

UNIVERSITÀ DEGLI STUDI DI PALERMO

FACOLTÀ DI LETTERE E FILOSOFIA

---

DOTTORATO DI RICERCA IN FILOSOFIA DEL LINGUAGGIO,  
DELLA MENTE E DEI PROCESSI FORMATIVI

*Sintassi, Prosodia e Socialità:  
le Origini del Linguaggio  
Verbale*

*Tutor:*  
*Prof. Marco Carapezza*

*Tesi di Dottorato di:*  
*Piera Filippi*

*Co-tutor:*  
*Prof. William Tecumseh Fitch*

*Coordinatore del dottorato:*  
*Prof. Franco Lo Piparo*

Settore disciplinare M-FII/05

---

Ciclo XXIII - Anni Accademici 2009-2011

# Sintassi, Prosodia e Socialità: Le Origini del Linguaggio Verbale

<b>Introduzione</b>	p. 4
<b>SEZIONE I: “SINTASSI”. UN APPROCCIO COMPARATIVO</b>	p. 18
<b>1. <i>Umwelt</i> e sintassi: il mondo specie-specifico</b>	p. 19
1.1 La nozione di <i>Umwelt</i> : il mondo non è una scatola	p. 19
1.2 Se ho un martello, vedo chiodi	p. 21
1.3 Connessioni: la specificità delle <i>Umwelten</i>	p.28
<b>2. Canti, vocalizzi, e proposizioni: similarità oltre le differenze</b>	p. 33
2.1 L’ipotesi innatista	p. 36
2.2 Un povero stimolo	p. 45
2.3 Grammatiche per lo studio della sintassi	p. 48
<b>3. Percezione di pattern fono-sintattici: apprendimento statistico e acquisizione di regole</b>	p. 52
3.1 Apprendimento statistico negli umani	p. 52
3.2 L’apprendimento di regole negli infanti pre-verbali	p. 67
3.3 Acquisizione di regole e apprendimento statistico nelle scimmie	p. 70

<b>4. Percezione di pattern fono-sintattici: il paradigma chomskiano</b>	p. 77
4.1 Facoltà comuni e innovazioni evolutive	p. 85
4.2 Percezione di grammatiche artificiali nelle scimmie	p. 89
4.3 Percezione di grammatiche artificiali negli uccelli canori	p. 94
<b>5. Comprensione dei pattern proposizionali, ovvero la co-determinazione tra sintassi e uso</b>	p. 105
5.1 Vedere una regola: comprensione o categorizzazione percettuale?	p. 105
5.2 Scheletri computazionali: similarità e differenze logico-sintattiche tra regole percettuali, algebriche e fraseologiche	p. 110
5.3 Strutture fraseologiche. Critiche al paradigma chomskiano	p. 117
<b>6. Sintassi: studi comparativi sulla produzione di pattern fonici</b>	p. 124
6.1 Strutture fonologiche. Requisiti anatomico-neurali	p. 124
6.2 Strutture protosintattiche fra le vocalizzazioni delle scimmie	p. 127
6.3 Strutture fonologiche fra i canti degli uccelli	p. 143
6.4 Dai canti alle parole-note	p. 150
 <b>SEZIONE II: IL VALORE (SOCIO-)COGNITIVO DELLA SINTASSI PROSODICA</b>	 p. 157
<b>7. L'ipotesi di Darwin: alle origini era il canto</b>	p. 158
7.1 Duetti e legami di coppia	p. 160
7.2 La selezione dei canti	p. 163
7.3 Vincoli ed analogie adattativi: il ruolo delle interazioni sociali	p. 168
<b>8. Strutture prosodiche alle origini del linguaggio</b>	p. 176
8.1 Animali non umani: riconoscimento e uso di indizi prosodici	p. 176
8.2 La processazione dei segnali linguistici: indizi che fanno la differenza	p. 185
8.3 La portata emozionale e cognitiva del <i>Motherese</i>	p. 190
8.4 L'ipotesi del <i>bootstrapping</i> prosodico	p. 203
8.5 Prosodia e contesti cross-situazionali	p. 213
<b>9. Prosodia, musica e facoltà di linguaggio: coincidenze evoluzionarie?</b>	p. 225
9.1 Musica: somiglianze di famiglia (in specie diverse)	p. 228
9.2 Prosodia: la musica dei discorsi	p. 240
9.3 Linguaggio e musica: affinità selettive	p. 250
9.4 Co-evoluzione tra mente sociale e linguaggio: il ruolo della modulazione musicale	p. 261
9.5 Musica: adattamento o dessert?	p. 274

