nel tempo assegnato per effettuare una trascrizione forense. Spesso (tranne nella fasi di indagini quindi con incarichi ricevuti dalla Procura), il Giudice assegna la trascrizione durante le fasi del processo: in udienza; il Pubblico Ministero non procede ad una cernita del materiale da trascrivere ma preferisce richiedere la trascrizione di tutto il materiale intercettato; la consegna della trascrizione preferibilmente deve avvenire prima della scadenza dei termini di custodia cautelare e comunque prima dell'inizio del dibattimento, pertanto il perito dovrà trascrivere tantissime ore di intercettazione di scarsa qualità in pochissimo tempo.

> 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. 111, comma 4, della Costituzione) quella secondo cui le prove devono formarsi nel contraddittorio[15].

Ne deduciamo quindi che il giudice nella sua figura di *peritus peritorum* (cfr. 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.

Le sentenze seguenti, invece, in qualche modo rinnegano il concetto di *integrale* riportato in precedenza. Si ritiene infatti che non sia necessario un verbale scritto (una trascrizione cartacea) ma sia sufficiente una deposizione testimoniale (quindi una interpretazione sommaria del contenuto di una intercettazione):

> il contenuto delle conversazioni intercettate può essere provato anche mediante deposizione testimoniale, **non essendo necessaria la trascrizione delle registrazioni nelle forme della perizia**, atteso che l'art. 271 comma 1 c.p.p, non richiama la previsione dell'art. 268 comma 7 c.p.p, tra le disposizioni la cui inosservanza determina l'inutilizzabilità e che la mancata trascrizione non è espressamente prevista né come causa di nullità, né è riconducibile alle ipotesi di nullità di ordine generale tipizzate dall'art. 178 c.p.p.[16]

In tema di intercettazioni telefoniche, il contenuto delle conversazioni intercettate può essere provato anche mediante deposizione testimoniale, non essendo necessaria la trascrizione delle registrazioni nelle forme della perizia, atteso che la prova è costituita dalla bobina o dalla cassetta, che l'art. 271, comma 1, c.p.p. non richiama la previsione dell'art. 268, comma 7, c.p.p. tra le disposizioni la cui inosservanza determina

[15] Cassazione penale, sez. VI, 28/03/2018, n. 24744.

[16] Cassazione penale, sez. III, 20/02/2018, n. 16040.

l'inutilizzabilità e che la mancata trascrizione non è espressamente prevista né come causa di nullità, né è riconducibile alle ipotesi di nullità di ordine generale tipizzate dall'art. 178 c.p.p.[17]

In queste ultime sentenze l'interpretazione del trascrittore ascoltato mediante deposizione testimoniale è ovviamente molto attiva.

Altro concetto riguarda l'interpretazione di *in forma intelligibile* presente nell'art. 268 comma 7 del Codice di Procedura Penale - Esecuzione delle operazioni citato all'inizio di questo lavoro. Consultando i dizionari della lingua italiana si rileva che intelligìbile (o intellegìbile) è un concetto, un contenuto accessibile e comprensibile. Quindi definire la trascrizione "integrale […] ovvero la stampa in forma intelligibile delle informazioni contenute nei flussi di comunicazioni informatiche"[18] risulta essere quanto meno ambiguo. Non può sicuramente far riferimento al contenuto, in quanto il trascrittore riporta quanto detto da un parlante intercettato e non può assolutamente interpretare per rendere maggiormente *intelligibile* il contenuto di una comunicazione; quindi deve essere interpretato come *chiaro* dal punto di vista grafico, cioè senza l'utilizzo di simboli o caratteri particolari (come quelli dell'Alfabeto Fonetico Internazionale ad esempio).

A definire ancora meglio la natura della trascrizione forense sono però le seguenti sentenze:

> La trascrizione deve consistere, […] nella mera riproduzione in segni grafici corrispondenti alle parole registrate[19].
> **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

[17] Cassazione penale, sez. VI, 20/02/2014, n. 25806.
[18] Art. 268 comma 7 del Codice di Procedura Penale - Esecuzione delle operazioni.
[19] Cassazione penale, sez. I, 24/04/1982, n. 805.

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[20].

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, c.p.p. 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**)[21].

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 contenuto**, procedendo direttamente al loro ascolto o disponendo una nuova perizia (Fattispecie in cui la perizia era stata disposta dopo il rinvio a giudizio dell'imputato da giudice, quindi, incompetente)[22].

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

[20] Cassazione penale, sez. VI, 30/10/1992, in Mass. Pen. Cass. 1993, fasc. 6,12 (s.m.).

[21] Cassazione penale, sez. VI, 06/11/2008, n. 2732.

[22] Cassazione penale, sez. VI, 15/03/2016, n. 13213.

> 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[23].
>
> 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[24].

Potremmo citare tante altre sentenze più recenti che richiamano quelle qui riportate, ma quello che è possibile dedurre da questa analisi (non tenendo conto dell'art. 268 comma 7 del c.p.p.) è che secondo la prassi giudiziaria la trascrizione è una operazione esclusivamente materiale, tecnica, non necessaria. Per *integrale* bisogna intendere *tutte, e solo, le parole presenti nella registrazione,* o un progressivo *per intero* e non una sua parte, e per *intelligibile* l'utilizzo di una scrittura chiara e facilmente comprensibile, quindi senza l'utilizzo di caratteri speciali o particolari[25]. La giurisprudenza ritiene che tutto il carico informativo di

[23] Cassazione penale, sez. VI, 3/11/2015, n. 44415.

[24] Cassazione penale, sez. VI, 22/1/2016, n. 3027.

[25] Ci è noto, ovviamente, il procedimento del *libero convincimento del giudice,* cioè la possibilità nella veste di *peritus peritorum* di accedere direttamente alle fonti di prova e quindi ad esempio di voler ascoltare direttamente una intercettazione. In nessuna sentenza da noi consultata però, emerge che siano stati valutati andamenti intonativi, pause o oltre informazioni presenti nel segnale acustico e assenti nella trascrizione. Al contrario ci risulta invece che il giudice può disconoscere la trascrizione effettuata dal perito trascrittore (esperto): "ciò premesso, va rammentato che '*la trascrizione, anche quella peritale, non costituisce la prova diretta di una conversazione, ma va considerata solo come un'operazione rappresentativa in forma grafica del contenuto di prove acquisite mediante la registrazione*

una discussione sia presente e veicolata solo attraverso le parole[26]. Il perito, o meglio il trascrittore forense, non sarebbe quindi un esperto, bensì un tecnico capace di riportare su carta le parole ascoltate e, in alcuni casi, con facoltà di scegliere le porzioni più attinenti[27].

fonica' (C.Cass., n. 4892/03). Nella sentenza della Sezione II, n. 12991/2013, i Giudici di Legittimità, nel respingere le doglianze dei ricorrenti relativamente all'ascolto diretto (in quel caso da parte dei Giudici d'Appello) dei *files* audio relativi agli originali delle intercettazioni ambientali, pur dopo apposite perizie foniche, ribadivano la possibilità del Giudice di valutare una prova tecnica *in modo difforme da quello suggerito dal perito*. Il Giudice, quale *peritus peritorum*, ben può '*esprimere il proprio giudizio in motivato contrario avviso rispetto a quello dei periti*'. L'importante è che il giudizio venga raggiunto non '*in base ad apodittiche od arbitrarie intuizioni, ma grazie all'ascolto diretto delle intercettazioni per il quale nessuna norma di legge prescrive o vieta l'uso di normali cuffie o di altro particolare accorgimento tecnico*'. Nella sentenza della Sezione I, n. 22062/2013 i Giudici di Legittimità hanno evidenziato che '*è sempre consentito al giudice l'ascolto in camera di consiglio dei supporti analogici o digitali recanti le registrazioni, debitamente acquisite e trascritte e l'utilizzo ai fini della decisione dei risultati dell'ascolto medesimo. (nella specie, la Corte ha ritenuto pienamente utilizzabili i risultati dell'ascolto dei supporti digitali, contenenti le copie delle registrazioni, allegate a corredo della perizia trascrittiva)*'" Sentenza p.p. 110/13 RGNR e 3914/14 RG GIP, Tribunale di Asti.

[26] Un ulteriore problema che non affronteremo in questa sede, riguarda le intercettazioni di un parlato dialettale. In questi casi il codice di procedura penale non prevede interventi particolari, non essendo il dialetto considerato una lingua ufficiale. Le sentenze invece riportano, ancora una volta, dichiarazioni contrastanti. La sentenza della Corte di Cassazione del 1982 precedentemente citata riporta che "nell'ipotesi, invece, di comunicazioni effettuate in lingua straniera o in un dialetto scarsamente intelligibile, la traduzione consiste **in due distinte operazioni**: la prima relativa alla riproduzione integrale degli elementi fonetici raccolti nella registrazione; la seconda nella vera e propria traduzione in lingua italiana *ex* art. 326 c.p.p." (Cass. pen., 1 sez., 24.4.82, n. 805); invece la Corte di Appello di Bologna qualche anno dopo in una propria sentenza riporta che "nell'ipotesi di […] conversazione, […] gli interlocutori utilizzino, […] talune espressioni ed inflessioni dialettali, non è necessaria la previa translitterazione dalle espressioni dialettali alla lingua italiana seguita, poi, dalla successiva traduzione, essendo sufficiente una traduzione unica, in lingua italiana, dell'intero contenuto della conversazione intercettata" (App. BO 9.11.90 II 502) assegnando al trascrittore una grande libertà di interpretazione nonché di traduzione (cfr. Cronin *et al.* 2013).

[27] Cassazione penale, sez. VI, 04/06/1993: "in materia di intercettazioni telefoniche, l'omissione da parte del perito della trascrizione di quelle conversazioni non attinenti, a giudizio del medesimo, ai fatti oggetto del processo, costituisce una mera irregolarità non sanzionata da alcuna espressa comminatoria di nullità né della parte non tradotta né della trascrizione parziale. Ciò che rileva ai fini del diritto della difesa è che, nell'espletamento della trascrizione siano osservati modi, forme e garanzie previsti per la perizia. L'imputato, inoltre, ha la facoltà di nominare un consulente tecnico (art. 225 c.p.p.), il quale può svolgere osservazioni circa l'omessa o incompleta trascrizione di parti di conversazioni ritenute ri-

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"[28]. Il giudice potrebbe ascoltare tutte le intercettazioni per giungere al proprio convincimento, così come tutte le parti e poi, in dibattimento, discutere e *formare* la prova in base alle proprie opinioni nate dall'ascolto globale del segnale intercettato.

Personalmente ritengo che questa sia una possibilità da tenere in alta considerazione. Tutte le parti presenti in un'aula di tribunale ascoltano (in maniera autonoma) il segnale sonoro nella sua interezza e ne deducono informazioni a favore o contro la propria linea difensiva/accusatoria. Ciò sarebbe in qualche modo in linea con quanto i linguisti affermano e con l'idea che il flusso comunicativo non sia fatto solo di parole. È notorio che durante una conversazione, il parlante trasferisce in prima istanza l'aspetto emotivo che deve essere considerato come la chiave di lettura per una corretta interpretazione delle parole e del messaggio che sta producendo. Ad esempio nello scambio:

> Lei: scusa, scusa, ho fatto tardi;
> Lui: Ti ammazzerei!

l'espressione *ti ammazzerei*, in base all'intonazione e all'aspetto emotivo comunicato, deve essere interpretata come un richiamo affettuoso e non certo come una minaccia[29].

levanti per la difesa e, ove non sia stato nominato un consulente tecnico, il difensore può estrarre copia delle trascrizioni e fare eseguire la trasposizione delle registrazioni su nastro magnetico (art. 268 comma 8 c.p.p.) onde accertare specifiche incompletezze o omissioni pregiudizievoli per la difesa che ben possono essere indicate anche in sede di dibattimento".

[28] Cassazione penale, sez. I, 26/03/2009, n. 26700.

[29] Nel processo denominato *no global* pp. n. 3997/01 RGNR e 3618/02 RG GIP, il giudice sulla base di una trascrizione effettuata dagli operatori di PG (Polizia Giudiziaria) e di una ordinanza di custodia cautelare del pubblico ministero, iscrive al registro degli indagati alcune persone. L'ordinanza di custodia cautelare scritta dal pubblico ministero si fonda sulla lettura di una trascrizione e sulla mancata conoscenza del contesto e del

Quanto detto in questo paragrafo porta a concludere che la considerazione del processo di trascrizione da parte della giurisprudenza, la poca competenza dei trascrittori forensi (per i motivi ampiamente trattati), la totale assenza di norme, procedure e opinioni comuni e la bassa qualità degli strumenti tecnici utilizzati crea una situazione quanto meno discutibile.

3. *Le intercettazioni*

Esistono tanti tipi di intercettazione dalle caratteristiche molto differenti. Queste possono sommariamente essere divise in due macro categorie: intercettazione di telecomunicazioni, comunemente definita intercettazione telefonica o *clear recording*, e intercettazione di conversazioni fra presenti, nota come intercettazione ambientale o *poor recording* (cfr.

dialetto utilizzato. Nell'ordinanza si legge: "questa telefonata è molto importante perché dimostra che l'argomento relativo alla manifestazione di Genova ed a quello che si sarebbe dovuto fare in quella sede è stato 'gestito' non solo da [omissis] ma anche da [omissis] i quali parlando di '*sparata su Genova e chiudiamo il discorso*' fanno capire che il loro intendimento su quello che si sarebbe dovuto fare a Genova era già molto chiaro".

La trascrizione quasi integrale della telefonata riportata nell'ordinanza di custodia cautelare, cioè la nr. 481 del 16/5/2001 alle ore 20,49 direttamente tradotta in lingua italiana, è la seguente: "U1= e noi la domenica invece approfondiamo aspetti più meridionali, vertenze …; U2= si, si perfetto. Primo punto all'ordine del giorno, Genova di modo che …; U1= perfetto, *per cui ci facciamo una sparata su Genova e chiudiamo il discorso*; U2= esatto, esatto, diciamo che il sabato è il giorno che insomma…; U1= è massima la presenza dei compagni, quindi diamo la priorità … ; U2= è chiaro, a me mi sembra scontato, considerato gli appuntamenti a seguire che ci sono..; U1=prima che mi scordo, siccome vengono tante persone e noi logisticamente … insomma stiamo vedendo di sistemare tutte le cose; U2= si; U1= però se voi portate i sacchi a pelo…; U2 = si mo avviso i compagni; U1= è un aiuto; U2= si certo mo avviso i compagni; U1= poi non è detto che debbano servire, però… insomma chi ce l'ha lo portasse; U2= perfetto d'accordo, se so qualcosa su Bari io stesso te lo faccio sapere, se no ti faccio telefonare insomma, se non ti telefoniamo non vengono […]". Secondo il pubblico ministero e il giudice, in questa telefonata si organizza quanto avvenuto nelle tristi giornate genovesi durante il G8. La conversazione avviene in dialetto, e la frase è stata prodotta durante l'organizzazione di un dibattito sindacale. "Fatti sta sparata" in dialetto calabrese vuol dire fai ciò che devi fare o dì ciò che devi dire in fretta. Quindi la corretta traduzione italiana della frase incriminata "per cui ci facciamo una sparata su Genova e chiudiamo il discorso" sarebbe stata "per cui parliamo velocemente di Genova e poi chiudiamo il discorso".

Fraser 2003; Galatà 2013). Nella prima macro categoria il materiale registrato, sia su rete fissa che mobile[30], ha una durata limitata e generalmente è privo di rumori *additivi*[31]. Anche quando ci troviamo in presenza di rumori ambientali, la distanza ravvicinata tra la fonte sonora e quella di registrazione (capsula microfonica del telefono), garantisce una buona qualità della registrazione. La conversazione fra parlatori avviene *in assenza*, e quindi i partecipanti sono chiamati a cooperare attivamente alla buona riuscita dello scambio comunicativo, sopperendo all'assenza di un *controllo visivo* e di un *canale complementare* (cfr. Bazzanella 2008; Goffman 1987). In questo tipo di intercettazione, all'interno di un progressivo, troviamo normalmente una conversazione completa, con un inizio ed una fine spesso dichiarati da un "pronto" e un "ciao". La seconda macro categoria è costituita dalla registrazione in ambienti chiusi o aperti: mezzi di trasporto, abitazioni, uffici, negozi o istituti penitenziari[32]. Nella maggior parte dei casi la registrazione è fortemente disturbata, dal momento che in essa confluiscono rumori di sottofondo di qualsiasi tipo; inoltre il numero degli interlocutori coinvolti può essere potenzialmente infinito, gli scambi conversazionali non rispettano un avvicendamento dei turni regolare, lo spazio fisico che intercorre fra il dispositivo di intercettazione e i soggetti indagati è altamente variabile e può anche compromettere il segnale in maniera significativa. Nel caso specifico di

[30] Attualmente, l'intercettazione su rete fissa avviene tramite smistamento del traffico telefonico (da e verso il numero intercettato) adoperato in tempo reale da centri di commutazione numerica; le comunicazioni su rete mobile (associate alla/e SIM e al codice identificativo dell'apparecchio, l'IMEI) sono direttamente inviate dai centri di interconnessione (*Mobile Switching Centre*, MSC) dell'operatore telefonico a un server della Procura della Repubblica, dove il segnale viene intercettato e registrato (cfr. Galatà 2013).

