



Classificazione Decimale Dewey:

410.1 (23.) LINGUISTICA. Filosofia e teoria

FRANCESCO CUTUGNO

INTENDI?

COME GLI UMANI RICONOSCONO IL PARLATO NATURALE

Prefazione di

FEDERICO ALBANO LEONI





ISBN
979-12-218-2268-7

PRIMA EDIZIONE
ROMA 31 OTTOBRE 2025

al mio splendore

Mi intendi? ¿*Entiendes?* in spagnolo, *M'he ntisə?* In dialetto napoletano...

Intendere v. tr. [dal lat. intendere, comp. di in- e -tendere «tendere, rivolgere, mirare a»]. – Verbo di largo uso e di molteplici significati, che si possono ricondurre fondamentale-mente a tre: a) *volgere i sensi o la mente a qualche oggetto*; b) tendere a un fine, e quindi avere intenzione, volere; c) *percepire e comprendere, che sono in certo modo l'effetto del tendere l'attenzione*. [corsivo mio]

Vocabolario Treccani online, Istituto dell'Enciclopedia Italiana, www.treccani.it.

INDICE

11	<i>Prefazione. L'altra metà del cielo</i> di FEDERICO ALBANO LEONI
19	<i>Introduzione</i>
25	Capitolo I Fisiologia uditiva, psicofisica e psicoacustica 1.1. Premessa, 25 – 1.2. Vie neurali della funzione uditiva, aree corticali, 27 – 1.3. Psicofisica e psicoacustica, 30 – 1.3.1. <i>I metodi di indagine della psicofisica</i> , 31 – 1.3.2. <i>Psicoacustica</i> , 34 – 1.3.3. <i>Loudness, pitch e durate dei toni puri</i> , 36 – 1.3.4. <i>Psicoacustica dei suoni complessi</i> , 37 – 1.3.5. <i>Memoria e attenzione</i> , 42 – 1.3.6. <i>La Gestalt. L'udito e il linguaggio</i> , 44 – 1.4. Conclusioni, 46.
47	Capitolo II Elementi di fonetica percettiva 2.1. Premessa, 47 – 2.2. Metodologie della fonetica percettiva, 48 – 2.3. Percezione categorica e percezione continua, 52 – 2.4. Rapporti fra produzione fonica e percezione fonetica, 59 – 2.4.1. <i>Percezione dei tratti segmentali</i> , 59 – 2.4.2. <i>Percezione dei tratti soprasegmentali</i> , 64.
69	Capitolo III I modelli della percezione del parlato 3.1. Premessa, 69 – 3.2. Le dimensioni del problema, 70 – 3.2.1. <i>Dimensioni temporali: l'unità minima di analisi</i> , 70 – 3.2.2. <i>Invarianza e variabilità</i> , 70 – 3.2.3. <i>Top-down o bottom-up?</i> , 71 – 3.2.4. <i>Modelli attivi o modelli passivi</i> , 73 – 3.2.5. <i>Il</i>

ruolo del contesto, 73 – 3.3. I modelli, 74 – 3.3.1. *La teoria acustica di Fant* (1960, 1973), 75 – 3.3.2. *La Motor Theory di Liberman con i suoi successivi sviluppi*, 77 – 3.3.3. *Analysis by synthesis di Halle & Stevens* (1962), 79 – 3.3.4. *La Quantal Theory of Speech di Stevens*, 81 – 3.3.5. *Hyper- & hypo- speech di B. Lindblom* (1990), 83 – 3.3.6. *La Fuzzy theory di Massaro*, 86 – 3.3.7. *Polysp di Hawkins & Smith*, 88 – 3.3.8. *Dei suoni e dei sensi di Albano Leoni* (2009), 90 – 3.3.9. *Abstracts and exemplars*, 93 – 3.4. Tempo, ritmo memoria e attenzione, 96 – 3.4.1. *Il ritmo del parlato*, 100 – 3.4.2. *Psicolinguistica e ritmo*, 107 – 3.4.3. *Per una visione asincrona dei processi temporali*, 111 – 3.5. Word recognition e lexical access, 113 – 3.5.1. *Il modello delle Coorti - Cohort Model*, 113 – 3.5.2. *TRACE*, 115 – 3.5.3. *Shortlist and Shortlist B*, 116 – 3.5.4. *Fine-Tracker*, 118 – 3.5.5. *Earshot, LDL-Auris, DIANA*, 119 – 3.6. Tempo e marcatori acustici, 120 – 3.6.1. *Il tempo nel lexical access*, 121 – 3.6.2. *Processi cognitivi per la segmentazione del parlato continuo*, 123.