<b>SEZIONE III: LA MATRICE SOCIALE DEL LINGUAGGIO</b>	p. 280
<b>10. Azioni congiunte: l'origine della mente sociale</b>	p. 281
10.1 La comunicazione "intenzionale" dei primati	p. 282
10.2 Le origini della intenzionalità condivisa	p. 293
10.3 Una questione di vita o di morte: la selezione di gruppo	p. 303
10.4 L'altruismo adattivo	p. 312
10.5 Strategie di coesione: gossip, promesse e rituali	p. 314
<b>Conclusioni</b>	p. 329
<b>Bibliografia</b>	p. 331

## **Introduzione**

Che cos'è il linguaggio? Cos'è la mente? E in che modo la realtà vissuta ne desume il proprio statuto ontologico? Sono queste le domande che hanno da sempre animato la mia passione per la filosofia conducendo il mio percorso di studio fino alla presente ricerca di dottorato. Non ho ancora trovato (naturalmente) una risposta definitiva a tali quesiti, ma la mia esperienza di ricerca mi ha mostrato la possibilità di un metodo efficace, in funzione del quale porre domande fertili: crediamo che al fine di conoscere entità come il linguaggio o la mente - di cui la nostra stessa forma di vita sipregna fino a renderle invisibili, sia assai fruttuoso il tentativo teorico di risalirne alle origini evolutive. Passando per l'esame di temi recentemente emersi in merito alla problematizzazione dell'origine e della dinamica evolutiva del linguaggio umano, dunque, il fine ultimo della

nostra ricerca è di maturare la comprensione di quanto di umano ci sia in esso, e viceversa, in che modo il linguaggio proposizionale distingue gli esseri umani in quanto tali. Nell'evoluzione del pensiero dall'antichità ai nostri giorni, lo studio dell'origine del linguaggio, insinuatosi con alterne vicende lungo la storia della letteratura filosofico-filologica, ha conosciuto più di un momento di gloria. Ma, nonostante l'interesse manifestato dagli studiosi nel corso dei secoli, non approdò mai a risultati scientifici plausibili, tant'è che nel 1866 i fondatori della *Société de Linguistique* di Parigi presero la decisione di non ammettere più, a norma di statuto, alcuna comunicazione che riguardasse tale ambito di ricerca. Un rinnovato interesse per i problemi riguardanti tale tema si è manifestato a partire dagli anni Sessanta del secolo scorso, anni in cui emersero dati assai importanti per lo studio del linguaggio sia sul piano sperimentale che teorico.

Uno dei nuclei tematici più fertili sul fronte del dibattito interdisciplinare intorno alle origini del linguaggio è costituito dal dibattito moderno inerente la fisionomia del cosiddetto “proto-linguaggio”: un supposto stadio “linguistico” intercorso tra i sistemi di comunicazione dell'ultimo antenato comune a uomini e scimmie antropomorfe e il sistema linguistico-proposizionale dell'*Homo Sapiens*. Le tesi dibattute sono in effetti tante, e spesso in disaccordo fra loro. In particolare, le diverse posizioni sono assimilabili a tre distinti modelli teorici generali: a) del

protolinguaggio lessicale/analitico, b) del protolinguaggio gestuale e c) del protolinguaggio olistico-musicale.

Secondo il modello di spiegazione analitico, le prime forme di linguaggio umano erano costituite essenzialmente da unità lessemiche singole, il cui crescere in complessità e numero sfociò *in seguito* nella nascita della sintassi. Uno dei teorizzatori più rappresentativi di tale posizione è Dereck Bickerton, il quale afferma:

Driven by climate changes into habitats where predators were fierce and common but food was scarce, at least one primate species began to exchange basic information about the environment in order to survive. But proto-language is not bee language. Once invented syllables begin to be used, they can be used to describe anything, consequently language can be adapted for social or any other purposes. So symbols multiplied but structure probably did not, prevented from developing by inadequate numbers of neurons and the right kind of connectivity. Once both of these had developed, things the brain could already do enabled protolanguage to develop quite rapidly into language as we know it today.<sup>1</sup>

Secondo il modello del protolinguaggio gestuale, invece, le prime forme di linguaggio presso gli umani consistevano in gesti che realizzati unitamente ad espressioni pantomimiche – avrebbero garantito una maggiore facilità di comunicazione iconica rispetto a quanto reso possibile dalla mera espressione vocale. In effetti, come notava già negli anni

---

<sup>1</sup> Bickerton, D., *Symbol and Structure: A Comprehensive Framework for Language Evolution*, in M. H. Christiansen, S. Kirby, (ed.), “Language evolution”, Oxford University Press, 2004, pp. 92-93.