[31] I rumori additivi sono quei rumori, quali traffico, passi, brusio, rumori elettrici, voci di sottofondo, che interagiscono con le frequenze del segnale sonoro modificandolo e riducendone l'intelligibilità (cfr. Romito 2013: 274).

[32] Questo tipo di intercettazione avviene per rete mobile, ovvero tramite apparecchi associati a un numero SIM e dotati di microfono e antenna (le microspie) che trasmettono il segnale sotto forma di telefonata (cfr. Galatà 2013). Oggi le intercettazioni più moderne avvengono tramite software spia installati su smartphone o tablet (ad esempio i *trojan* - dal nome del celebre inganno di Ulisse -). Questi software vengono inviati attraverso sms e installati su un qualsiasi dispositivo elettronico ricevente e sono in grado di impadronirsi di tutti i comandi dell'apparecchio. Il software è definito "captatore" se legale e frutto di autorizzazione da parte di una procura, "malware" (o *malicious software*) se inviato illegalmente.

intercettazioni nella sala colloqui di un carcere, possono subentrare disturbi volontariamente prodotti dai parlatori, consapevoli di essere intercettati e decisi a pregiudicare l'intelligibilità del segnale ottenuto (ultimamente si tenta di sopperire a questa limitazione effettuando delle intercettazioni audiovisive). L'intercettazione ambientale ripropone una conversazione *faccia a faccia* o *in presenza* che, a differenza della comunicazione verbale telefonica, tende a omettere molte delle informazioni esplicite e ordinate tipiche di una telefonata. La conversazione in presenza contenuta nelle intercettazioni è ricca di perdite (cfr. Sinatra 2014). Mancano infatti tutte le informazioni cinesiche e prossemiche legate a espressioni facciali, sguardi e gestualità, spesso sostitutive di qualsiasi manifestazione fonetico/linguistica, legate al cosiddetto *canale complementare* (cfr. Goffman 1987).

4. *La trascrizione in ambito linguistico*

In generale si identificano con il termine *trascrizione* operazioni molto differenti tra loro, come ad esempio gli appunti di uno studente in aula durante una lezione di filosofia o di un giornalista durante una conferenza stampa o di uno studioso durante un convegno, ma anche l'elenco della spesa dettato per telefono o il verbale di una riunione, di un interrogatorio; la trascrizione di un film per la sottotitolazione, la trascrizione di un logopedista durante una seduta con un paziente affetto da una qualche patologia meccanica (non fluente) del linguaggio[33], la sceneggiatura di un attore; un'indagine dialettologica, sociologica ecc. Ovviamente queste forme di trascrizione sono molto differenti tra loro e non hanno nulla a che vedere con la trascrizione linguistica: potrebbero, infatti, essere classificate come sintesi scritta di testi orali in quanto riporteranno su carta solo i dati utili allo scopo del trascrittore. Così la lista della spesa riporterà solo gli oggetti da comprare e non i commenti sul prezzo o sulla qualità, la lezione trascritta dagli studenti riporterà ciò che lo studente ritiene essere importante per lo studio della disciplina e

[33] In questo caso la trascrizione attraverso la semplice percezione di una forma sonora indica al foniatra o al logopedista la strada da intraprendere in caso di riabilitazione o di intervento. Si veda Vernero & Romano (2017).

per il superamento dell'esame[34], ecc. Nei diversi dizionari di lingua o di linguistica italiana[35], la trascrizione viene confusa addirittura con la traslitterazione, che significa tradurre i caratteri grafici di una lingua nei caratteri di una lingua diversa, ad esempio i caratteri cirillici o greci antichi nell'alfabeto latino.

La trascrizione linguistica invece è un'analisi scientifica, utilizzata solo da specialisti, che mira a riportare su carta tutte le informazioni presenti nel canale sonoro. Questa si differenzia in base al settore di studio[36] e di applicazione. Ad esempio la trascrizione fonetica è una operazione molto complessa[37] con diversi livelli di precisione e di profondità (trascrizione stretta o larga), che non si sofferma sui significati né parziali né globali dei segnali sonori, ma esclusivamente sui singoli suoni prodotti; la trascrizione prosodica invece sofferma la propria attenzione sugli aspetti intonativi del parlato.

Sarebbe quindi necessario differenziare la trascrizione intesa come la mera trasposizione delle parole ascoltate e interpretate su carta, dalla trascrizione linguistica o fonetica[38]. La trascrizione linguistica o fonetica pone

[34] Si veda Romito (2013: 69).

[35] Treccani Vocabolario online: "rappresentazione dei fonemi di una lingua o di un dialetto in un sistema grafico diverso o comunque non usuale per quella lingua o per quel dialetto". Treccani Enciclopedia online: "1. L'azione e l'operazione di trascrivere, […] t. di un testo, […]; a. Rappresentazione grafica dei fonemi di un contesto, di una lingua o di un dialetto, in un sistema di scrittura diverso". Dubois *et al*. (1979: 302): "trascrivere significa far corrispondere, termine per termine le unità discrete della lingua parlata con le unità grafiche". Cardona (1989: 307): "ogni procedimento che registri un enunciato verbale per mezzo di un sistema di segni grafici (distinto dalla traslitterazione che è dallo scritto allo scritto)".

[36] Scegliendo di fatto di privilegiare un canale della comunicazione omettendo o trascurando volontariamente gli altri.

[37] In Minissi (1990: 3) si legge "le scritture fonetiche offrono questi simboli arbitrari ma costanti. Tutte le norme che esse danno circa l'utilizzazione dei simboli stessi riguardano la coerenza interna del sistema di simboli, non la problematica linguistica; allo stesso modo come le convenzioni della simbologia algebrica (per esempio […] la differente funzione dei differenti tipi di parentesi, il diverso significato delle cifre a seconda che siano in linea o come esponenti ecc.) costituiscono l'ordine grafico e non l'ordine concettuale dell'algebra".

[38] Minissi (1990: 103) "un alfabeto fonetico offre solo i simboli e non le soluzioni ai problemi della trascrizione. La trascrizione è essenzialmente un'operazione linguistica non una operazione grafica e consiste nel connotare l'interpretazione fonetica o la sistemazione fonematica dei fatti di lingua considerati mediante l'uso di simboli arbitrari di valore costante".

la propria attenzione sulla *forma* del messaggio. Ad esempio nell'Analisi Conversazionale che mira ad individuare le azioni sociali messe in atto dagli interlocutori, l'attenzione non si sofferma sulle parole dette ma sulla loro modalità di produzione (sospiri, risate, pause, sovrapposizioni ecc.)[39]. Nelle trascrizioni forensi (che vedremo nel § successivo) invece, l'attenzione è rivolta alla comprensione del contenuto, e quindi vengono ritenute poco importanti la presenza di pause, cambi di intonazione o altro.

5. *Le trascrizioni e le verbalizzazioni in ambito forense*

Oltre a quanto già accennato sulla differenza tra le diverse trascrizioni[40], è necessario introdurre il concetto di *verbalizzazione*, cioè la trascrizione di una registrazione nella quale tutti i parlanti partecipano e collaborano al buon raggiungimento di uno scopo, producendo un parlato altamente intelligibile che dovrà diventare verbale o documento scritto di un fascicolo e *trascrizioni forensi* dove, al contrario, tutti o una parte dei parlanti, volontariamente o perché inconsapevoli di essere registrati, sono poco intelligibili e poco collaborativi.

Nel caso delle *verbalizzazioni* come quelle di udienze, di una riunione di condominio o di un consiglio di amministrazione, la correttezza della trascrizione viene affidata al singolo parlante intervenuto attraverso l'approvazione del verbale di trascrizione nella seduta successiva.

Al contrario, nelle *trascrizioni forensi*, chi parla ha lo scopo esattamente contrario a quello dell'avvocato di udienza o del membro di un consiglio di amministrazione. Il suo scopo è infatti quello di farsi capire solo da chi ha di fronte (si vedano le intercettazioni effettuate durante i colloqui in carcere); il parlante per l'effetto chiamato *Romito*[41] nel

[39] Si veda a questo proposito la notazione introdotta da Jefferson (2004) e per l'applicazione dell'analisi conversazionale in ambito forense si veda Romito *et al*. (2016).

[40] Per un approfondimento sui tipi di trascrizione si veda Romito (2013: 67-79).

[41] Romito (2013: 264) L'effetto contrario è definito *effetto Lombard* o (*Lombard reflex*), scoperto da Etienne Lombard (otorinolaringoiatra francese) nel 1909. È la tendenza dei parlanti ad aumentare l'intensità della loro voce in presenza di un rumore di fondo in modo da sovrastare il rumore con la voce. Lo sforzo e il cambiamento non riguarda solo l'intensità ma anche altre caratteristiche come il *pitch* o la durata sillabica. Questa compensazione ha come risultato l'aumento nell'ascoltatore del rapporto segnale rumore e quindi nella maggiore comprensione delle singole parole. Tale effetto si attua anche inconsciamente ad esempio quando parliamo mentre ascoltiamo musica con le cuffie.

dire una qualunque cosa (anche senza alcun pericolo di intercettazioni) che possa andare contro la morale, contro la legge o possa essere ritenuto lesivo o anche solo sconveniente, è involontariamente spinto ad abbassare il proprio tono di voce al di sotto della soglia minima, cioè senza far vibrare le corde vocali e, producendo un parlato mormorato o bisbigliato, avvicina le labbra alle orecchie dell'ascoltatore o evidenzia il movimento delle labbra affinché l'ascoltatore possa supplire alla mancanza di informazioni acustiche con l'informazione dei movimenti labiali durante il processo di decodifica del segnale audio. In molti casi, affinché questo avvenga più facilmente, il parlante mette le mani intorno alla bocca per attirare l'attenzione dell'ascoltatore verso le labbra. Solitamente, durante una lezione in aula o un'arringa in udienza, il parlante posiziona la propria bocca davanti al microfono (si presenta prima di parlare) e collabora alla massima comprensione e, durante i passaggi importanti, alza il volume della voce e scandisce il concetto sillabandolo. Nel caso invece di una intercettazione, il parlante tenterà di nascondere un contenuto importante sussurrando le parole, tentando di dire il meno possibile affidandosi alle conoscenze pregresse e condivise e supportando le omissioni con la mimica e il codice gestuale (assente in una registrazione sonora).

Quindi, in base alla situazione comunicativa, il parlante programma in modo *acoustic oriented*, dove tutto il carico informativo dell'eloquio è affidato alla qualità acustica del segnale, oppure in modo *system oriented*, dove molta informazione dell'eloquio è affidata alle conoscenze pregresse e condivise, assegnando all'ascoltatore il compito di *ricostruire* parte del segnale/informazione non esplicitato/a (cfr. Lindblom 1990). Normalmente la *verbalizzazione* riguarda eloqui tendenzialmente *acoustic oriented*, mentre invece la *trascrizione forense* riguarda eloqui tendenzialmente *system oriented.* Le metodologie e le procedure per i due tipi *verbalizzazione* e *trascrizione forense* sono e devono essere molto differenti tra loro[42].

Un'ulteriore variabile utile per differenziare la *verbalizzazione* e la *trascrizione forense* riguarda la qualità della registrazione del segnale

[42] Questa sottolineatura è necessaria perché nella prassi giuridica spesso il trascrittore di udienza viene anche nominato dal giudice trascrittore forense. Nella maggior parte dei casi l'approccio metodologico è lo stesso e le relazioni consegnate sono strutturalmente simili.

sonoro[43] che dovrà essere trascritto. Il linguista ricercatore[44] effettua le registrazioni in un ambiente silenzioso e, nel caso non sia soddisfatto della registrazione, può chiedere al proprio informante di ripetere una parola o una frase. Nelle intercettazioni forensi la microspia registra tutto ciò che viene prodotto, i parlanti inconsapevoli si muovono nell'ambiente rendendo il segnale registrato a volte saturo e a volte impercettibile, sovrappongono le proprie voci ed entrano ed escono dalla conversazione senza alcun preavviso[45]. Inoltre la registrazione a scopo di ricerca viene eseguita da esperti, mentre nel caso delle intercettazioni a scopo forense le registrazioni vengono effettuate da società private vincitrici di una gara di appalto, che spesso si avvalgono di personale poco tecnico.

Riassumendo, risultano essere variabili importanti per la corretta differenziazione di una *trascrizione forense* rispetto ad una *verbalizzazione*: la qualità della registrazione e degli strumenti utilizzati, la competenza di chi effettua la registrazione, la collaborazione dell'intervistato. Pertanto il trascrittore non può e non deve avere lo stesso atteggiamento durante il processo di *verbalizzazione* e il processo di *trascrizione forense*.

---

[43] Non approfondiamo in questa sede la qualità degli strumenti di registrazione; è sufficiente sapere che quando la registrazione era analogica e gestita direttamente dallo Stato attraverso le Procure, venivano utilizzate bobine a 4 piste con una velocita di registrazione pari a 2,38 cm al secondo (contro i 22 cm al secondo utilizzati durante una registrazione per indagine linguistica o dialettologica). Una bassa velocità di registrazione comporta una scarsa qualità di registrazione, una notevole perdita di segnale ma permette di registrare oltre 5 ore di parlato per ogni bobina. Quando la registrazione diventò digitale, lo Stato delegò ai privati l'intercettazione. Ogni Procura ha quindi nella propria sala di intercettazioni diverse ditte esterne che si occupano di intercettare con propri strumenti, con propri formati proprietari, con propri software di consultazione. Anche in questo caso, così come quando la registrazione era analogica, l'obiettivo non è la qualità della registrazione ma l'economia e quindi la possibilità di contenere in una memoria (server) più registrazioni possibili. Questo è possibile campionando ad una bassa frequenza (8000 Hz rispetto ai 44000 normalmente utilizzati per una normale registrazione casalinga) e comprimendo il segnale (con elevatissime perdite di informazioni). Non approfondiamo neppure il settore delle intercettazioni video a scopo investigativo. Ci limitiamo a riportare che nel fascicolo viene inserita solo la trascrizione della traccia audio.

[44] Quanto detto vale anche per il tecnico addetto alle registrazioni di udienza. Spesso infatti l'operatore invita l'avvocato a parlare al microfono, a dire il proprio nome e cognome oppure ad alzare il volume della voce.

[45] Per maggiori informazioni sulle caratteristiche delle intercettazioni ambientali si veda Romito (2013).

6. *La professionalità del trascrittore*

Come già detto nel paragrafo precedente, le trascrizioni linguistiche vengono effettuate da linguisti o da persone con un percorso formativo linguistico, invece per il trascrittore forense il giudice deve riferirsi all'art. 221 del c.p.p. *Nomina del perito,* che recita: "il giudice nomina il perito scegliendolo tra gli iscritti negli *appositi albi* o tra persone fornite di particolare competenza nella *specifica disciplina*".

L'albo dei trascrittori purtroppo non esiste[46], così come non esiste istituzionalmente la Linguistica forense in Italia, quindi di norma il trascrittore forense è una persona che non ha una adeguata formazione[47] linguistica e il suo percorso formativo è tra i più variegati, oscillando dal titolo di scuola media inferiore fino al dottorato di ricerca in fisica[48]. Inoltre, mentre la *verbalizzazione* di una registrazione in udienza, come già detto, viene validata dagli stessi interlocutori nell'udienza successiva, la *trascrizione forense* riporta una realtà volutamente mantenuta nascosta dagli interlocutori registrati. La *trascrizione forense* è un indizio o una prova[49],

[46] Per approfondimenti riguardo il ruolo del perito forense si veda Romito (2010; 2013).

[47] Anche in questo caso la prassi è quella di confondere i titoli scientifici o di formazione con la competenza sul campo: Tribunale di Napoli, verbale di trascrizione di udienza, Esame del perito. Giudice – […] sommariamente vuole indicare alla Corte quali sono le sue competenze, […] quali sono i suoi titoli e le sue qualifiche?; Perito – io opero nel settore fonico da oltre 40 anni. Tribunale di Perugia, verbale di trascrizione di udienza, Esame del perito. Giudice – di professione esperto in? Perito – […] prima ero dirigente di azienda informatica, dal '91 […] mi sono messo a fare queste attività peritali. Il controllo sui titoli attestanti la formazione del perito potrebbe comunque essere svolto in aula dalle parti (Pubblico Ministero 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 curriculo 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".

[48] Cfr. Romito & Galatà (2008).