125 Capitolo iv

Riconoscere il parlato

4.1. Premessa, 125 – 4.2. Processi metalinguistici, ruolo della memoria, 126 – 4.2.1. *L'ossessione per la ricerca dell'invarianza*, 126 – 4.2.2. *Fonetica e metafonetica percettiva*, 128 – 4.2.3. *Psicolinguistica, priming e monitoring*, 131 – 4.3. Il parlato naturale e i processi di riconoscimento, 134 – 4.3.1. *Com'è fatto il parlato naturale?*, 134 – 4.3.2. *Moduli, pensieri lenti e veloci, memoria e attenzione*, 135.

139 Capitolo v

Le tecnologie per il riconoscimento del parlato e i modelli di percezione

5.1. Premessa, 139 – 5.2. L'approccio classico basato su modelli markoviani e Language Models, 140 – 5.3. Le reti neurali come modello “nascosto” dei processi percettivi, 143 – 5.3.1. *I primi sistemi DNN-based*, 144 – 5.3.2. *I sistemi totalmente End-to-end (E2E)*, 145 – 5.4. I modelli matematici della acquisizione del linguaggio nei bambini, 148 – 5.5. Multimodalità e acquisizione del linguaggio negli umani e nei sistemi IA, 152.

155 Conclusioni. La complessità del parlato come oggetto di indagine

161 Ringraziamenti

163 Bibliografia

PREFAZIONE

L'ALTRA METÀ DEL CIELO

Credo che ci siano pochi dubbi sul fatto che l'agire linguistico degli umani abbia come primo e ultimo fine quello di dire qualche cosa a qualcuno e di capire quello che qualcuno ci ha detto o, per dirla con l'Autore di questo libro, di intenderlo. Questo qualche cosa può essere l'espressione di un desiderio o di uno stato d'animo, un ordine, una domanda, un'informazione, una descrizione di un oggetto o di una situazione, una dichiarazione d'amore, o qualsiasi altra cosa ci possa venire in mente. Questo "qualche cosa" può essere scritto o parlato o segnato o disegnato o fischiato, ma è sempre un "qualche cosa" rivolto a qualcuno che sta lì davanti a noi o anche a qualcuno lontano, come in una lettera, o a qualcuno ignoto come il lettore di un romanzo. Insomma, il monologo è solo una finzione letteraria o teatrale e parlare è sempre un dialogare in cui a turni alterni si cerca di dire e si cerca di capire, come si legge nei testi di due giganti del pensiero linguistico e filosofico: Humboldt e Benveniste.

Questo che sto dicendo è una banalità, che però ci è confermata dalla nostra esperienza quotidiana, e anche dal buon senso: se nessuno parla non c'è niente (di linguistico) da ascoltare e se nessuno ascolta è inutile parlare. Gli psicoanalisti ci insegnano che perfino l'endofasia, cioè il parlare tra sé e sé, il parlato interiore è in realtà il dialogo tra le due parti di un io sdoppiato. Dunque, si potrebbe dire che tra il parlante e l'ascoltatore c'è un rapporto di presupposizione reciproca: senza ascoltatore non c'è parlante e senza parlante non c'è ascoltatore e un'analisi che tenga conto esclusivamente di uno dei due attori è fatalmente monca.

Ma forse si potrebbe aggiungere anche un'altra cosa, e cioè che il rapporto tra i due è leggermente squilibrato, a vantaggio dell'ascoltatore. Chi è infatti che garantisce che quello che ho detto è qualche cosa di sensato? Chi è che garantisce al bambino che fa i suoi primi esperimenti che il suono olofrastico che ha emesso, magari insieme a un gesto, è efficace? Il garante è l'ascoltatore che capisce, è l'adulto che soddisfa il bisogno che il bambino cerca di esprimere; sono loro che ci confermano la sensatezza del nostro parlare e che dunque ci inducono a ripeterlo e a stabilizzarlo nella memoria. Né si può dimenticare che il primo ascoltatore di me stesso sono io che sento quello che dico, sia attraverso le orecchie, sia attraverso la trasmissione ossea delle vibrazioni, e posso così controllarmi e eventualmente correggermi, operazione fondamentale nell'apprendimento della lingua. Infine, si sa che i cosiddetti sordomuti sono in realtà sordi e non parlano perché è mancato loro lo stimolo da imitare.