settanta Gordon Hewes<sup>2</sup> e, più di recente, il gruppo di ricerca di Micheal Tomasello - tale tesi guadagna plausibilità se si guarda alle modalità comunicative presso le grandi scimmie (le specie a noi filogeneticamente più vicine), le quali, pur mancando della capacità di controllo motorio per l'articolazione dei suoni vocali, adottano una varietà di gesti manuali sia per fini imperativi che per richiamare l'attenzione su qualcosa nell'ambiente circostante. In altri termini, secondo Tomasello, "la strada verso la comunicazione cooperativa umana inizia con la comunicazione intenzionale delle grandi scimmie che si manifesta principalmente nei gesti".<sup>3</sup> Su tale scia si collocano altresì le idee di Micheal Corballis:

The overall evolutionary scenario may have gone something like this. With the emergence of bipedalism, the early hominins evolved more sophisticated ways to gesture to one another than their immediate primate ancestors. But these gestures may still have consisted of relatively isolated signs until around two millions years ago, when brain size increased and migrations out of Africa began. This may have led to the combining of gestures to form new meanings, and perhaps the beginnings of narrative. Thus, eventually, was syntax born.<sup>4</sup>

---

<sup>2</sup> Hewes, G. W., *Primate Communication and the gestural origin of language*, "Current Anthropology", 14, 1973, pp. 5-24.

<sup>3</sup> Tomasello, M., *Origins of Human Communication*, Massachusetts Institute of Technology, 2008; trad. it. di S. Romano, *Le origini della comunicazione umana*, Raffaello Cortina, Milano, 2009, p. 278.

<sup>4</sup> Corballis, M. C., *From Hand to Mouth: The Gestural Origins of Language*, in M. H. Christiansen, S. Kirby, (ed.), "Language evolution", Oxford University Press, 2004, p. 216. Vedi: Corballis, M. C., *Did Language Evolve from Manual Gestures?*, in A. Wray, A. (ed.), "The Transition to Language", Oxford University Press, 2002, pp. 161-79; Arbib, M. A., "The Evolving Mirror System: A Neural Basis for Language Readiness", M. H. Christiansen, S. Kirby, (ed.), "Language evolution", cit., pp. 182-200.



Posta dunque la modalità primariamente gestuale delle prime forme di linguaggio presso i nostri primi antenati umani, gli autori che aderiscono a tale paradigma teorico assumono che il passaggio alla dimensione vocale (evidentemente pervasiva nella comunicazione verbale oggi) abbia preso piede per via dei vantaggi che apportava: la comunicazione a lunghe distanze o al buio, la possibilità di parlare lasciando le mani “libere” di lavorare nel contempo o di focalizzare l’attenzione visiva su altri oggetti. Tuttavia, i sostenitori di tale tesi mancano di specificare in che modo sia avvenuto tale passaggio, da un punto di vista cognitivo-funzionale.

Nell’attuale dibattito intorno alla evoluzione del linguaggio, questi due modelli di teorizzazione sul protolinguaggio si scontrano dialetticamente con il modello olistico-musicale, secondo il quale le prime forme di linguaggio consistevano non di singole etichette lessicali o gestuali usate per riferirsi ad oggetti dell’ambiente esterno utili alla sopravvivenza – bensì di unità vocali olistiche, ovvero di concatenazioni di suoni mancanti di una struttura grammaticale interna ben definita, ma nondimeno dotati di un’intrinseca portata comunicativa.

Il padre di tale teoria è identificabile in Charles Darwin, il quale afferma che le prime forme di linguaggio presso gli uomini dovevano consistere in espressioni con cadenza musicale evolutisi - come nel caso degli uccelli canori o dei gibboni - per ragioni legati alla difesa del

territorio, o alla conquista del partner sessuale<sup>5</sup>. In tale direzione vanno anche le idee di Otto Jespersen secondo cui le prime espressioni proferite dalla neo-specie umana dovevano presentare una forma musicale priva di una mappatura articolata tra le parti costitutive del segnale e le parti del riferimento esterno:

[...] it is not so wild to imagine as the first beginning something which can be *translated* into our languages by means of a sentence, but which is not 'articulated' in the same way as such a sentence. [...] it is just units of this character [...] whose genesis we can most easily imagine on the supposition of a primitive period of a meaningless singing.