[49] Alcune sentenze di Cassazione ritengono che la trascrizione sia una prova, in altre invece solo un indizio, in altre ancora come in quella riportata nella nota 19 solo un documento tecnico senza alcun valore.

su una delle tante possibili ricostruzioni della realtà, sulla quale le parti fondano la propria accusa o difesa. Il trascrittore non si limita semplicemente a trascrivere il materiale sonoro (come riportano le sentenze della Corte di Cassazione) ma a volte *interpreta* un segnale acustico, ricostruendo una realtà o indirizzando le indagini[50]. Questo avviene quando la qualità della registrazione è bassa e il segnale non è intelligibile. Lo stesso vale quando il perito trascrittore interpreta alcuni rumori presenti nella registrazione come nel caso della seguente trascrizione effettuata dagli operatori di PG durante una indagine:

> S: *Ora quando lo vedo a Martino lo scherzo* (fonetico: cugghjiuniu [ndr. la corretta traduzione sarebbe *lo prendo in giro, lo sbeffeggio*); A: *Gli dici che … me la prendo io la macchina che spara…o tu? … gli devi dire*; S: *Si*; A: *quello ha capito … inc…*; S: *Si … adesso la prendo io la macchina che spara!;* A: *Se lo vedi … se viene*; S: *inc… nel paese … questa qua!* (alle ore 17:29:39 si sente maneggiare qualcosa che provoca un rumore metallico e che, a parere di questo ufficio, corrisponde al rumore prodotto dall'arretramento dell'otturatore di una pistola semi automatica per inserire il colpo in canna)*… no, non è male come macchina. Per questo me la sono presa …che è tenuta bene. Che vuoi fare. Qualche rigatina, qualche cazzata*[51].

---

[50] La Polizia Giudiziaria del Tribunale di Asti durante una intercettazione di una automobile proprietà di alcuni sospetti trascrive questa conversazione: "U1= *Su levaru* U2= *Ah?*; U1= *<<parole incomprensibili>>…già ne ho ammazzato uno…*; U2= *U sacciu, c'era <<parole incomprensibili>> non lo ammazziamo <<parole incomprensibili >>*; U1 = *No, lui e <<parole incomprensibili >>[ridono]*". Nonostante le tante parole incomprensibili e quindi probabilmente la scarsa qualità della registrazione il giudice per le indagini preliminari decide per il fermo cautelare in attesa del processo procedendo all'arresto dei due ragazzi (trad.it dell'autore U1= *lo hanno portato via* U2= *Ah?*; U1= *<<p. inc.>>…già ne ho ammazzato uno…*; U2= l' so, *c'era <<p. inc.>> non lo ammazziamo << p. inc.>>*; U1= *No, lui e <<p. inc.>>) [ridono]*. Il giudice per le indagini preliminari nomina un perito trascrittore il quale consegna la seguente trascrizione: U1= *l'ulivaru*; U2= *Ah?*; U1= *Pizzicai n'animali sutta i ruoti, già 'n'ammazzai dui*. U2= *U sacciu, cj'era… <<1P>> a vipera a 'mmazzai*; U1 [ride] (trad.it dell'autore U1= *l'ulivo*; U2= *Ah?*; U1= *Ho beccato un animale sotto le ruote, già ne ho ammazzati due*. U2= *Lo sò, c'era… <<1P>> la vipera l'ho ammazzata*; U1 [ride]). I sospettati vengono scarcerati.

[51] p.p. nr. 8570/14 RGDDA, Procura della Repubblica c/o il Tribunale di Catanzaro. La trascrizione di un documento simile offre interpretazioni diverse da parte del lettore: avvocato, pubblico ministero, operatore di Polizia giudiziaria e giudice. È ovvio che il rumore metallico può essere l'inserimento della cintura di sicurezza o qualunque altra cosa. Dal rumore di due metalli che si urtano si può dedurre solo che sono due metalli e non certo la dettagliata descrizione riportata nel verbale di trascrizione.

Il ricorso ad interpretazioni personali basate sulla conoscenza del caso ad esempio, così come la propria professione (agente di polizia o investigatore), può introdurre un pericoloso elemento di soggettività: quello che Goodwin (1994) chiama *propria visione professionale*.

Spesso la trascrizione forense è redatta in lingua italiana, mentre invece il sonoro è in lingua dialettale, quindi il perito trascrittore effettua una doppia interpretazione prima delle informazioni mancanti ed in seguito della possibile/probabile traduzione dal dialetto all'italiano, con tutte le conseguenze del caso quando il dialetto non è molto noto (come riportato nelle note 29 e 50). 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. In un dialetto della Calabria settentrionale le due frasi seguenti sono differenziate dalla presenza di un suono occlusivo velare sordo [k] o affricato palatale sordo [tʃ] rispettivamente: [tʃajuˈditːu] 'l'ho detto a lui' e [kajuˈditːu] 'che cosa ho detto'; nel trascrivere in dialetto utilizzando i caratteri dell'alfabeto italiano le due frasi diventano entrambe <c'haju dittu> creando ambiguità nella comprensione. Un altro esempio è l'opposizione fonologica tra il suono occlusivo alveolare sonoro /d/ e il suono retroflesso sonoro /ɖ/ in parole come [viˈdiːku] 'vi dico' e [viˈɖiːku] 'ombellico' ecc. Le differenze potrebbero essere segnalate riportandosi ad una norma della grafia dialettale o anche semplicemente inserendo nella relazione introduttiva una legenda con l'esplicitazione dei diversi simboli utilizzati (cfr. Romito & Frontera 2017; Tarasi *et al*. 2019).

Il giudice non *torna* quasi mai sulla bobina o sulla registrazione come invece dovrebbe fare, ma quasi sempre considera il verbale di trascrizione una prova veritiera di quanto accaduto.

Siamo consapevoli che nessuna trascrizione potrà mai contenere tutte le informazioni presenti in una comunicazione orale, eppure in ambito forense si preferisce la trascrizione alla traccia audio. Abbiamo già detto che la traccia audio è caratterizzata da un numero indefinito di voci, che l'argomento della discussione è noto solo agli interlocutori, che spesso è di bassa qualità e avviene in un ambiente rumoroso, che l'intensità delle voci degli interlocutori oscilla in relazione alla loro posizione rispetto alla microspia e che ogni trascrizione è frutto di un'interpretazione. È necessario, infatti, identificare gli interlocutori e assegnare

loro i singoli turni di parole[52], è necessario comprendere parole pronunciate sottovoce, coperte da rumore o da altre voci[53], comprendere nomi, soprannomi e toponimi repertorio delle conoscenze condivise degli interlocutori e normalmente svincolate dal contesto[54] (cfr. Romito 2005), riconoscere ed interpretare differenti modelli intonativi[55] e tradurre tutte

[52] Nel p.p. nr 2139/14 RGNR mod 21 del Tribunale di Siracusa, viene contestata l'attribuzione della voce in una trascrizione della Guardia di Finanza: "In relazione a quanto delegato con la nota posta in riferimento, si precisa che è stata acquisita la traccia audio relativa all'interrogatorio di [omissis], al fine di un confronto con l'audio relativo alle frasi che lo stesso avrebbe proferito nell'incontro che si teneva in data […]. Da un attento riascolto effettuato delle conversazioni e comunicazioni tra presenti sembrerebbero non attribuibili al predetto solamente le seguenti frasi: a) verbale di trascrizione integrale delle conversazioni e comunicazioni ambientali – progressivo n. 22".

[53] Si veda ad esempio il caso del processo denominato *Gioco d'azzardo* e riportato nella nota 59 dove le perizie effettuate sulla stessa porzione di segnale sono almeno 7, o il p.p. nr. 2507/2015 r.g.n.r. Tribunale di Catanzaro, dove per la stessa identica porzione di segnale, gli operatori di PG trascrivono: "non ve lo eravate messi il capuccio?", il consulente del pubblico ministero trascrive: "*un vilati misu u capucciu*"; la Squadra Mobile trascrive: "non vi siete messi il capuccio?"; uno dei consulenti della difesa (linguista) scrive: "nel loro insieme, le analisi effettuate e i dati ottenuti portano alla conclusione che il segnale in cui era ipoteticamente contenuta la frase «un vilati misu u capucciu» non permette la trascrizione oggettiva di nessun tipo di frase"; un secondo consulente della difesa (ingegnere) riporta: "il rapporto segnale/rumore del segnale non permette alcuna trascrizione"; e infine un terzo consulente (trascrittore) della difesa scrive: "pure per me si è incapucciato!".

[54] Nel p.p. nr. 102/2018 del Tribunale di Parma viene contestata la seguente trascrizione: "U1 = non trovano niente neanche in sede su Zara; adesso l'avvocato andrà … andrà avanti col penale eccetera eccetera eccetera" perché "su Zara" è stato erroneamente confuso con il toponimo Suzzara. La corretta trascrizione quindi dovrebbe essere: "U1= non trovano niente neanche in sede a Suzzara; adesso l'avvocato andrà … andrà avanti col penale eccetera eccetera eccetera".

[55] L'aspetto intonativo è molto importante nel parlato e i segni di interpunzione utilizzabili nello scritto spesso sono insufficienti. Nell'esempio: "U1= *Sai che è morto Giovanni?* U2= *No*" se nella traccia sonora, il "no" di U2 è stato prodotto con un allungamento vocalico e un andamento prosodico lungo e costante (informazione non presente nella trascrizione), l'interpretazione (o la volontà del parlante) non è la negazione, ma piuttosto l'incredulità e lo sconcerto, quindi come un *non lo sapevo, non me lo aspettavo, mi dispiace moltissimo.* Il modello intonativo può anche essere erroneamente interpretato dal trascrittore e quindi essere riportato in maniera fuorviante nella propria trascrizione attraverso una errata posizione dei simboli di interpunzione. Nel seguente esempio presente in Azzalini (2017: 112): (1) ma se parte la prima botta a Francesco, lo stermino con tutta la famiglia; (2) ma se parte la prima botta, a Francesco lo stermino

queste informazioni su carta[56], cercando di fornire un'obiettiva ricostruzione dell'accaduto[57].

Non è facile effettuare una trascrizione forense ed è per questo motivo che è necessario che venga effettuata da un esperto. La comprensione del significato delle singole frasi assume una rilevanza fondamentale in ambito giudiziario ed investigativo; gli eventi, i fatti e le situazioni che la lingua descrive nella trascrizione di un'intercettazione possono diventare importanti indizi, se non prove di colpevolezza o innocenza (cfr. Romito 2013: 315). La delicata operazione di passaggio dall'oralità alla scrittura, come è noto, porta con sé diversi limiti sia oggettivi che soggettivi. La convinzione che la trascrizione sia un verbale completo e obiettivo del parlato nasce dal considerare il parlato come una versione sonora dello scritto. Ciò che sentiamo o percepiamo è sempre frutto di un'interpretazione inconsapevole. L'atto del trascrivere comporta un'analisi dei dati a disposizione e rappresenta, in questo caso, un compromesso tra il segnale acustico, le conoscenze pregresse, le competenze e le aspettative (cfr. Romito 2013: 271).

Nonostante quanto appena riportato, la trasposizione dal parlato allo scritto è un'attività indispensabile nel procedimento penale, perché ogni azione deve trasformarsi in atto scritto, in documento, quindi è necessario che anche l'accademia italiana si occupi della trascrizione forense e della formazione dei periti trascrittori, nonché delle parti in causa.

con tutta la famiglia, in (1) Francesco *riceve la prima botta* e l'interlocutore minaccia di sterminare una terza persona e tutta la sua famiglia, mentre in (2) la minaccia di morte è indirizzata proprio a Francesco e alla sua famiglia.

[56] Sappiamo benissimo che il parlato è spesso confuso, i gesti articolatori si sovrappongono, i verbi non si accordano e i concetti non sempre vengono espressi completamente. Non è quindi facile ordinare il parlato attraverso la sequenza di semplici parole e l'uso standardizzato della punteggiatura (cfr. Romito & Frontera 2017).

[57] La trascrizione forense oltre alle parole, deve necessariamente contenere una chiave di lettura per il lettore. Nell'esempio seguente probabilmente l'intensità della voce associata al suono del clacson spinge il trascrittore ad *interpretare* che U1 parla con una terza persona all'esterno dell'automobile e che quindi il nome di U2 non è Giuseppe: "U1= Ohu /../ assa u viju comu a mentisti (!) (-) […]; U2=Aund'è <<PP>> (?) sa' chi ffai (?); U1= eh (?) U2= stringila bene (!); [colpo di clacson, rivolto a qualcuno all'esterno] U1= Oh Giuseppe! Roba di elettrica ne vendi tu?" (U1= ehi /../ fammi vedere come l'hai messa (!) (-) […]; U2=dov'è <<PP>> (?) sai che cosa fai (?); U1= eh (?) U2= stringila bene (!); [colpo di clacson, rivolto a qualcuno all'esterno] U1= Oh Giuseppe! Materiale elettrico ne vendi tu? trad. it. dell'autore). Senza il commento del trascrittore l'interpretazione del lettore sarebbe completamente differente.

7. *Le linee guida per una corretta trascrizione forense*

La trascrizione forense, fin dalla sua prima apparizione nelle aule dei tribunali, è stata considerata un procedimento così semplice da non richiedere studi approfonditi o specializzazioni proprie. Non essendo necessario in questo caso l'uso di particolari metodiche o di sofisticata strumentazione, è invalsa la prassi che chiunque, purché munito di registratore, cuffia e buon orecchio, possa espletare in modo soddisfacente qualsiasi trascrizione. In realtà trascrivere una comunicazione orale comporta una serie di problemi che solo un esperto riesce ad intuire e a controllare, come abbiamo notato nel paragrafo precedente.

Sarebbe necessario identificare due tipi di trascrizione forense: la trascrizione forense *integrale* (nell'accezione giuridica già discussa) e la trascrizione delle *produzioni controverse,* cioè parole o frasi con grande carico informativo e incriminante, caratterizzate da una bassa qualità e spesso origine di differenti interpretazioni e trascrizioni[58].

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.

Nel caso della trascrizione delle *produzioni controverse,* l'esperto-trascrittore deve associare un'analisi acustico-linguistica che motivi le proprie scelte, come ad esempio l'individuazione di una certa vocale attraverso la misurazione acustica delle frequenze formantiche, la compatibilità di una parola attraverso la presenza di suoni fricativi o affricati con fruscìo ad alta frequenza che normalmente sovrasta i rumori di fondo, o la compatibilità della durata di una parola comparata con la velocità dell'eloquio o con la velocità di articolazione del parlante che si sta trascrivendo.

Entrambe le trascrizioni devono essere introdotte da considerazioni sulla qualità del segnale registrato, sull'ambiente, sul numero dei parlanti e sull'attendibilità[59].

---

[58] Per alcuni esempi reali si veda Romito *et al*. (2017: 126-129).

[59] Il concetto di valutazione e di accuratezza di una trascrizione, in un'aula di tribunale, assume nuovi significati. Dovrà essere l'esperto a definire il grado di attendibilità o di

È necessario che i trascrittori e le parti del processo (avvocati e giudici) condividano una stessa norma[60], oggi inesistente in ambito forense, perché il trascrittore non ha un percorso formativo uniforme[61] e in secondo luogo perché il committente (cioè il tribunale) sottovaluta il compito da assegnare, poiché nomina come perito trascrittore individui dalle competenze molto diverse e spesso discutibili.

È noto a tutti che oggi il pubblico ministero utilizzi parti di trascrizioni per la propria richiesta di custodia cautelare, la difesa utilizzi spezzoni di trascrizione per la propria arringa o relazione difensiva e infine il giudice riporti nella propria sentenza frasi o parole estrapolate dalle trascrizioni di intercettazioni.

La competenza dell'esperto-trascrittore e una norma condivisa tra tutte le parti di un processo dovrebbero garantire una lettura quanto più possibile oggettiva di una perizia di trascrizione. In molti atti giudiziari si nota una lettura "interpretata" della trascrizione forense. Ad esempio nel verbale di trascrizione presente nel documento di richiesta di custodia cautelare presentato dal pubblico ministero nell'ambito del p.p. nr 466/18, Tribunale di Lamezia Terme si notano molti omissis, ma soprattutto una propria interpretazione riguardo il significato di *singole parole* decontestualizzate.

accuratezza di una porzione di segnale e nei casi dubbi definire una porzione non trascrivibile. Tale decisione dovrà essere vincolante anche nei confronti del giudice e delle singole parti, in quanto una *opinione* su un segnale degradato non ha alcun fondamento scientifico. I criteri per tale definizione dovrebbero essere pienamente condivisi così da scongiurare casi come quello del processo denominato *Gioco d'azzardo* del Tribunale di Reggio Calabria, Catanzaro, Roma e Lecco, dove le nove diverse trascrizioni effettuate sullo stesso segnale sonoro intercettato contengono da 3485 parole a 415 o addirittura alla dichiarazione di "segnale non trascrivibile". La dichiarazione di segnale non trascrivibile deve essere effettuata attraverso dati acustici ed oggettivi, ad esempio misurando l'indice di intelligibilità, o il rapporto tra segnale informativo e segnale disturbante (Romito 2005).

[60] Un esempio che spiega bene la mancanza di norma è l'utilizzo dei puntini sospensivi nelle trascrizioni forensi. Questi vengono utilizzati per segnalare porzioni incomprensibili di segnale, intonazione sospensiva, allungamenti, pause, ecc. Nel p.p. nr 1369/2018 RGNR mod. 21 del Tribunale di Catanzaro, i puntini sospensivi vengono inseriti nella trascrizione di intercettazione alla fine di una parola lasciando intendere che la stessa abbia intonazione sospensiva, in altri casi invece sono posti alla fine di una porzione incomprensibile o all'inizio di una frase, cosa incompatibile con il livello intonativo: "C1= Lei ti ha visto, Lei ti ha visto...; C1= Andate a vedere *incomprensibile* … ; C1 = …. Io purtroppo … come … poi me l'ha spiegato ….".