Cutugno ci ricorda dunque che il primato dell'ascolto è anche nella nostra filogenesi e ontogenesi: l'apparato uditivo dell'*homo sapiens* è un apparato primordiale, arcaico, che nel corso della nostra storia anatomica documentata non ha subito modifiche rilevanti, che ha come fine primario l'udire i rumori del mondo per sopravvivere. Inoltre, è un apparato che non si chiude mai e che non ha bisogno di addestramento: noi sentiamo e basta, e sentiamo perfettamente già prima della nascita.

Viceversa, gli organi con i quali parliamo, il cosiddetto apparato fonatorio, hanno come funzioni primarie, vitali, quella della respirazione e della nutrizione e si sono poi progressivamente adattati a produrre i suoni attraverso un lungo processo evolutivo che ha modificato la nostra anatomia e per farli funzionare è necessario un lungo addestramento. Questo lunghissimo processo è stato guidato dall'orecchio; noi abbiamo sviluppato la capacità di produrre suoni selezionando quelli che il nostro orecchio percepisce meglio. In teoria, il nostro apparato fonatorio avrebbe anche potuto svilupparsi in modo da produrre ultrasuoni, come è nei pipistrelli, ma ciò non è avvenuto perché il nostro orecchio non li avrebbe percepiti.

Dunque, osservando le cose per come sono, si potrebbe dire che l'ascoltatore (che, non si dimentichi, deve anche capire, cosa ben più complessa del semplice percepire) certamente occupa la metà del cielo

dell'agire linguistico e dunque merita di essere studiato come il parlante. Se poi si pensa, come è giusto che sia, che forse ne occupa un po' più della metà, i motivi per studiarlo diventano più pressanti.

Tuttavia, se passiamo a guardare il mondo e gli strumenti istituzionali della linguistica, per esempio i manuali che sono il luogo in cui i risultati della ricerca diventano sapere condiviso e dunque concorrono a formare il senso comune di chi li legge, vediamo un'anomalia. Le istituzioni linguistiche si dedicano soprattutto a studiare il testo, cioè il prodotto oggettivato e concretamente percepibile dell'agire linguistico, ad analizzarlo, segmentarlo, classificarlo; in qualche caso si dedicano anche a studiare il produttore, per esempio in fonetica articolatoria, ma mai, o molto raramente l'ascoltatore. Dunque, da questo punto di vista la nostra tradizione di studi linguistici è sbilanciata.

Credo che il motivo dello sbilanciamento sia semplice e risiede nella diversa natura e diversa fenomenologia dei processi del parlare e i processi del sentire/capire. Io posso osservare chi parla, posso vedere alcuni dei suoi movimenti, posso percepire me stesso che parlo e percepire alcuni dei miei movimenti, posso udire il risultato di quell'agire e posso metterlo per iscritto e da qualche tempo anche registrarlo e studiarlo. L'udito è invece un processo tutto interiore, sfuggente, invisibile dall'esterno e per essere studiato richiede strumenti che non fanno parte del laboratorio del linguista. La matrice dello squilibrio è in questa differenza: studiare ciò che posso osservare direttamente è più facile che studiare ciò che non posso osservare semplicemente, come del resto mostra abbondantemente questo libro. Per non dire poi del pregiudizio a lungo vigente secondo il quale i processi della produzione e quelli della comprensione sarebbero perfettamente simmetrici e che dunque per capire basterebbe compiere a ritroso e passivamente il percorso che ha seguito chi parla.

Ora, le difficoltà che presenta lo studio dell'udire e del comprendere sono fuori dubbio, ma non possono essere la giustificazione di un silenzio.