[61] Cfr. Romito (2016).

Dalle intercettazioni si evince infatti che esistono luoghi di occultamento sicuri, comuni e/o disponibili a tutti i partecipanti: *solito posto; nella macchina; li ho cacciati dalla macchina; nel sellino; al muretto scalinata*; che viene utilizzato un linguaggio codificato e noto ai clienti: *un caffè; un caffettino; sigaretta; babà, rilassarsi*; che le somme vengono richieste sempre in maniera codificata: *dieci minuti; venti minuti* ecc.[62]

Inoltre, al fine di supportare le parti nella consultazione delle trascrizioni, riteniamo sia fondamentale l'uso di software dedicati che permettano un accesso immediato alla traccia sonora e la riformulazione del concetto di documento in ambito forense che, in relazione alle trascrizioni, deve effettuare il passaggio da cartaceo a multimediale.

8. *Conclusioni*

La competenza di un trascrittore forense non si limita alle tecniche di trascrizione o di verbalizzazione, ma per assolvere al proprio compito deve comprendere competenze differenti e interdisciplinari. Per definire l'originalità di una registrazione[63], l'esperto dovrà avere competenze acustiche e informatiche; per identificare i singoli locutori[64] presenti in

[62] A questo proposito la Cassazione penale, sez. VI, 24/03/2010, n. 16823, riporta che il giudice può consultare le trascrizioni effettuate dalla polizia giudiziaria: "in sede di giudizio abbreviato, il giudice può valutare le trascrizioni sommarie compiute dalla polizia giudiziaria circa il contenuto di conversazioni telefoniche oggetto di intercettazione (cosiddetti "brogliacci"), essendo utilizzabili ai fini della decisione tutti gli atti che siano stati legittimamente acquisiti al fascicolo del pubblico".

[63] Si veda ad esempio l'incarico dato ad un perito trascrittore nell'ambito del p.p. nr 570/16 RGNR dal Tribunale Ordinario di Lamezia Terme: "trascrizione di tutte le conversazioni di cui agli elenchi in atti […] ponendo in particolare ulteriore quesito al perito sull'originalità delle registrazioni in suo possesso ed obbligo di informare il Tribunale in caso di risposta negativa".

[64] P.p. 2139/14 RGNR e 10981/15 RG GIP Tribunale di Siracusa incarico estrapolato da verbale di trascrizione di udienza: "dopo l'esecuzione della misura cautelare alcuni indagati hanno contestato il contenuto della trascrizione, […] che […] ha fatto la Guardia di Finanza, che è l'organo investigativo, quindi l'oggetto del suo incarico è duplice, da un lato è necessario che […] si attribuisca la voce, le voci della intercettazione ad un indagato piuttosto che a un altro o persone che non risultano essere tra gli indagati in questo procedimento, […] dall'altro è necessario […] procedere alla trascrizione del contenuto delle conversazioni".

una conversazione, l'esperto dovrà avere competenze di fonetica per identificare qualche particolarità individuale durante la produzione di un suono da parte di un parlante o la pronuncia di una certa variabile dialettale o sociolinguistica; dovrà avere competenza di acustica di ambienti per intuire, attraverso analisi indirette, il luogo in cui è avvenuta una registrazione; e ancora dovrà avere competenza in analisi del segnale per identificare i diversi rumori ed eventualmente filtrarli per migliorare l'intelligibilità senza alterare il segnale[65]; competenze giuridiche per poter svolgere al meglio il proprio compito senza commettere errori formali.

Concludendo, riteniamo fondamentale istituire una nuova figura professionale[66], un percorso formativo istituzionale[67] e l'inserimento della disciplina della linguistica forense nei corsi di studi in giurisprudenza.

Riferimenti bibliografici

Azzalini, Irene. 2017. Il brogliaccio d'ascolto: passaggio dall'orale allo scritto nelle indagini. In Romito, Luciano & Frontera, Manuela (a cura di), *La scrittura all'ombra della parola*, 105–121. Milano: Officinaventuno.

Bazzanella, Carla. 2008. *Linguistica e pragmatica del linguaggio. Un'introduzione*. Bari: Laterza.

Cardona, Giorgio. 1989. *Dizionario di Linguistica*. Roma: Armando Editore.

CLIPS www.clips.unina.it/it/ (consultato il 31/10/2020).

Cronin, Michael & Romito, Luciano & Albanese, Maria. 2013. La traduzione. In Romito, Luciano (a cura di), *Manuale di Linguistica forense*, 307–320. Roma: Bulzoni.

Dubois, Jean & Giacomo, Mathée & Guespin, Louis & Marcellesi, Christiane & Marcellesi, Jean-Baptiste & Mével, Jean-Pierre. 1979. *Dizionario di Linguistica*. Bologna: Zanichelli.

[65] Proc. n.01/2019 CAA Tribunale di Perugia: "effettui il perito la trascrizione delle tracce audio […] previo filtraggio delle stesse".

[66] Al momento soltanto sei regioni (Toscana, Lazio, Marche, Abruzzo, Basilicata e Calabria) hanno riconosciuto ufficialmente le figure professionali di "Tecnico di (o addetto alla) analisi e trascrizione di segnali fonici" e "Tecnico di (o addetto alla) gestione della perizia di trascrizione in ambito forense", proprio grazie ad iniziative promosse dall'Osservatorio sulla Linguistica Forense. Ovviamente la figura professionale garantisce anche un compenso adeguato per il lavoro svolto.

[67] In Italia, al momento sono attivi solo corsi privati a pagamento.

Fraser, Helene. 2003. Issues in transcription: Factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics* 10(2). 203–226.

Galatà, Vincenzo. 2013. Aspetti tecnici sulle intercettazioni: analisi dei segnali e dei supporti. In Romito, Luciano (a cura di), *Manuale di Linguistica forense*, 123–172. Roma: Bulzoni.

Goffman, Erving. 1987. *Forme del parlare*. Bologna: Il Mulino.

Goodwin, Charles. 1994. Professional vision. *American Anthropologist* 96(3). 606–633.

Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Lerner, Gene (ed.), *Conversation Analysis: Studies from the first generation*, 13–31. Amsterdam: John Benjamins.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H Theory. In Hardcastle, William J. & Marchal, Alain (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer Academic Publishers.

Mehrabian, Albert. 1972. *Non-verbal communication*. New Brunswick: Aldine Transaction.

Minissi, Nullo. 1990. *La scrittura fonetica*. Firenze: La Nuova Italia Scientifica.

Olsson, John. 2004. *Forensic Linguistics: An introduction to Language, Crime and the Law*. London: Continuum.

Romito, Luciano. 2005. Il contesto, l'intelligibilità, il rapporto segnale–rumore. In Cosi, Piero (a cura di), *Misura dei parametri. Aspetti tecnologici ed implicazioni nei modelli linguistici*, 539–566. Torriana: EDK Editore.

Romito, Luciano. 2010. Le intercettazioni. In Caligiuri, Mario (a cura di), *Cultura della Legalità*, 207–217. Catanzaro: Rubbettino.

Romito, Luciano. 2013. La linguistica forense. In Romito, Luciano (a cura di), *Manuale di linguistica forense*, 173–306. Roma: Bulzoni.

Romito, Luciano. 2016. La competenza linguistica nelle perizie di trascrizione e di identificazione del parlatore (a margine, alcune riflessioni sul difficile rapporto tra scienza e processo). *Diritto Penale Contemporaneo*. https://archiviodpc.dirittopenaleuomo.org/upload/1453648881ROMITO_2016a.pdf. (consultato il 25.11.2020).

Romito, Luciano & Galatà, Vincenzo. 2008. Speaker Recognition in Italy: Evaluation of Methods used in Forensic Cases. *Language Design* 1. 229–240.

Romito, Luciano & Ciardullo, Maria Assunta & Frontera, Manuela & Bianchi, Francesca. 2016. Analisi conversazionale e (a)simmetria dei ruoli nel parlato intercettato. In Andorno, Cecilia & Grassi, Roberta (a cura di), *Le dinamiche dell'interazione Prospettive di analisi e contesti applicativi*, 333–342. Milano: AItLA.

Romito, Luciano & Frontera, Manuela. 2017. La trascrizione forense di intercettazioni ambientali: una proposta di metodologia procedurale. In Romito, Luciano & Frontera, Manuela (a cura di), *La scrittura all'ombra della parola*, 121–139. Milano: Officinaventuno.

Romito, Luciano & Tarasi, Andrea & Graziano, Elvira. 2017. Un modello per l'annotazione di fatti prosodici nelle trascrizioni forensi. In Romito, Luciano & Frontera, Manuela (a cura di), *La scrittura all'ombra della parola*, 139–155. Milano: Officinaventuno.

Sinatra, Chiara. 2014. Il passaggio dall'oralità alla scrittura in ambito forense e giudiziario. *Cuadernos AISPI* 4. 197–212.

Tarasi, Andrea & Graziano, Elvira & Romito, Luciano. 2018. Il complesso rapporto tra parlato e scritto nei dialetti calabresi. In Carbonara, Valentina & Cosenza, Luana & Masillo, Paola & Salvati, Luisa & Scibetta, Andrea (a cura di), *Il parlato e lo scritto: aspetti teorici e didattici*, 387–399. Pisa: Pacini Editore.

Vernero, Irene & Romano, Antonio. 2017. La trascrizione del parlato patologico. In Romito, Luciano & Frontera, Manuela (a cura di), *La scrittura all'ombra della parola*, 11–32. Milano: Officinaventuno.

Jacopo Saturno
(Università degli studi di Bergamo)

La trascrizione di dati linguistici – istruzioni di base

1. *Come trascrivere*

Il testo che segue propone un'introduzione di base alla trascrizione di dati linguistici utilizzando un numero ridotto di risorse elettroniche, più o meno specialistiche ma tutte disponibili gratuitamente[1]. Il suo obiettivo non è tanto di trattare in maniera approfondita tutte le funzionalità dei programmi considerati, quanto piuttosto di descrivere alcune semplici procedure di base utili a trascrivere dati linguistici e presentarli in modo appropriato nell'ambito di testi stampabili, quali ad esempio una tesi di laurea. Particolare attenzione è dedicata al programma ELAN, il quale permette di allineare ciascun enunciato alla corrispondente porzione di traccia audio o video, nonché di rappresentare in modo graficamente efficace anche dinamiche interazionali complesse.

1.1 Operazioni di base: trascrizione di testi monologici

Il caso più semplice di trascrizione è quello in cui si debba trascrivere un testo monologico senza la necessità di ancorarlo alla traccia audio o video. A questo scopo è naturalmente possibile utilizzare programmi specifici come ELAN, trattato approfonditamente nelle sezioni seguenti, oppure affidarsi a metodologie improvvisate e poco efficaci che prevedano l'uso di strumenti nati per altri scopi, quali un programma di videoscrittura e un *player* audio o video. Tuttavia, se l'obiettivo è la sbobinatura di una traccia, può risultare conveniente affidarsi a

[1] In particolare le istruzioni si riferiscono ai programmi Atom (https://atom.io) e alla suite Libreoffice (https://it.libreoffice.org, in particolare Writer e Calc). Le medesime procedure si applicano con minime modifiche anche ad altri programmi equivalenti (es. la suite MS Office).

strumenti appositi, ma non per questo complessi. In questa sede presentiamo il sito https://otranscribe.com, il quale gratuitamente e senza registrazione offre la possibilità di caricare la traccia audio desiderata (oppure importarla da *youtube*) e trascriverla mediante un'interfaccia estremamente semplice (Figura 1). Molto utile risulta la possibilità di utilizzare dei semplici comandi da tastiera per avviare/fermare la traccia audio e per spostarsi avanti e indietro all'interno di essa (rispettivamente Esc, F1, F2). Il testo può essere poi copiato e incollato in qualsiasi programma di videoscrittura per la formattazione. È anche possibile inserire dei riferimenti temporali per ancorare almeno in parte il testo alla traccia audio, così da stabilire punti di riferimento potenzialmente utili per orientarsi in registrazioni particolarmente estese.

Fig. 1. *Interfaccia del servizio online* https://otranscribe.com

Nella pratica della ricerca linguistica, tuttavia, l'obiettivo del processo di trascrizione è di norma più complesso rispetto alla semplice sbobinatura. Per esempio, può capitare di dover trascrivere un'interazione tra più persone, la quale quasi certamente prevede fenomeni difficili da rendere graficamente in un programma di videoscrittura, come le sovrapposizioni di turno; oppure per un singolo enunciato possono rendersi necessarie annotazioni pertinenti più livelli di analisi (es. trascrizione fonetica, classi di parole); oppure ancora può risultare utile ancorare ciascun enunciato trascritto alla porzione corrispondente di traccia audio, così da poter facilmente riascoltare il frammento in questione. In tutti questi casi è opportuno affidarsi a uno strumento specializzato come ELAN (Brugman & Russell 2004). Si tratta di un programma utilizzabile liberamente[2], a condizione che sia debitamente citato nel prodotto finale della ricerca. Come già anticipato, questo capi-

[2] https://archive.mpi.nl/tla/elan/download

tolo presenterà solo una piccola gamma delle potenzialità offerte dal programma[3]: per una trattazione più completa si rimanda invece al manuale ufficiale[4].

Una volta installato e lanciato il programma, per creare un nuovo documento scegliamo "File > New". In ambiente MacOS, nella schermata che si apre, facendo clic su *"Add Media File"* è possibile selezionare uno o più file audio o video salvati sul computer. Per l'esempio proposto in questo capitolo, scegliamo il file "lista_parole1.wav" e facciamo clic su *"OK"* (Figura 2).

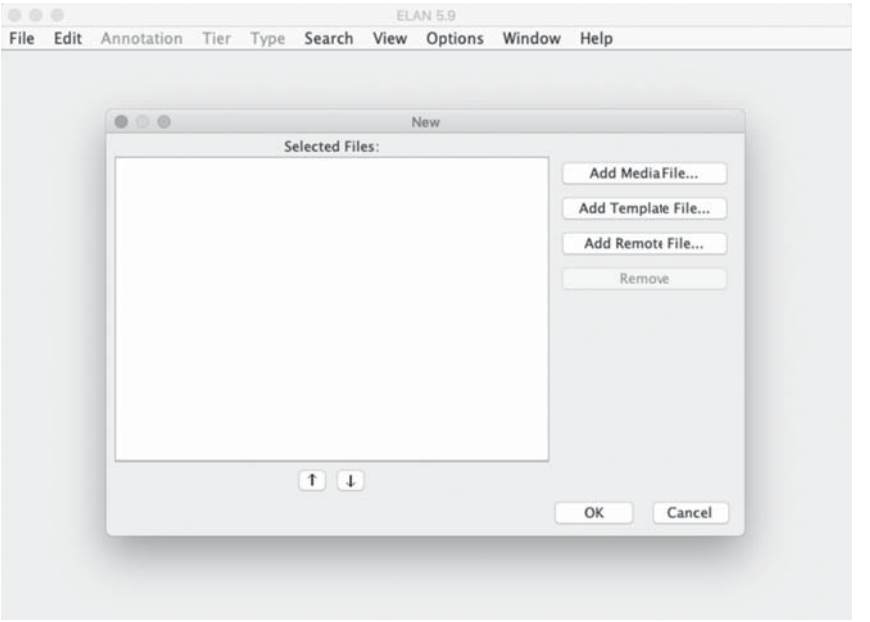


Fig. 2. *Collegamento di una traccia audio o video.*

Gli utenti Windows potranno invece creare un nuovo documento a partire dalla schermata presentata nella Figura 3. Agendo sul menu a tendina "*look in*" è possibile risalire alla cartella in cui è localizzato il file audio o video da trascrivere, di cui si può specificare il formato nel

[3] Il presente testo si riferisce alla versione 5.9.
[4] https://archive.mpi.nl/tla/elan/documentation

menu a tendina "*file format*". Una volta selezionato il file desiderato, lo si può aggiungere alla trascrizione utilizzando le doppie frecce a destra. Con il pulsante "OK" si confermano le scelte effettuate.

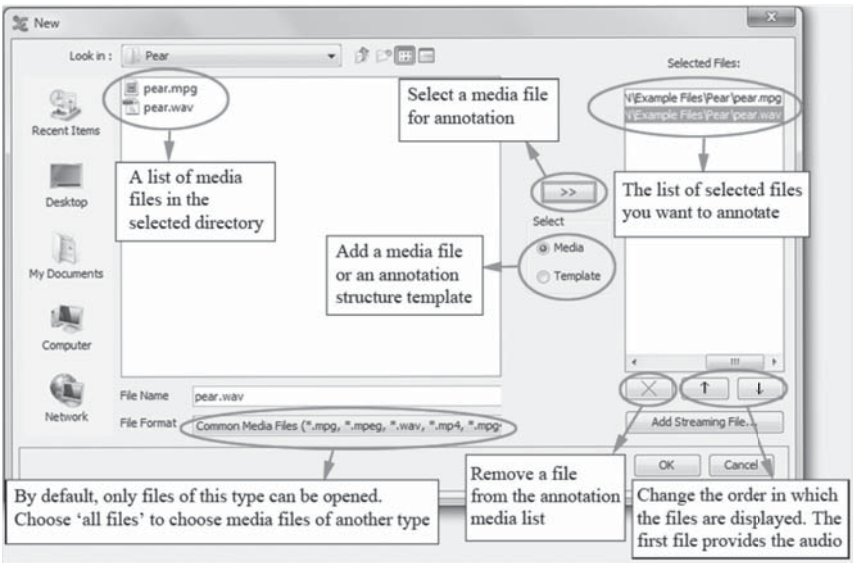


Fig. 3. *Creazione di una nuova trascrizione su Windows*[5].

Si aprirà a questo punto l'interfaccia principale del programma, nella quale si possono individuare sette grandi sezioni orizzontali, identificate nella Figura 4 dai rettangoli contrassegnati e dalle lettere A-G:

A. barra dei menu;
B. cursori che regolano rispettivamente il volume e la velocità di riproduzione della traccia;
C. comandi di spostamento nella traccia e fra le annotazioni;
D. forma d'onda della traccia audio. NB: per evitare problemi di compatibilità è consigliabile che le tracce siano in formato .wav;
E. area delle trascrizioni;
F. barra di spostamento orizzontale;
G. cursore dell'ingrandimento orizzontale.

[5] Immagine tratta dal manuale ufficiale di Elan: https://www.mpi.nl/corpus/html/elan/ch01s02s02.html

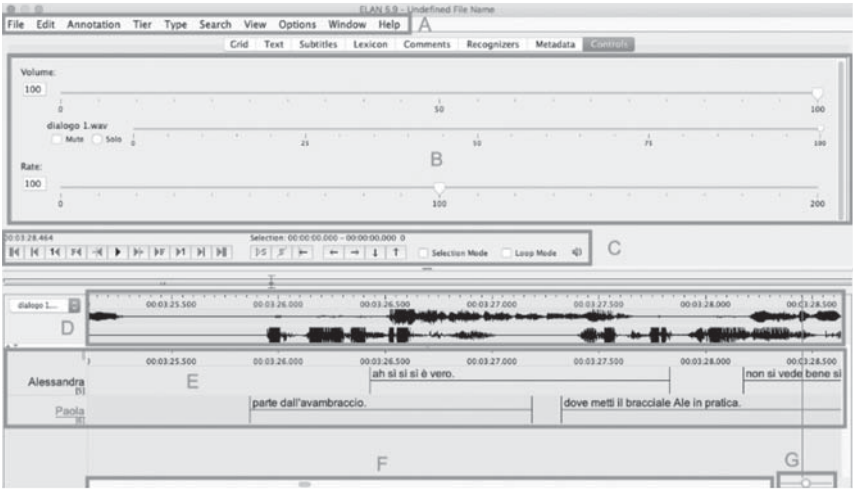


Fig. 4. *Interfaccia di ELAN.*

Il cursore dell'ingrandimento orizzontale regola l'ampiezza della porzione di traccia mostrata in una singola schermata. Minore il livello di ingrandimento, più ravvicinate tra loro appariranno le annotazioni. È utile modificare questo valore quando il testo di un'annotazione non è visibile interamente (Figura 5).

Fig. 5. *Zoom orizzontale.*

È possibile modificare la traccia audio o video associata al documento, così come anche aggiungerne di nuove. A questo scopo, nella barra dei menu è necessario selezionare *"Edit > linked files"* (Figura 6).

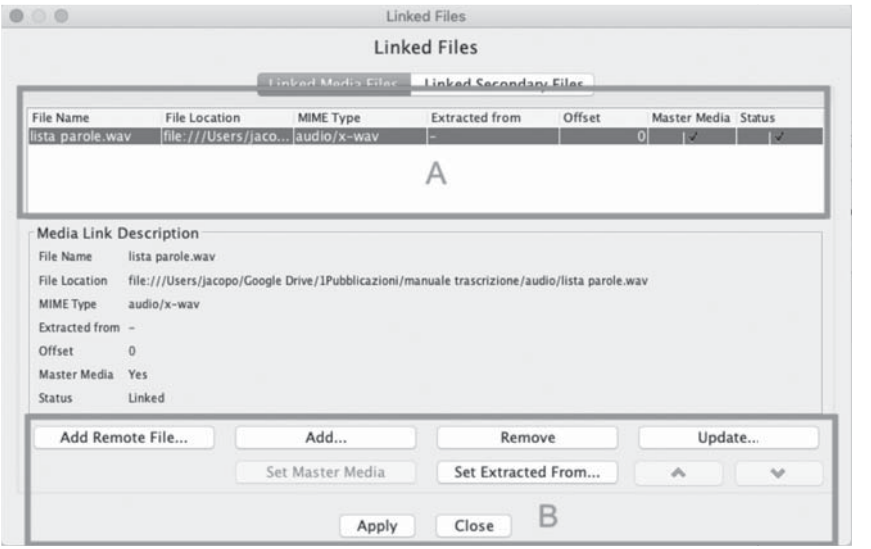


Fig. 6. *Modifica dei file audio e/o video associati alla trascrizione.*

L'unità di base di un documento di ELAN è la "annotazione" (ingl. *annotation*), cioè la trascrizione di una porzione di traccia audio. La definizione dei suoi confini dipende dalle scelte del trascrittore e dagli obiettivi della ricerca: può essere una singola parola, un enunciato (frase), un turno di parola etc. Per inserire un'annotazione è necessario dapprima agire sul mouse per selezionare nella forma d'onda la porzione di traccia corrispondente, che si potrà ascoltare premendo il pulsante *"Play selection"* (Figura 7). È possibile ripetere questo comando più di una volta per riascoltare la porzione di traccia, a condizione che rimanga selezionata la porzione di forma d'onda corrispondente.

Potremo poi digitare il contenuto facendo clic su *"Annotation > New Annotation Here"* (Figura 8) e premendo "Invio" dopo averne digitato il contenuto.

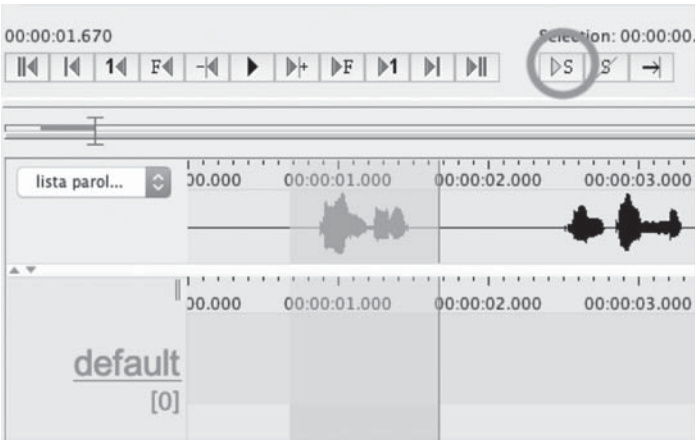


Fig. 7. *Selezione di una porzione di forma d'onda mediante il mouse.*

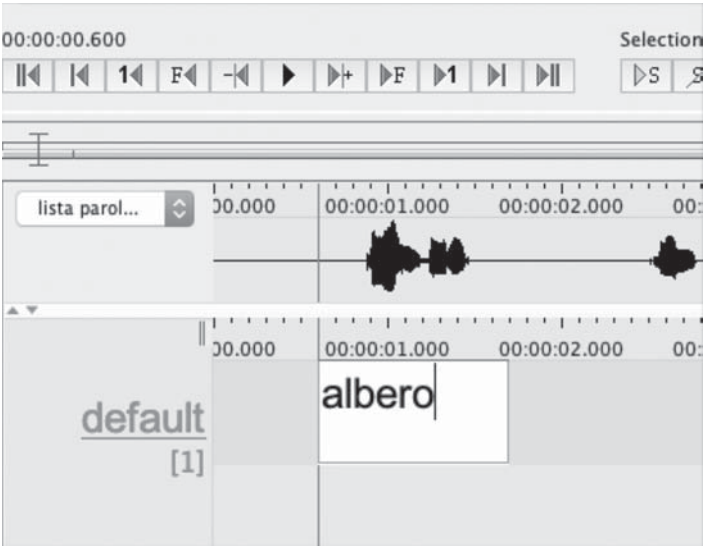


Fig. 8. *Inserimento di una nuova annotazione.*

Una volta inserita un'annotazione è possibile modificarne il contenuto facendo doppio clic su di essa (Figura 9).

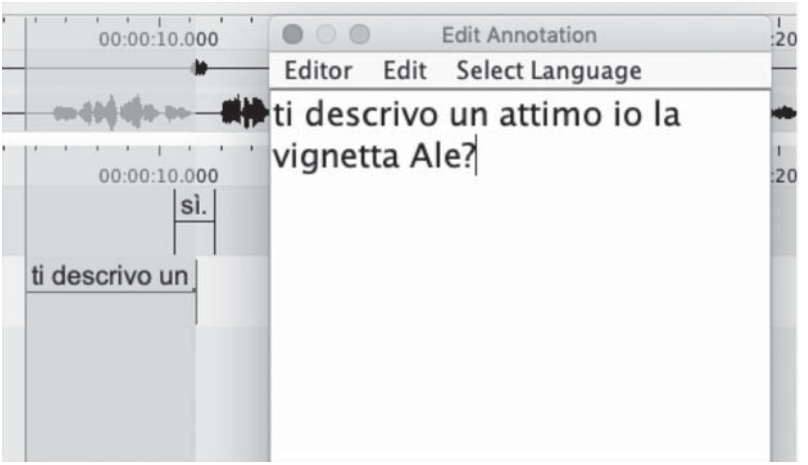


Fig. 9. *Modifica di un'annotazione già inserita.*

Selezionando la casella *"Selection Mode"* (cerchio a destra nella Figura 10) è possibile attivare una modalità di selezione alternativa, particolarmente utile per trascrivere rapidamente testi in cui gli enunciati si susseguono senza troppe sovrapposizioni tra partecipanti. In primo luogo, dalla barra dei menu scegliamo *"Edit > Preferences > Edit Preferences > Editing"* e verifichiamo che sia selezionata l'opzione *"Clear selection after creating or editing an annotation"*. Tornati alla schermata principale, facendo clic sul pulsante *"Play / Pause the media"* (cerchio a sinistra nella Figura 10) si fa partire la registrazione: spostandosi verso destra, il cursore evidenzierà automaticamente la porzione di traccia percorsa. Una volta raggiunta una selezione soddisfacente, fermiamo la riproduzione agendo sul medesimo pulsante e inseriamo il contenuto dell'enunciato scegliendo dalla barra dei menu *"Annotation > New Annotation Here"*. Premiamo poi nuovamente *"Play / Pause the media"* per proseguire nello stesso modo.

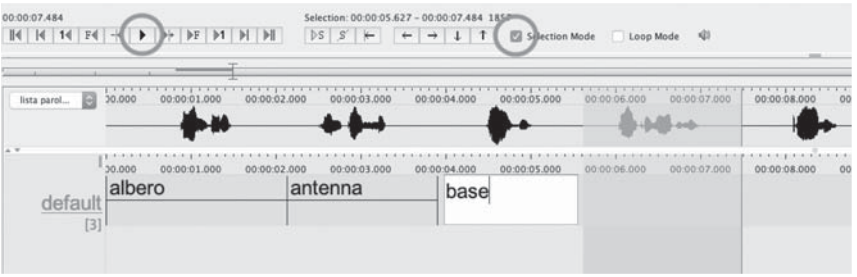


Fig. 10. *Trascrizione in modalità "Selection Mode".*

La velocità del processo di trascrizione può essere aumentata esponenzialmente utilizzando le scorciatoie da tastiera corrispondenti alle operazioni eseguite mediante il mouse. Scegliendo dalla barra dei menu *"Edit > Preferences > Edit Shortcuts"* è possibile visualizzare le scorciatoie predefinite, nonché modificarle secondo le proprie preferenze. A questo scopo, scegliamo il pannello *"Annotation Mode"* nella parte superiore della schermata, nel quale sono contenuti i comandi trattati finora. Una volta individuato quello che si intende modificare (es. *"Play / Pause the media"*), è sufficiente selezionarlo facendo clic su di esso e poi scegliere *"Edit Shortcut"* in basso a sinistra. Nella schermata che si apre è ora possibile inserire la scorciatoia da tastiera desiderata (es "F1") e confermare la scelta con *"Apply"* (Figura 11).

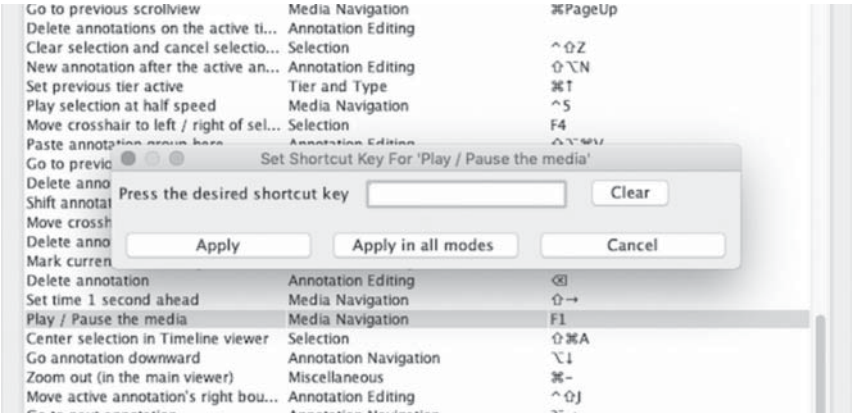


Fig. 11. *Personalizzazione delle scorciatoie da tastiera.*

La trascrizione in modalità *"Selection Mode"* potrà ora essere svolta con rapidità di gran lunga maggiore grazie a due sole scorciatoie opportunamente selezionate, cioè *"Play / Pause the media"* e *"New Annotation Here"*. Per fare ciò sarà sufficiente far partire la riproduzione, attendere che il cursore abbia evidenziato la porzione di audio desiderata, e premere il tasto corrispondente al comando *"New Annotation Here"*. Una volta terminato di digitare il contenuto, si potrà far ripartire la riproduzione, e così via.

Per salvare il file di trascrizione (formato ".eaf") scegliamo nella barra dei menu *"File > Save"* e indichiamo la cartella in cui desideriamo conservare il documento. È sempre consigliabile salvare il file non appena creato e poi ripetere frequentemente l'operazione.

1.2 Testi dialogici

ELAN possiede una struttura simile a quella di una partitura orchestrale, in cui ad ogni partecipante (nella metafora accomunato a uno strumento dell'orchestra) è associata una riga. Come nello spartito, le voci dei partecipanti possono sovrapporsi nella rappresentazione grafica, ricreando le sovrapposizioni di suono presenti nella traccia audio.

Per esplorare questa possibilità, creiamo un nuovo documento ELAN utilizzando una traccia audio in cui intervengono due interlocutrici, Alessandra e Paola. Impostiamo ora una linea di trascrizione per ciascun partecipante. La riga *"default"* utilizzata negli esempi precedenti è presente in ogni nuovo documento di ELAN. Per cambiarne il nome, nella barra dei menu selezioniamo *"Tier > Change tier attributes"*. Nel menu a discesa *"Select Tier"* scegliamo ora la riga che intendiamo modificare (Figura 12).

Nel campo *"Tier Name"* modifichiamo il nome della riga *"default"* in "Alessandra". Facciamo poi clic su *"Change"* e infine *"Close"*.

Per aggiungere una nuova riga, dalla barra dei menu scegliamo *"Tier > Add New Tier"*, provocando l'apertura di una finestra del tutto simile a quella precedente. Nel campo *"Tier Name"* indichiamo il nome "Paola" e facciamo clic su *"Add"*, poi *"Close"*. Abbiamo ora una riga per ciascun partecipante all'interazione (Figura 13). Per selezionare la riga su cui trascrivere è sufficiente fare doppio clic sul nome corrispondente oppure agire sulle scorciatoie da tastiera appropriate.