Naturalmente, guardando al mondo della ricerca, si vede che l'altra metà del cielo non è affatto abbandonata o negletta: ci sono fior di studiosi che la osservano con attenzione: audiologi, fisici acustici e psicoacustici, psicologi, neuroscienziati e ora anche gli informatici e gli

studiosi di intelligenza artificiale (perché i sistemi intelligenti devono parlare e soprattutto devono capire). Ma ognuno di queste specialità studia una porzione del complesso processo che parte dalla forma d'onda del segnale, attraversa i meccanismi della trasformazione dello stimolo in impulsi nervosi, e arriva al costituirsi di una rappresentazione mentale e infine alla comprensione e non è detto che ognuna di queste specialità conosca quello che fanno le altre. Quindi il problema, pratico ma anche teorico, è quello di una sintesi che aiuti tutti gli studiosi a capire meglio quello che fanno essi stessi e soprattutto quello che fanno gli altri, e a mettere a frutto nel proprio dominio i preziosi saperi degli altri

Il libro di Francesco Cutugno è appunto un contributo in questa direzione, incentrato, come dice il titolo, sul problema complesso e anche un po' misterioso della comprensione, senza la quale il nostro parlare sarebbe inutile e le nostre tassonomie sarebbero giochi autoreferenziali. In questo l'Autore è aiutato dalla sua formazione: in partenza fisico, poi studioso dei processi uditivi e della modellizzazione del rapporto fra evento acustico e rappresentazione uditiva e infine linguista, ponte fra saperi anche molto diversi e che spesso hanno difficoltà a entrare in contatto. Cutugno è infatti studioso del parlato, della fonetica acustica e uditiva; è un critico della semplificazione che le fonologie operano sulla complessità dei processi della produzione e comprensione dei suoni nel parlato naturale e quest'ultimo punto è ormai come un attrattore della sua ricerca.

In questo libro Cutugno mette a frutto le sue competenze tecniche e, cosa non frequente, le indirizza appunto al problema della comprensione, cioè a un problema che è, o dovrebbe essere, centrale nella ricerca sul funzionamento delle lingue.

Credo quindi di poter dire che le poche cose che ho scritto in questa prima parte della mia prefazione sono un condensato di quello che si trova nel libro. Infatti Cutugno, per ciascuno dei domini disciplinari che ho ricordato, non solo passa in rassegna il complicato stato dell'arte, ma non dimentica mai le relazioni tra le componenti.

Il libro è costruito a partire dalla convinzione che ciò che mette in moto il processo è un segnale acustico che viene percepito. Di conseguenza sono descritti con accuratezza i meccanismi psicofisici della

percezione del segnale, della discriminazione tra segnali e e della individuazione delle sue proprietà, oggettive e soggettive, sempre determinate dalla nostra anatomia e dalla nostra neurologia.

Naturalmente l'Autore è molto attento ai metodi e non si può non apprezzare la sua prudenza, a volte quasi la diffidenza, nel valutare l'efficacia e la capacità esplicativa dei test, strumento principe della ricerca in psicoacustica, tanto più quando il materiale sperimentale è prodotto in laboratorio e non estratto da parlato naturale. Altrettanto notevole è la sua attenzione per il ritmo e l'intonazione e per la centralità percettiva della sillaba.

È infine da osservare che Cutugno non elude una delle *cruces* dell'analisi linguistica e cioè quella della indeterminatezza delle lingue, della loro intrinseca ambiguità, che attraversa tutti i livelli, fonologico, morfo-sintattico, semantico-lessicale e che pone, anche alla fonologia problemi che nessuno dei modelli correnti riesce a domare. Infatti Cutugno, raccogliendo e mettendo a frutto indirettamente la lezione di Zipf e di Bühler e direttamente quella di Lindblom, abbandona decisamente l'ottica, fino a poco fa dominante, del logocentrismo e riconosce che la cognizione linguistica, parte importante della capacità cognitiva generale, si costruisce non a partire da un testo in sé, o da un enunciato ben formato, ma a partire dall'intreccio tra testo e contesto, tra enunciato e esperienza sensoriale e mentale dl mondo: è da qui che parte la costruzione e la comprensione del senso, come ormai molti, me compreso, ritengono.

Un aspetto veramente rilevante di questo libro è, come si vedrà leggendolo, lo spazio e l'attenzione dedicati alla memoria, proprietà di tutti i viventi, ancora un po' misteriosa dal punto di vista neurologico, che si manifesta in modo eccezionale e pervasivo negli umani nei quali, insieme con la capacità di attenzione, è condizione necessaria per ogni forma di cognizione, di conoscenza e per ogni forma di vita individuale e sociale.