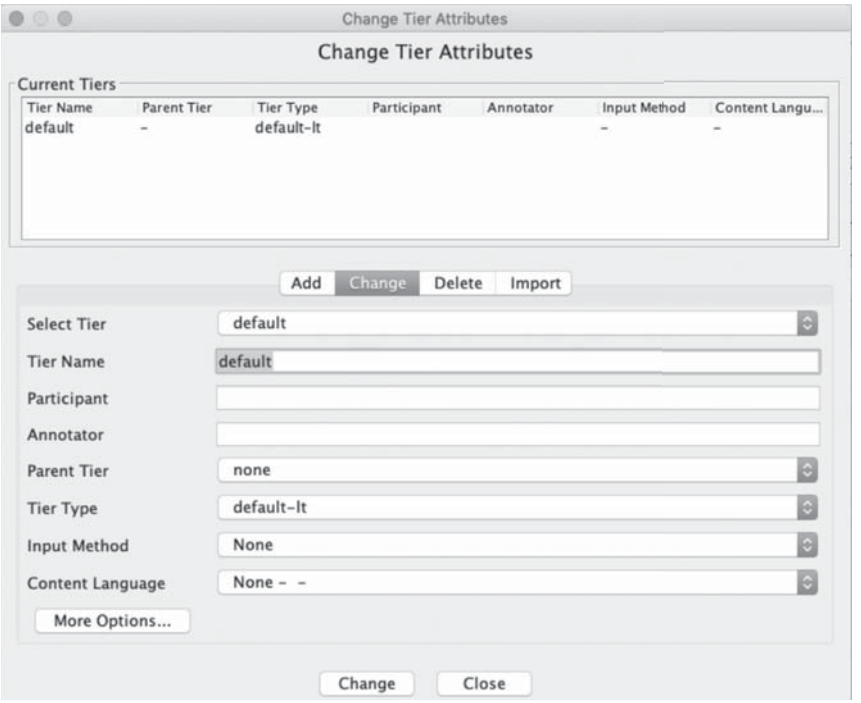


Fig. 12. *Modifica delle caratteristiche di una riga di trascrizione.*

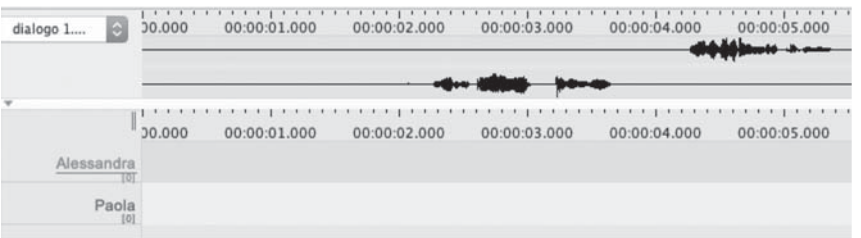


Fig. 13. *Partitura con righe multiple.*

Come si vede, la trascrizione su righe parallele offerta da ELAN è molto efficace per rappresentare graficamente le sovrapposizioni di turni (Figura 14).

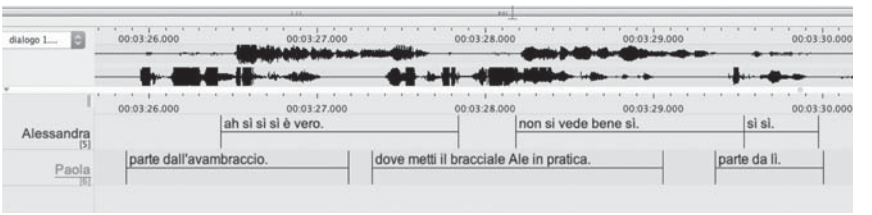


Fig. 14. *Trascrizione di scambi interazionali.*

Le righe di trascrizione non necessariamente devono corrispondere agli enunciati di un interlocutore. Al contrario, esse possono includere informazioni di qualunque altro tipo, quali commenti del trascrittore, trascrizioni fonetiche, traduzioni etc. A questo scopo è sufficiente aggiungere nuove righe di trascrizione applicando la procedura indicata in precedenza. Per allineare una nuova annotazione a un'altra già esistente è necessario in primo luogo selezionare la riga su cui si intende inserire la nuova annotazione. Facendo poi clic sull'annotazione già esistente si evidenzia la corrispondente porzione di forma d'onda (Figura 15).

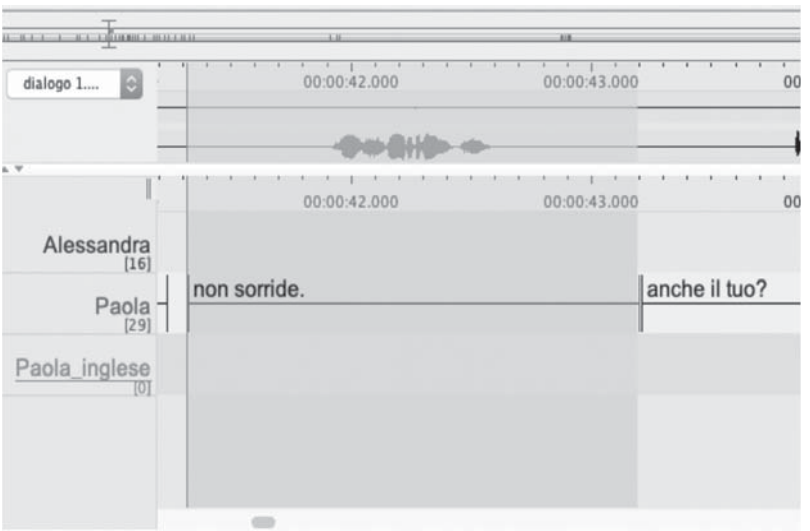


Fig. 15. *Selezione di un'annotazione esistente.*

A questo punto, scegliendo dalla barra dei menu *"Annotation > New Annotation Here"* o la scorciatoia da tastiera corrispondente, si potrà procedere a inserire il contenuto dell'annotazione sulla nuova riga, ma in corrispondenza dell'annotazione già esistente (Figura 16).

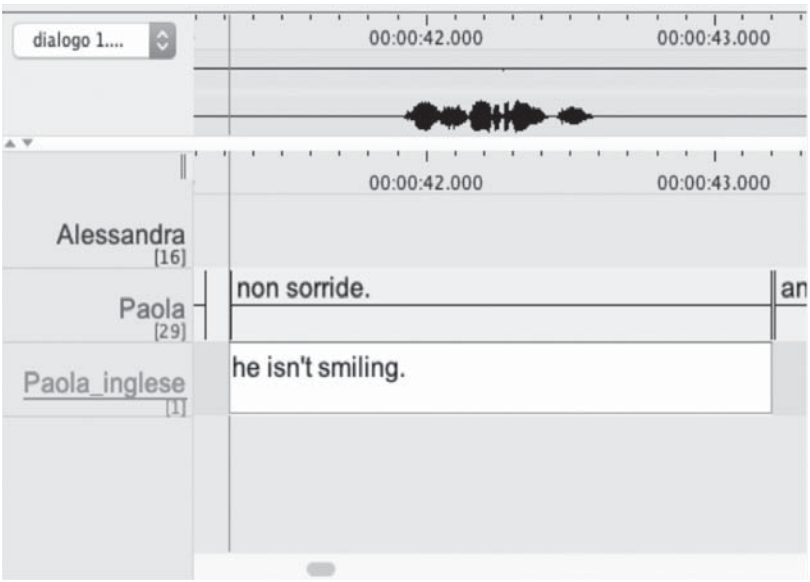


Fig. 16. *Inserimento di una nuova annotazione allineata a una esistente.*

1.3 Esportazione di trascrizioni

Per quanto il formato a partitura di ELAN sia estremamente efficace per rappresentare graficamente le dinamiche interazionali, lo sviluppo orizzontale delle sue trascrizioni risulta poco pratico per la stampa, per la quale invece è più appropriato uno sviluppo verticale. È tuttavia possibile esportare la trascrizione prodotta da ELAN in numerosi altri formati, molti dei quali associati a programmi specialistici. In questa guida ci si limiterà a mostrare una semplice procedura per esportare la trascrizione in un formato testuale, al fine di includerla in testi quali articoli scientifici o tesi di laurea.

Dopo aver completato la trascrizione, dalla barra dei menu scegliamo *"File > Export As > Traditional Transcript File"* (Figura 17). Ciò ci permetterà di creare un documento di testo semplice in formato ".txt" in cui gli enunciati di tutti i partecipanti sono ordinati verticalmente in base al momento in cui sono stati prodotti.

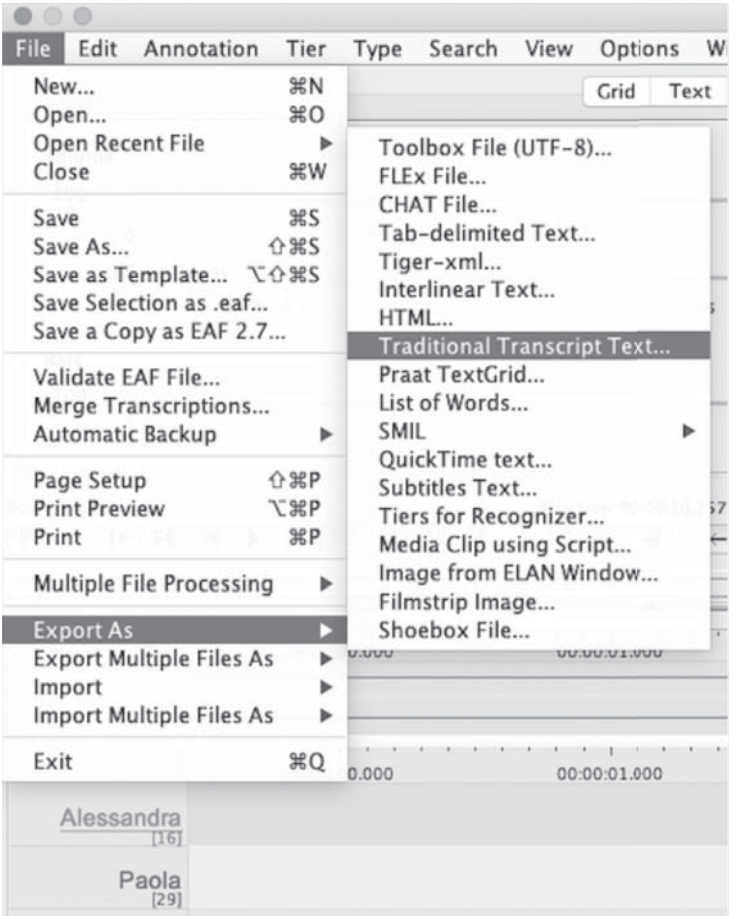


Fig. 17. *Esportazione della trascrizione.*

Nella schermata che si apre (Figura 18) selezioniamo le righe di trascrizione che ci interessa esportare (qui entrambe quelle corrispondenti alle partecipanti all'interazione, Alessandra e Paola). Spuntiamo inoltre l'opzione "*Wrap lines*" e inseriamo il valore "200". Tale parametro determina il numero di caratteri dopo il quale il testo va a capo: impostando un valore elevato, ci assicuriamo che una volta esportate nel pro-

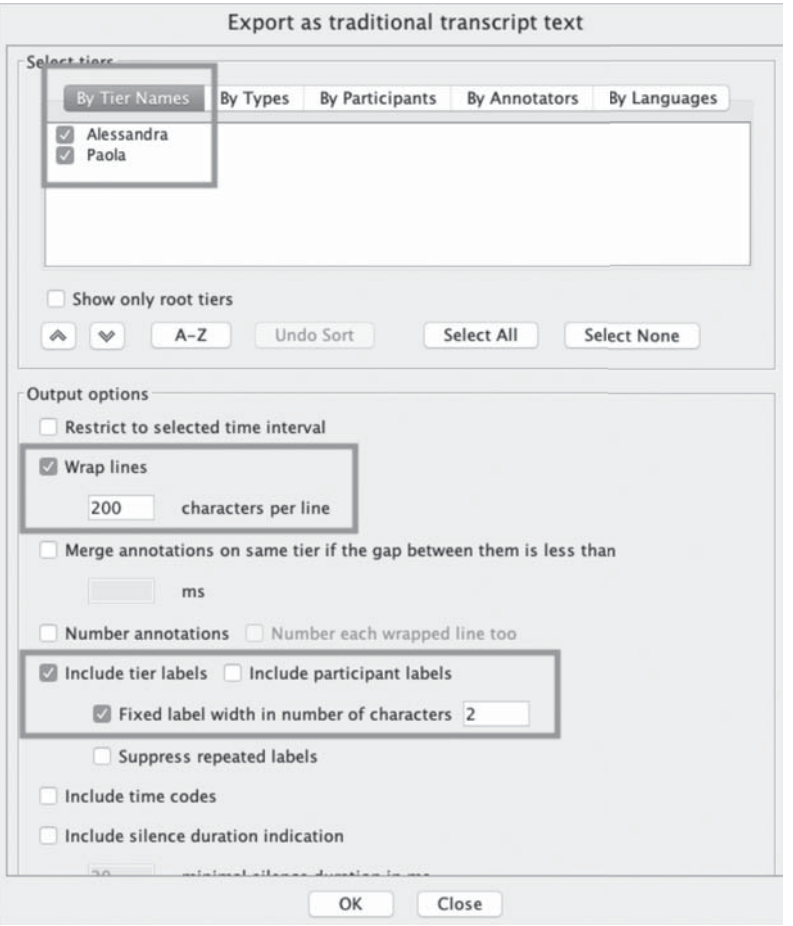


Fig. 18. *Esportazione come "Traditional Transcript Text".*

gramma di videoscrittura, le righe occuperanno tutta lo spazio disponibile in larghezza. Infine, spuntiamo le opzioni "*Include Tier Labels*", grazie alla quale gli enunciati saranno preceduti dal nome della riga corrispondente, e "*Fixed label width in number of characters*", impostando per quest'ultima il valore "2". In questo modo i nomi delle righe (qui, "Alessandra" e "Paola") saranno troncati dopo il numero di caratteri specificato (rispettivamente "Al" e "Pa").

Facciamo clic su *"OK"* e salviamo il documento nella posizione più comoda, facendo attenzione al fatto che il nome sia seguito dall'estensione ".txt" (Figura 19).

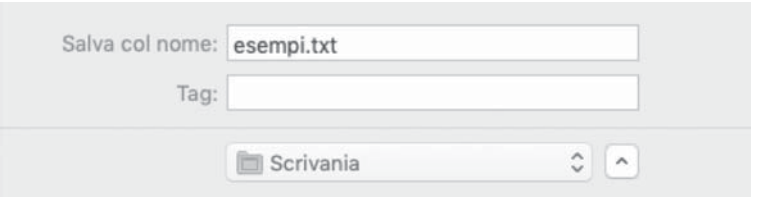


Fig. 19. *Salvataggio in formato ".txt".*

Nella finestra successiva verifichiamo che il menu a tendina presenti la dicitura "UTF-8" e facciamo clic su *"OK"* (Figura 20).

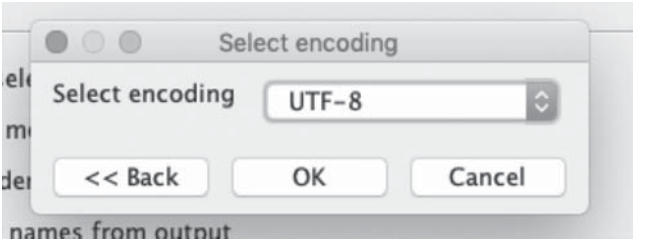


Fig. 20. *Salvataggio con codifica "UTF-8".*

Apriamo ora il documento esportato utilizzando l'editor di testo *Atom* (Figura 21).

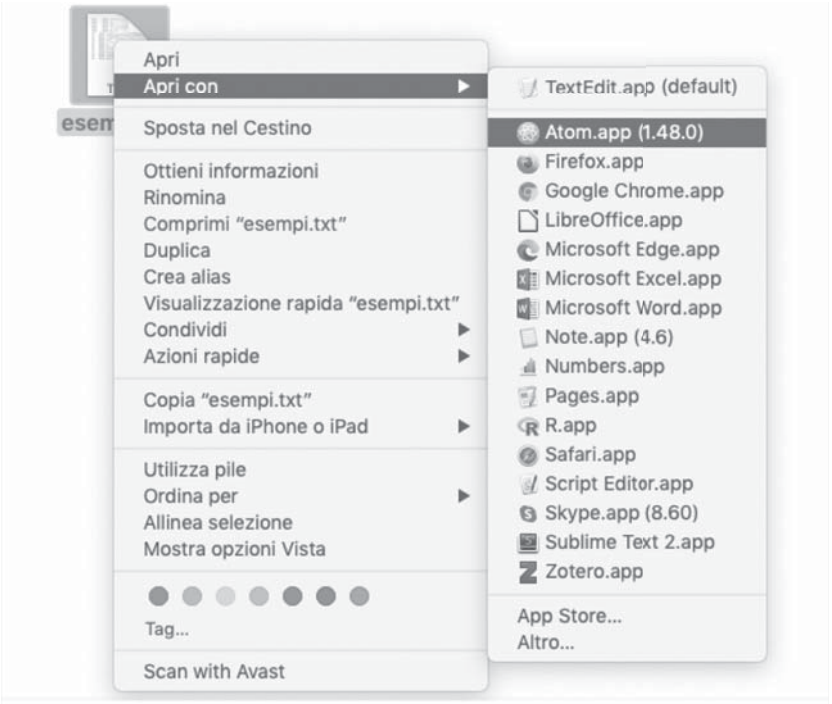


Fig. 21. *Apriamo il file ".txt" con l'editor di testo.*

Dal menu "*Find*" scegliamo "*Replace in Buffer*"[6] (Figura 22). Nella maschera che si apre in basso, scriviamo nel campo superiore (dedicato alla stringa da cercare nel documento) il nome di un partecipante (qui, per esempio, "Pa") seguito da tre spazi, i quali corrispondono ai caratteri utilizzati nel file in formato .txt per separare la colonna contenente i nomi delle righe da quella riportante il contenuto delle annotazioni. Nel campo inferiore (dedicato alla stringa con cui sostituire quella ricercata) scriviamo il nome completo del partecipante (qui, "Paola"; possiamo anche ripetere il nome abbreviato, se opportuno), seguito da "\t" (tabulazione: attenzione all'inclinazione della barretta!). Premiamo il pulsante "*Replace All*" e ripetiamo l'operazione per tutti i partecipanti.