Le pagine di Cutugno a questo proposito sono quanto mai opportune. Vorrei ricordare che gli antichi sapevano benissimo quanto la memoria sia importante: Mnemosyne era una dea, madre delle Muse; i trattati di oratoria latini, dalla anonima *Rhetorica ad Herennium* a Cicerone e Quintiliano considerano la memoria una delle componenti

basilari della orazione; in Europa sono stati scritti trattati sulla memoria di grande fascino, come quello di Giordano Bruno, e innumerevoli lavori di mnemotecnica.

Ma i linguisti sembrano ignorare o aver dimenticato tutto questo e la memoria non è mai presente tra i parametri da considerare per capire come funzionano i parlanti e dunque le lingue. Eppure, non sarebbe difficile vedere, con un piccolo gioco di immaginazione, che senza la memoria tutto il nostro edificio mentale non solo linguistico si affloscerebbe come un castello di carte.

Invece, in controtendenza rispetto ai canoni della descrizione del funzionamento delle lingue, il tema della memoria è largamente presente nel libro e ad esso sono dedicate due sezioni specifiche (1.2.5. *Memoria e attenzione*; 3.3. *Tempo, ritmo memoria e attenzione*) nella seconda delle quali esso si intreccia con una questione cruciale nei modelli di riconoscimento del parlato e cioè se i sistemi di riferimento temporale del parlante e dell'ascoltatore siano sincroni o asincroni. La questione si manifesta nel ritmo del segnale dove si può ritenere, come fanno le fonologie metriche, che la successione delle salienze si disponga in maniera regolare, condivisa da chi parla e da chi ascolta (modello sincrono) o se essa si sviluppi nel tempo in modo irregolare (modello asincrono). In ambedue i casi, naturalmente, è attiva la memoria a breve termine, senza la quale il ritmo non potrebbe essere percepito o riconosciuto perché non sarebbe possibile correlare una salienza alle salienze precedenti. L'autore sostiene la seconda opzione (3.3.3. *Per una visione asincrona dei processi temporali*), in base a un principio, valido anche per il funzionamento neurale del recupero dell'informazione, che un processo regolare, e dunque prevedibile, è meno potente, meno ricco di informazione, e dunque meno efficace, di un processo irregolare.

Nella presentazione di modelli, dati, teorie Cutugno non è mai perentorio e un ragionevole dubbio accompagna le sue analisi, e il libro si conclude con l'individuazione di due problemi aperti e non da poco.

Il primo consiste nel fatto che, assodato che la comprensione linguistica è un processo complesso che risulta da meccanismi e conoscenze anche molto diversi, rimane da chiedersi dove risieda e come sia fatto il centro mentale nel quale avviene il coordinamento di questi meccanismi.

Il secondo è la domanda su come questi meccanismi e conoscenze locali si correlino alla conoscenza enciclopedica del mondo, senza la quale essi sarebbero inefficaci. Infatti, è essa che rende possibili i processi inferenziali, cioè il riferimento a ciò che non è detto ma è probabile, ricordando che questi processi sono di importanza cruciale per muoversi in un mondo linguistico e cognitivo in parte caotico e al cui interno c'è ben poco di discreto e ci sono ben pochi rapporti biunivoci.

Dunque, il libro di Cutugno è un invito a continuare a studiare, perché c'è molto da fare per tutti.

Federico Albano Leoni

INTRODUZIONE

Questo libro approfondirà i significati a) e c) della definizione del verbo “*intendere*” indicate in corsivo in epigrafe. “Intendere” un enunciato pronunciato a voce vuol dire attivare una catena di processi sensoriali, cognitivi e semantici. Il segnale vocale, oltre al messaggio in sé, veicola significati pragmatici, paralinguistici e sociali. A partire dalla elaborazione uditiva, la nostra mente, facendo ricorso a tutte le facoltà cognitive, non solo quelle linguistiche, ci permette di comprendere il senso del messaggio, e il messaggio recepito modifica il nostro stato mentale. Questo libro, si pone come obiettivo di riassumere, probabilmente in maniera incompleta, gli studi del settore e, al contempo, di fornire nuove chiavi di lettura.

Questo libro si sarebbe dovuto chiamare “Dai suoni ai sensi”, ma questo titolo e alcune sue varianti, sono già state usate in molte circostanze, sarebbe stato poco originale. È stato usato come titolo di un grande progetto europeo nell’ambito delle reti tematiche Marie Curie (*Sound to sense*) e ha caratterizzato una felice stagione di studi in tutta Europa sul tema della percezione e del riconoscimento del parlato; è stato usato (Dei suoni e dei sensi) da Federico Albano Leoni (2009). Questo titolo viene anche usato correntemente da chiunque studi un processo che non si limiti a comprendere come riconosciamo i singoli suoni della lingua o le strutture prosodiche che si articolano sul livello soprasegmentale, ma anche di come, da sequenze di suoni che generano una complessa, ridondante, polisistemica catena di informazione, si

ricostruiscano significati, si comunichi con gli altri, si esprimano emozioni, si confezionino atti linguistici tradotti direttamente in funzioni pragmatiche.

I significati a) e c) della voce di dizionario del verbo intendere indicano proprio questo: volgere i sensi, quello dell'udito in particolare, ma, come vedremo non solo quello, e di conseguenza la mente, a qualcosa che riceviamo; percepire, comprendere e quindi rivolgere l'attenzione al messaggio ricevuto. Questo messaggio è complesso, la parola parlata è il nostro principale mezzo di comunicazione, tutti gli altri sono ancillari rispetto ad essa. La comunicazione parlata, insieme al pollice opponibile, ci pone al vertice della catena evoluzionistica, crea le basi per l'essere sociale. Il problema di trovare un titolo originale per questo mio saggio è condiviso con Tullio De Mauro, che già nel 1994 (*Intelligenti pauca* -De Mauro 1994) faceva notare la asimmetria nel grande numero di parole, in italiano come in molte altre lingue, che esprimono concetti di produzione della parola rispetto a quanto poche siano le parole che significhino invece l'«udire» e in particolare l'udire riferito a parole. Similmente, nella comunità dei linguisti italiani, si deve aspettare il 1990 perché il tema del ricevente occupi sufficiente spazio nel dibattito scientifico, al punto da dedicare al tema un convegno della Società di Linguistica Italiana. In ambito internazionale, al contrario, si potrebbe partire dalle considerazioni storiche dei padri della linguistica, che avevano contro la difficoltà di svolgere esperimenti di ascolto e valutazione soggettiva che richiedevano la disponibilità di attrezzature e metodi che hanno iniziato a svilupparsi in maniera fattiva solo nel dopoguerra. Le idee proposte fino agli anni '40 del secolo scorso erano basate su ipotesi o su pareri empirici su come si potessero spiegare i meccanismi di percezione e riconoscimento del parlato. La possibilità di registrare dei suoni e di riprodurli ripetutamente, con parametri sperimentali che potessero essere tenuti sotto controllo in modo da consentire l'osservabilità dei fenomeni e la ripetibilità dei risultati, risale alla metà del '900 e, pur sapendo di fare un enorme torto al pensiero scientifico precedente, questo libro partirà da studi che possiamo considerare "moderni" con le prime testimonianze che si attestano intorno al 1950.

Le discipline che coprono i temi del riconoscimento e della comprensione del parlato sono molteplici. Da un lato dobbiamo conoscere

bene la forma e le strutture del segnale acustico, della sua organizzazione fonotattica, delle proprietà spettrali dei segmenti e delle componenti soprasegmentali, dall'altra dobbiamo conoscere il sistema uditivo, sia nella sua componente periferica che in quella neurale che dal trasduttore sensoriale veicola informazione al cervello seguendo un percorso articolato che porta vie nervose fino alla corteccia cerebrale. Poi dobbiamo fare un salto funzionale, per giungere alla descrizione delle funzioni cognitive corticali e modulari fino a prendere in considerazione un possibile modello della mente che ascolta, che comprende e che prende decisioni sulla base di quello che ha... inteso...

Questo volume non può e non vuole essere un compendio di tutti questi argomenti, e, sebbene quindi possa apparire incompleto, lascerà al lettore il compito di integrare le competenze necessarie alla comprensione di alcuni passaggi.

I temi trattati in questo volume sono “nuovi” nel panorama italiano, raramente trattati in altre pubblicazioni e, comunque, mai affrontati con la sistematicità che ho cercato di produrre. Si è reso quindi necessario proporre nei primi tre capitoli, una rassegna che aiuti a comprendere le mie osservazioni scientifiche presentate nei capitoli finali, dove cercherò di presentare le mie idee sul tema del rapporto fra i suoni e il riconoscimento dei significati.