[6] In altri editor di testo questo comando può essere denominato "(*Find and) Replace*" in inglese, "*(trova e) sostituisci*" in italiano.

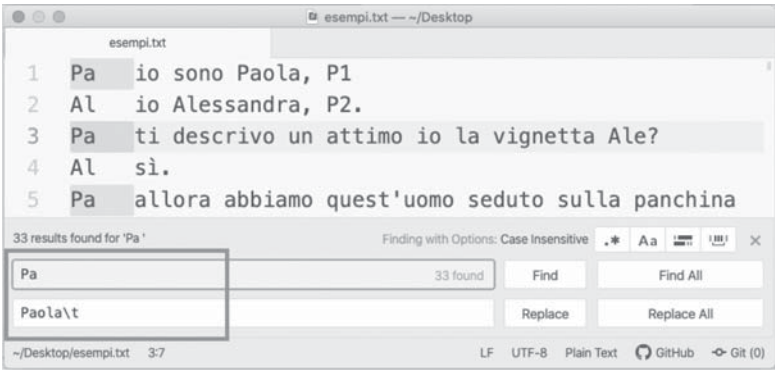


Fig. 22. *Trova e sostituisci in Atom.*

Selezioniamo quindi l'intero testo scegliendo "*Edit > Select All*" dalla barra dei menu e copiamolo con *"Edit > Copy"* (Figura 23).

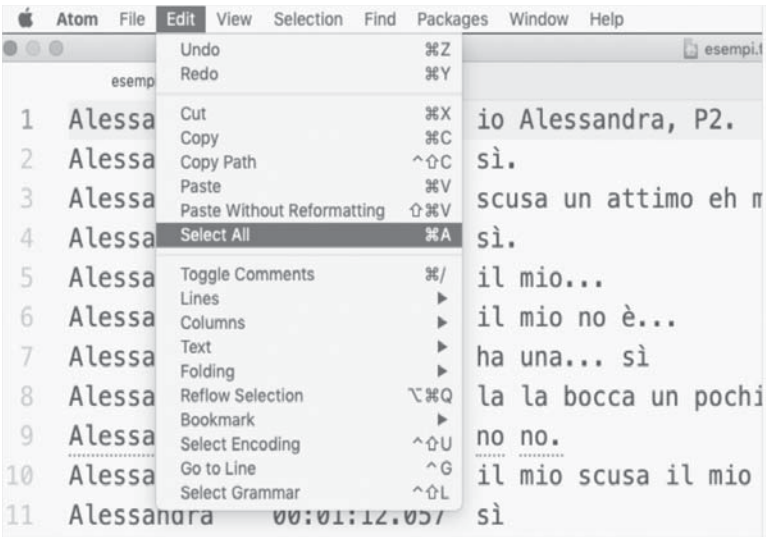


Fig. 23. *Seleziona tutto in Atom.*

Possiamo ora incollare il testo copiato in qualunque programma di videoscrittura, come LibreOffice Writer. Avendo utilizzato una tabulazione per separare il nome delle righe dal contenuto delle annotazioni, possiamo formattare il testo a nostro piacere. Per allineare in maniera uniforme le annotazioni è possibile agire sui due cursori visibili sul righello in alto (evidenziati dal rettangolo nella Figura 24), dei quali quello superiore indica la posizione della prima parola di ogni paragrafo (cioè del nome della riga, nella nostra trascrizione), mentre il secondo corrisponde alla posizione di tutte le parole seguenti.

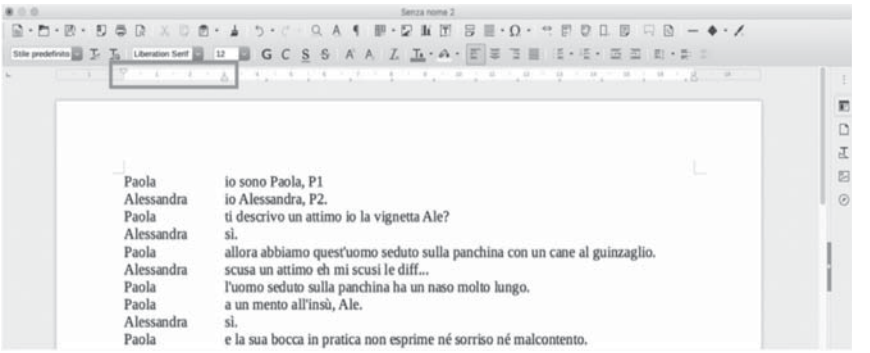


Fig. 24. *Formattazione del testo mediante tabulazioni.*

Alternativamente è possibile formattare la trascrizione come una tabella di due colonne, la prima contenente il nome del partecipante, la seconda il testo. A questo scopo, evidenziamo tutte le righe della trascrizione e dalla barra dei menu scegliamo *"Tabella > Converti > Testo in Tabella"* (ingl. *"Table > Convert > Text to Table"*) (Figura 25).

Dopo essersi assicurati che come separatore di testo siano impostate le tabulazioni (Figura 26), facciamo clic su *"OK"*.

Una volta esportata la trascrizione nel programma di videoscrittura è possibile elaborare il testo segmentandolo in più esempi o eliminando eventuali righe non pertinenti. Se opportuno, la trascrizione può essere formattata secondo le convenzioni di particolari standard, come ad esempio la *Conversation Analysis* (cfr. 1, Jefferson 2004: 14).

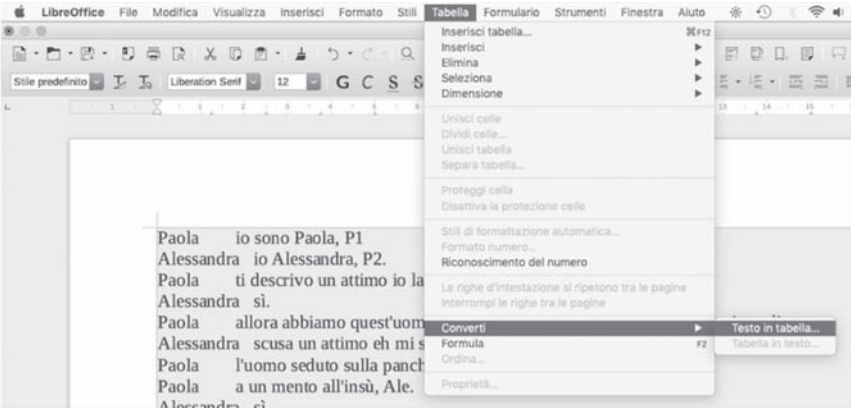


Fig. 25. *Converti testo in tabella.*

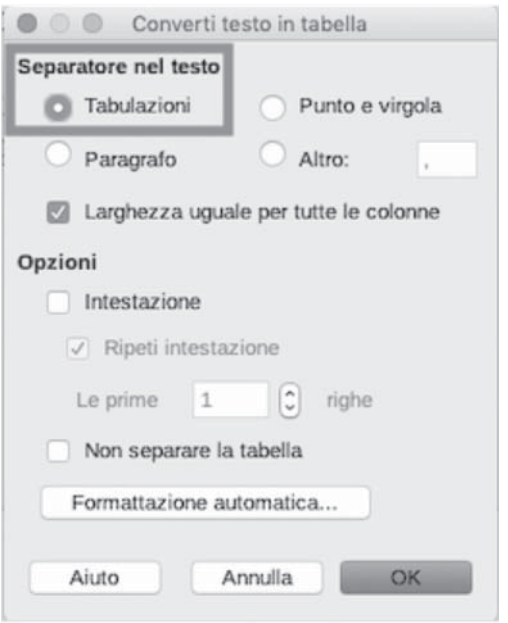


Fig. 26. *Tabulazione come separatore di testo.*

(1) Ken: I started workin etta buck thirty en hour
(0.4)
Ken: en'e sid that if I work fer a month: yih getta buck,h ·h thi[rty ↓fi:ve=
(Dan): [((sniff))
Ken: ='n hour en (·) ev'ry month he uh ( ) he rai[ses you]˚( )˚]
Dan: [How'dju]getth]e jo:b

La modalità di esportazione descritta fin qui si limita a ordinare le annotazioni in ordine cronologico. Si perdono però le potenzialità offerte dalla struttura "a partitura" delle trascrizioni di ELAN, fra cui in primo luogo la possibilità di inserire più annotazioni parallele relative al medesimo frammento di traccia audio (per esempio una trascrizione ortografica e una fonetica), oppure la rappresentazione grafica delle sovrapposizioni di turno, le quali in un normale testo organizzato in verticale sono ricostruibili solo in parte e a costo di un notevole sforzo, come ben si vede nell'esempio (1).

Esiste però una modalità di esportazione alternativa che permette di ricreare almeno in parte la struttura originale della trascrizione, traslandola per così dire dal piano orizzontale a quello verticale. In altre parole, laddove nel documento di ELAN avevamo diverse righe parallele, nel documento di testo avremo altrettante colonne parallele. Per fare ciò, dalla barra dei menu di ELAN scegliamo "*File > Export As > Tab-delimited Text*". Nella schermata che si apre selezioniamo le righe di trascrizione che ci interessa esportare (qui "Alessandra" e "Paola") e attiviamo le opzioni "*Separate column for each tier*","*Include time column for > Begin Time*" e "*Include time format > hh:mm:ss:ms*" (Figura 27).

Di nuovo facciamo clic su "OK" e salviamo il documento nella posizione a noi più comoda, utilizzando l'estensione ".txt" e la codifica "UTF-8". Una volta aperto utilizzando l'editor di testo Atom, il contenuto può essere esportato in un programma di videoscrittura (come mostrato in precedenza) oppure in un foglio di calcolo come Libreoffice Calc, al fine di svolgere ulteriori operazioni di elaborazione e analisi sulle quali non è possibile soffermarsi in questo manualetto (Figura 28).

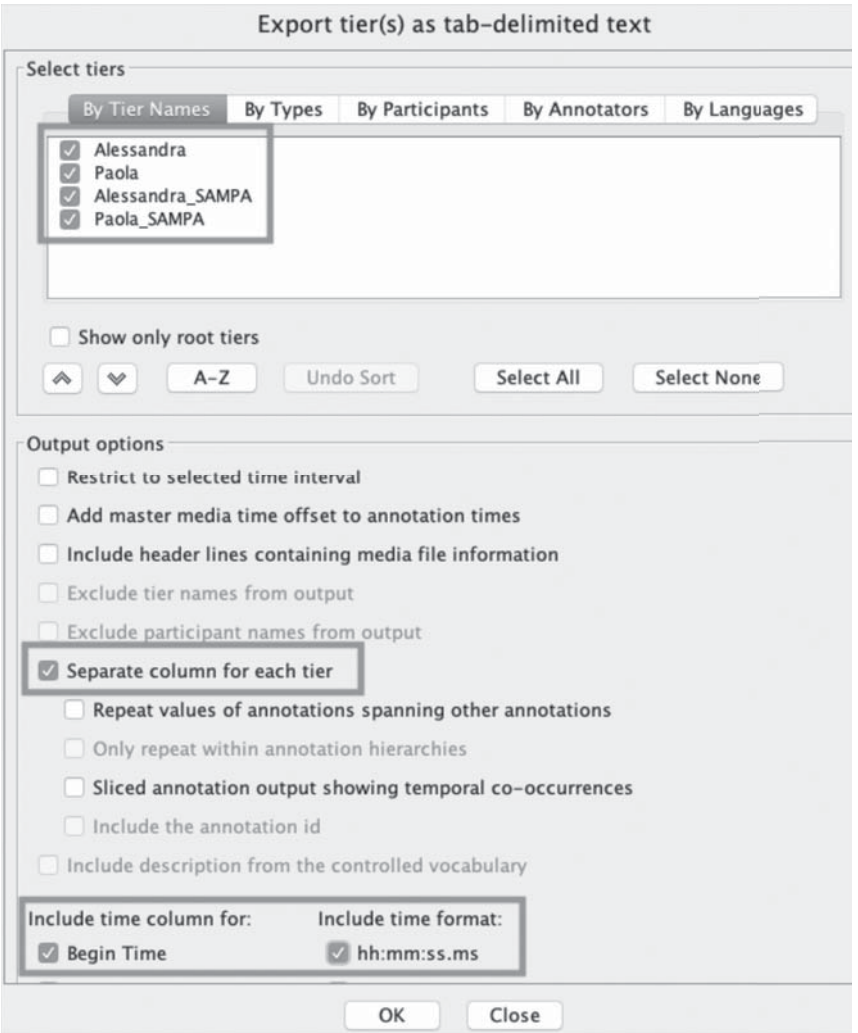


Fig. 27. *Esportazione in formato "Tab-delimited Text".*

A	B	C	D	E
Begin Time - msec	Alessandra	Paola	Alessandra_SAMPA	Paola_SAMPA
62126		cosa guarda alla televisione il tuo?		'koza 'gwarda 'al:a televi'zjone il 'tuo?
63827	il mio scusa il mio ha un mento	invece che va all'ingiù.	il 'mio 'skuza il 'mio a un 'mento	'iFvetSe ke va al:in'dZu
67420		va all'in giù il tuo mento?		va al:in'dZu il 'tuo 'mento?
68878		ah il mio va all'insù.		a il 'mio va al:in'su
69633		cioè va dritto quindi facciamo una x sul nostro mento.		'tSo'e va 'drit:o kwindi fa't:Samo 'una iks sul 'nostro 'mento

Fig. 28. *Resa verticale della struttura "a partitura".*

2. *Cosa trascrivere*

La decisione di quale strumento utilizzare e quale contenuto immettere nelle annotazioni dipende in primo luogo dal tipo di dati e dall'obiettivo dello studio. Nel caso di una ricerca testuale su dati monologici di parlanti nativi, una semplice operazione di sbobinatura potrebbe rivelarsi sufficiente. Qualora invece sia rilevante il livello fonetico, come nel caso di uno studio sulla prosodia, oppure ancora ci si interessi di interazione o di acquisizione linguistica, può essere auspicabile produrre una trascrizione allineata alla traccia audio, così da poter facilmente risalire ai dati originali per aiutarsi nella sua interpretazione. Nel caso della trascrizione di varietà di apprendimento, inoltre, la natura fluida e la struttura parzialmente autonoma del sistema dell'interlingua impongono di presentare il dato fonetico grezzo, in modo da non condizionare il lettore con premature interpretazioni (si veda il contributo di Benazzo e Watorek in questo volume). Consideriamo a questo proposito il seguente esempio polacco (adattato da Saturno 2015: 125).

(2) [ˈprɔʒeʧ ˈjexaʧ ˈprosto ˈɔbok bibljoˈteka i ˈʃkɔwə].
proszę jechać prosto obok biblioteka i szkoł<?>.
'vada dritto vicino alla biblioteca e alla scuola'.

Il polacco è una lingua dotata di una ricca morfologia nominale, così che numerose classi di parola, tra cui i nomi come *biblioteka* 'biblioteca' e *szkoła* 'scuola', possono comparire in varie forme a seconda del contesto sintattico. Nell'esempio (2) la preposizione *obok* 'vicino' richiede il caso genitivo, rispettivamente *bibliotek-i* e *szkoł-y.* Nel caso di *biblioteka* è facile stabilire che la parola compare erroneamente nella forma del caso nominativo in *-a*; nel caso di *szkoła,* invece, l'apprendente produce una vocale centrale /ə/, la cui interpretazione non è altrettanto immediata: potrebbe trattarsi di un'approssimazione della forma del nominativo (errore), di quella del genitivo (forma bersaglio), o ancora di una strategia di evitamento, per cui l'apprendente, conoscendo le propria difficoltà con la morfologia della lingua bersaglio, produce intenzionalmente una forma ambigua. Alcune lingue, infine, come il tedesco, prevedono una regola fonologica per cui le vocali in posizione finale di parola tendono a centralizzarsi, così che l'esempio potrebbe rappresentare anche un caso di interferenza fonologica dalla lingua madre dell'apprendente. Tutte queste possibilità devono essere deducibili dalla trascrizione.