In questo volume, non saranno trattati i temi di fonetica acustica e articolatoria considerando che chi giunge su queste pagine ha già avuto modo di incontrare questi argomenti. Allo stesso modo, i principi della biolinguistica e delle mappe corticali e funzionali che tanto care sono ai neurolinguisti, saranno appena accennati e probabilmente, per certi aspetti, anche questo potrà far apparire incompleto il trattamento del tema principale, chi fosse interessato ad approfondire questi aspetti, troverà un'interessante rassegna nel volume “Il cervello fonologico” di Grimaldi (2019).

Il volume, quindi, descriverà le principali teorie di percezione fonetica e di psicolinguistica dei processi uditivi che sono state proposte negli ultimi settanta anni circa. Tenterò, sia pure in termini a volte sintetici, di metterne in evidenza tratti comuni, motivi di discussione e di opposizione, di mettere in risalto la visione generale, a volte unificante, che, nel tempo, è arrivata a manifestarsi attraverso processi

di avvicinamento di teorie talvolta simili e talvolta opposte; attraverso compromessi costruttivi che prendevano “il meglio” delle teorie da cui si partiva, per giungere ad una ragionevole spiegazione di fenomeni che, tuttavia, sono ancora ignoti anche alla luce di molte delle recenti scoperte.

Nella parte finale, quando il manuale, almeno nelle mie intenzioni, dovrebbe diventare saggio, parleremo poi di due aspetti cruciali per dare una visione completa del tema che ho deciso di trattare: i) il ruolo del parlato naturale, quello delle conversazioni di tutti i giorni, così diverso dal parlato che quasi sempre si usa negli esperimenti sulla percezione e riconoscimento del parlato; ii) delle nuove aspettative prodotte dalla elaborazione automatica delle lingue parlate nell’era dell’intelligenza artificiale. Per alcuni versi, gli incomprensibili meccanismi delle sofisticate *deep neural networks* ci offrono un ragionevole modello che potrebbe spiegare, in base a una sorta di processo imitativo, come gli umani apprendono prima e riconoscono poi le parole parlate.

Nel suo volume “Dei suoni e dei sensi” Federico Albano Leoni (spero non me ne voglia se sintetizzo troppo) insiste su alcuni punti chiave: i) il componente più piccolo individuabile all’interno della catena fonica, che, in maniera compositiva, porti alla comprensione del messaggio acustico, non può essere immediatamente riconducibile ad una unità astratta generata dalla nostra competenza di scriventi, ii) il contesto ci permette di disambiguare e dare un “senso” ad ogni segnale fonico anche in presenza di stimoli che in alcuni punti possono essere fortemente impoveriti dai fenomeni di coarticolazione, mancato raggiungimento del *target* articolatorio, erosione fonetica tipica del parlato spontaneo, iii) attraverso processi paralleli che agiscono anche su scale temporali differenti, lo stimolo sonoro viene “mappato” direttamente nei significati che convoglia.

In linea di principio condivido tutte queste affermazioni, ne abbiamo discusso insieme per anni, ma quella più difficile da accettare è la terza, perché prevede un “salto” funzionale che è contraddetto dalle attuali nostre conoscenze sull’organizzazione dell’intero sistema uditivo, dalla periferia alle aree primarie corticali e verso le aree monolaterali specializzate. Il segnale viene analizzato e riconfigurato ad ogni passaggio per le “stazioni” cognitive lungo il percorso, per potere disporre del

corretto contesto da usare per la “*reconstructio ad sensum*” serve un buffer di memoria ecoica e un principio di sincronia temporale fra ascoltatore e parlante. Serve, inoltre, un meccanismo di ricezione delle parti più significative dello stimolo su cui chi ascolta concentra massimamente. Sicuramente un processo di inferenza *top-down* si sovrappone a quello *bottom-up* strettamente legato all’analisi del segnale, ed è probabile che in questo senso il significato linguistico si costruisca in maniera parzialmente indipendente dal contenuto acustico dello stimolo, ma non basta. In pratica serve un’interfaccia fra suoni e sensi. Discuterò di questi aspetti, anche se non conosco la risposta a tutte le domande. Buona lettura.