Esistono diverse strategie per rappresentare graficamente eventuali pronunce devianti dalla forma richiesta. La più immediata prevede l'uso dell'ortografia della lingua bersaglio, es. **piedi sporche* per *piedi sporchi*. Un sistema ortografico come questo presenta però diversi inconvenienti. In primo luogo, è evidente che si presuppone la conoscenze delle norme ortografiche della lingua bersaglio. In secondo luogo, non è detto che sia possibile trascrivere combinazioni di suoni estranee alla specifica lingua per cui il sistema ortografico è progettato: come rendere la pronuncia di una parola come [ʌʧˈki] (russo 'occhiali') utilizzando le convenzioni ortografiche dell'italiano? Dal momento che la fonotassi di quest'ultimo non conosce il nesso /ʧk/, logicamente non ne prevede nemmeno una resa ortografica. Infine, in diverse lingue non esistono corrispondenze univoche tra ortografia e pronuncia, circostanza che rende tale approccio alla trascrizione fonetica poco pratico. Così, la prima parte dell'esempio inglese in (3) è più o meno interpretabile come *with, uh, one boat*, mentre la seconda è incomprensibile persino a specialisti madrelingua (Preston 1985: 329).

(3) *Wih A one Boat yuh: : : uhlon dohlenko -*

Il principio fondamentale della trascrizione fonetica è infatti la corrispondenza univoca tra suono e simbolo grafico (grafema), per cui a) ad ogni suono corrisponde uno e un solo grafema e b) a ogni grafema corrisponde uno e un solo suono.

Nell'ortografia delle lingue storico-naturali tale condizione non è di norma rispettata in maniera sistematica. È emblematico a questo proposito il caso dell'inglese: in questa lingua, ad esempio, il medesimo grafema <o> può essere pronunciato /wa/, come in *one* /wʌn/ 'uno', /ɪ/ come in *women* /wɪmən/ 'donne', /əʊ/ come in *no* /nəʊ/ 'no' etc. Allo stesso modo, il suono /k/ può essere rappresentato graficamente con <c>, es. *cat* /kæt/ 'gatto', <ch> es. *character* /kæraktə/ 'personaggio', <k> es. *key* /ki:/ 'chiave' etc. Dalla rappresentazione ortografica di una parola inglese non è dunque possibile ricostruirne la pronuncia, e viceversa.

Quest'ultima affermazione non si applica a una lingua come l'italiano, il cui sistema ortografico tuttavia non può essere considerato un alfabeto fonetico. Se infatti è vero che tra pronuncia e ortografia è possibile stabilire un legame non arbitrario, non si verifica comunque la condizione per cui a un grafema corrisponde uno e un solo suono e viceversa. Per

convincersene è sufficiente considerare il caso della lettera <c>, la quale a seconda del contesto grafico contribuisce a rappresentare graficamente tre suoni diversi: /ʧ/ quando è seguita dalle lettere <i, e> e non è preceduta da <s>, es. *ciao*, *cena*, /k/ quando è seguita dalle lettere <a, o, u, h>, es. *cane*, *anche*, e infine /ʃ/ quando è preceduta da <s> e seguita dalle lettere <e, i>, es. *scena*, *scimmia*.

Forse il più diffuso tra gli alfabeti fonetici è l'alfabeto fonetico internazionale (IPA, ingl. *International Phonetic Alphabet*, Landau et al. 1999), che rappresenta uno degli strumenti indispensabili nel bagaglio metodologico di qualunque linguista in quanto permette di eseguire la trascrizione fonetica di qualsiasi lingua. Laddove possibile, l'IPA utilizza uno specifico carattere Unicode per rappresentare un dato suono linguistico: l'elevato numero di questi ultimi fa sì che l'alfabeto comprenda un discreto numero di simboli indipendenti, oltre ad alcuni diacritici (caratteri modificatori di altri caratteri). Per questo motivo, l'immissione di caratteri IPA sul computer si rivela spesso complicata. Tra gli strumenti più utilizzati si può ricordare le tastiera a schermo, la quale può essere incorporata nel sistema operativo (es. Figura 28, "Visore caratteri" di Mac OS) oppure online (es. Figura 29). Si tratta di un sistema utile per inserire un numero ridotto di simboli in particolari contesti, come gli esempi di un articolo scientifico, ma poco pratico per trascrivere grandi quantità di testo. Oltre all'IPA, le tastiere virtuali permettono normalmente di inserire una grande varietà di simboli, quali caratteri con diacritici, alfabeti diversi da quello latino etc.

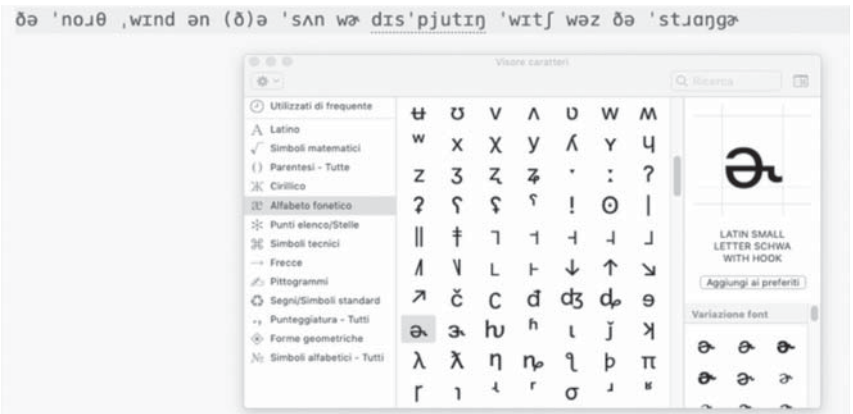


Fig. 29. *"Visore caratteri" di Mac OS.*

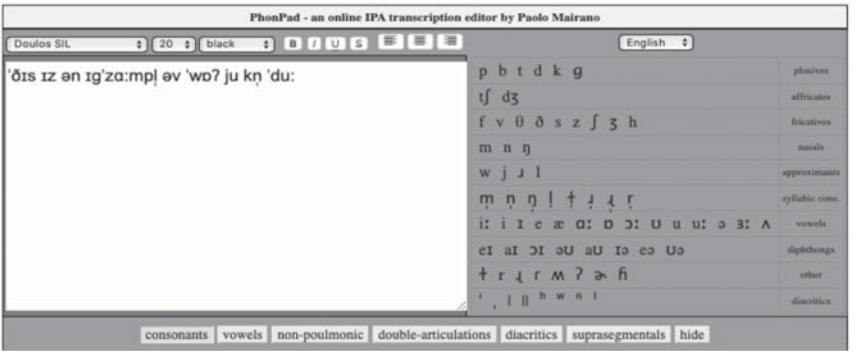


Fig. 30. *Tastiera online (Phonpad*[7] *di Paolo Mairano).*

Un'alternativa alla tastiera virtuale è l'installazione di uno specifico layout di tastiera. Si tratta di un software che associa ciascun simbolo IPA a una combinazione di tasti: per esempio, per inserire il simbolo /æ/, il layout "IPA Unicode 6.2 (ver. 1.4) MSK" di SIL[8] richiede di digitare la combinazione "<a". Si tratta di un sistema utile per inserire consistenti quantità di testo direttamente in IPA. Fra gli svantaggi si segnalano una certa macchinosità e la necessità di memorizzare diverse combinazioni di tasti (Figura 30).

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p p b b			t t d d		ʈ <t ɖ <d	c c ɟ =j	k k g <g	q q ɢ =G		ʔ =?
Nasal	m m	ɱ >m		n n		ɳ <n	ɲ =n	ŋ >n	ɴ =N		
Trill	ʙ =B			r r					ʀ =R		
Tap or Flap		ⱱ <v		ɾ >r		ɽ <r					
Fricative	ɸ =f β =b	f f v v	θ =t ð =d	s s z z	ʃ =s ʒ =z	ʂ <s ʐ <z	ç =c ʝ <j	x x ɣ =g	χ =x ʁ >R	ħ >h ʕ <?	h h ɦ <h
Lateral Fricative				ɬ =l ɮ >l							
Approximant		ʋ =v		ɹ =r		ɻ <R	j j	ɰ >w			
Lateral Approximant				l l		ɭ <l	ʎ <L	ʟ =L			

Fig. 31. *Combinazioni da tastiera, layout "IPA Unicode 6.2 (ver. 1.4) MSK" di SIL, consonanti.*

[7] es. http://phonetictools.altervista.org/phonpad/
https://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=uniipakeyboard

A questo proposito, va specificato che le combinazioni specificate nei manuali si riferiscono alla tastiera americana, la quale differisce da quella italiana nella collocazione di numerosi simboli, così che i caratteri stampati sulle tastiere italiane si rivelano spesso fuorvianti (Figura 31).

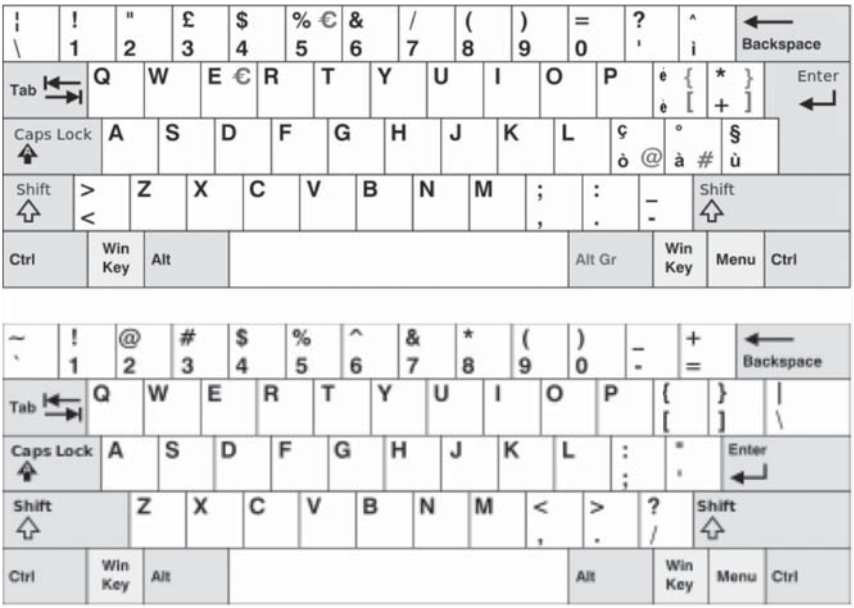


Fig. 32. Layout di tastiera italiana (in alto) e americana (in basso).

Alla luce di tali limitazioni, ai fini della trascrizione manuale è utile introdurre anche l'alfabeto SAMPA (Wells 1997) e la sua estensione X-SAMPA (Wells 1995), nati proprio per la trasmissione telematica dell'alfabeto fonetico in tempi precedenti alla diffusione dello standard Unicode. SAMPA utilizza i soli caratteri ASCII, tutti riportati direttamente sulla tastiera del computer: in particolare, esso comprende le 21 lettere minuscole e maiuscole dell'alfabeto inglese, i numeri da 0 a 9, alcuni segni di punteggiatura e diversi altri simboli (es. <£>, </> etc.). Ai fini dell'uso di tali caratteri come alfabeto fonetico, ciascuno (comprese le varianti minuscola e maiuscola della medesima lettera) è associato in maniera univoca a un suono, es <s> /s/, <S> /ʃ/; in più, se necessario, le lettere dell'alfabeto possono essere modificate da altri sim-

boli per rappresentare ulteriori suoni, es <s\> /ɕ/ (Figura 32). Quest'ultimo esempio permette di evidenziare il fatto che un grafema può essere composto anche di più simboli e ciononostante corrispondere ad un singolo suono (come d'altra parte avviene normalmente nei sistemi ortografici delle lingue storico-naturali, es. <gn> /ɲ/ in italiano).

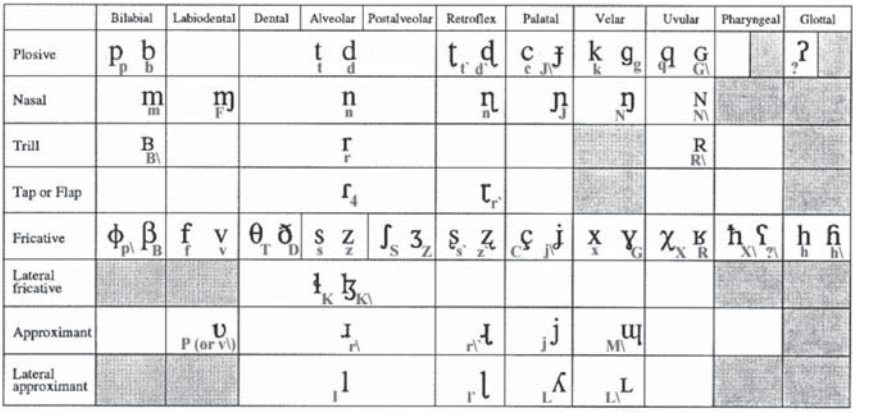


Fig. 33. *Combinazioni di tastiera, X-SAMPA, consonanti*[9].

Il punto (4) riporta la trascrizione della medesima frase inglese (4a) secondo le convenzioni IPA (4b) e SAMPA (4c).

(4) a. *The North Wind and the sun were disputing which was the stronger.*
b. [ðə nɔːθwɪnd ən ðə sʌn wə dɪsˈpjuːtɪŋ wɪtʃ wəs ðə ˈstɹɑŋgə]
c. D@ nO:TwInd @n D@ sVn w@ dIspju:tIN wItS w@s D@ strANg@

Rispetto all'IPA, tra i vantaggi del SAMPA c'è senza dubbio la rapidità di digitazione; non è inoltre necessario memorizzare le numerose combinazioni di tasti corrispondenti ai singoli grafemi. Tra gli svantaggi si segnala invece la minore leggibilità. Un utile compromesso perciò è di utilizzare il SAMPA durante la trascrizione manuale, per poi convertire il testo in IPA utilizzando lo strumento "trova e sostituisci" automatico. Dal momento che tanto IPA, quanto SAMPA presentano una rela-

[9] http://www.let.rug.nl/~kleiweg/L04/Tutorial/xsamchart.gif

zione univoca tra suoni e caratteri, l'operazione è puramente meccanica e non presenta alcuna difficoltà. L'unica cautela da osservare è di convertire prima i grafemi composti da più di un carattere (es. <s\>) e successivamente quelli composti da uno solo (es. <S>). Tale approccio permette anche al trascrittore di elaborare un alfabeto fonetico personalizzato per ragioni di comodità, a condizione che sia rispettata la relazione di univocità tra suoni e grafemi.

La frequenza del ricorso alla trascrizione fonetica dipende dal tipo di dati e dagli obiettivi della trascrizione. Se lo studio non si concentra specificamente sul livello fonetico e i dati non presentano un eccessivo scostamento dalla varietà bersaglio, il trascrittore potrebbe scegliere di privilegiare la leggibilità della trascrizione utilizzando l'ortografia standard laddove possibile, limitandosi a inserire la trascrizione fonetica di elementi particolarmente devianti oppure ambigui (es. 5, francese L2; Benazzo & Watorek, questo volume).

(5) *après je* [le fe] *un petit peu de ménage*

Nel caso invece di forti e sistematici scostamenti dal bersaglio atteso può essere consigliabile trascrivere l'intero testo in alfabeto fonetico (es. 6, polacco L2; Bernini 2018: 96).

(6) [ˈpɔm ʤeˈlone skɛnˈtaʧˈtɛʃ (...) skranˈʧaʧ]
'anche il signor Verdi salta'

È poi possibile utilizzare diverse convenzioni di trascrizioni, dedicate a specifici livelli di analisi, come gli atti linguistici o la prosodia. A questo proposito si consiglia di consultare la sezione del manuale di ELAN relativa ai "tipi linguistici" (ingl. "*type*") e al vocabolario controllato (ingl. "*controlled vocabulary*").

Bibliografia

Bernini, Giuliano. 2018. The sound pattern of initial learner varieties. *Linguistica e Filologia* 38. 85–110.

Brugman, Hennie & Russell, Albert. 2004. Annotating Multimedia/Multi-modal resources with ELAN. In Lino, Maria T. & Xavier, Maria F. & Ferreira,

Fátima & Costa, Rute & Silva, Raquel (eds.), *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, 2065–2068. Paris: ELRA.

Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Lerner, Gene H. (ed.), *Conversation analysis. Studies from the first generation*, 13–31. Amsterdam/Philadelfia: John Benjamins.

Landau, Ernestina & Lončarić, Mijo & Horga, Damir & Škarić, Ivo. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Preston, Dennis. 1985. The Li'l Abner Syndrome: Written Representations of Speech. *American Speech* 60(4). 328–336.

Saturno, Jacopo. 2015. La trascrizione di varietà di apprendimento iniziali. In Chini, Marina (a cura di), *Il parlato in [italiano] L2: aspetti pragmatici e prosodici/[Italian] L2 spoken discourse: pragmatic and prosodic aspects*, 117–138. Milano: Franco Angeli.

Wells, John C. 1995. Computer-coding the IPA: a proposed extension of SAMPA. https://www.phon.ucl.ac.uk/home/sampa/ipasam-x.pdf

Wells, John C. 1997. SAMPA computer readable phonetic alphabet. In Gibbon, Dafydd & Moore, Roger K. & Winski, Richard (eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Berlin/New York: Walter de Gruyter.

Questa pubblicazione è stata realizzata utilizzando carta fabbricata nel pieno rispetto dell'ambiente senza l'utilizzo di sostanze nocive e con l'impiego di prodotti ecocompatibili nella fase di stampa e confezione.

Finito di stampare
nel mese di luglio 2021
sestanteinc - Bergamo