[...] Primitive man came to attach meaning to what were originally rambling sequences of syllables in pretty much the same way as the child comes to attach a meaning to many of the words he ears from the elders, the whole situation in which they are heard giving a clue of their interpretation.<sup>6</sup>

Tale tesi è stata recentemente ripresa da Alison Wray, la quale ritiene che simili proto-espressioni linguistiche, erano essenzialmente volte ad indurre una reazione comportamentale nel ricevente, e che i riferimenti (ovvero il ricevente stesso, l'agente chiamato in causa, e l'oggetto predicato) fossero verosimilmente deducibili dal contesto pragmatico d'uso, e dall'apporto costitutivamente integrante della modalità gestuale

---

<sup>5</sup> Darwin, C., *The Descent of Man, and Selection in Relation to Sex*, John Murray, London, 1871, p. 56: "Primeval man, or rather some early progenitor of man, probably first used his voice in *producing true musical cadences, that is in singing*, as do some of the gibbon-apes at the present day."

<sup>6</sup> Jespersen, O., *Language: Its nature, developement and origin*, W.W. Norton & Co., New York, 1922, pp. 440-441

d'espressione<sup>7</sup>. Su tale scia teorica si collocano le idee di Tecumseh Fitch, il quale, con Wray, trova plausibile l'idea che i primi proferimenti degli *Homo* fossero effettivamente dotati di un intrinseco valore performativo e manipolativo. Tuttavia - ritiene l'autore - essi non dovevano essere del tutto privi di una struttura interna, ma al contrario presentare una regolata forma fonica interna, al pari, ad esempio, dei canti degli uccelli. In particolare, Fitch propone l'ipotesi di un percorso evolutivo a più tappe che ha condotto alla comunicazione verbale. La prima di queste sarebbe costituita da un "protolinguaggio prosodico", ovvero un sistema comunicativo preverbale che condivide determinate caratteristiche strutturali con le lingue odierne:

By analogy with the comparative method, the properties of this version of protolanguage would include the two main properties that modern human language and music share: a focus on the vocal/auditory channel, and the complex structure that is learned and culturally transmitted. [...] Vocal learning of *complex structured signals without propositional meaning* has evolved repeatedly among vertebrates, typically under sexual selection.<sup>8</sup>

---

<sup>7</sup> Wray, A., *Protolanguage as a holistic system for social interaction*, "Language & Communication", 18, 1998, pp. 47-67; Wray, A., *Holistic utterances in protolanguage: the link from primates to humans*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy, M. (ed.), "The Evolutionary emergence of language", Cambridge University Press, 2000, pp. 285-302; Wray, A., *Dual Processing in Protolanguage: Performance without Competence*, in A. Wray, (ed.), "The Transition to Language", Oxford University Press, 2002, pp. 113-37.

<sup>8</sup> Fitch, W. T., *The evolution of language: a comparative review*, "Biology and Philosophy", 20, 2005, pp. 220-21 (corsivo mio). Vedi Deacon T., W., *The Symbolic Species. The coevolution of Language and the Brain*, Norton & Company, New York, 1997; trad. it. di S. Ferraresi, *La specie simbolica. Coevoluzione di linguaggio e cervello*, Giovanni Fioriti, Roma, 2001, p. 351: "L'unica classe di funzioni paralinguistiche che probabilmente ha avuto la più lunga relazione coevolutiva con il linguaggio è la prosodia del parlato: le variazioni di ritmo, di volume e di tonalità delle parole, usate sia per dirigere l'attenzione sugli elementi che il parlante giudicasse salienti sia per comunicare un tono emotivo."

In altri termini, l'elemento di continuità che avrebbe segnato il passaggio dal protolinguaggio prosodico a quello proposizionale sono, secondo Fitch, quelle unità melodiche che comprendono in sé una struttura complessa data dal fatto che i passaggi, di per sé privi di significato, si uniscono intersecandosi, o ripetendosi, in base a regole strutturali determinate. Segnali, dunque, che presentano una complessità per così dire semplice, in quanto non dotata di contenuto referenziale, semantico.

Secondo l'autore, inoltre, il passaggio da tale prima fase di comunicazione prosodicamente strutturata allo stadio proposizionale del linguaggio è stato promosso da una seconda "tappa evolutiva", costituita dall'impulso alla comunicazione fra individui geneticamente imparentati<sup>9</sup>. L'idea dello studioso è che l'ipotesi di comunicare circa la presenza di cibo o di pericoli fra consanguinei, favorendo la perpetuazione dei geni comuni – sia stata funzionale alla loro stessa sopravvivenza. Riprendendo le affermazioni di Darwin, Fitch sostiene quanto segue:

Several lines of evidence lead to the hypothesis that an a-referential songlike communication system predated language in human evolutionary history. [...] This makes a hypothesis in which complex signals ("song") evolved first, and that meanings were added to these signals later.<sup>10</sup>

---

<sup>9</sup> Fitch, W. T., *The evolution of language: a comparative review*, cit., p. 221: "if we add to Darwin's prosodic protolanguage a second stage, in which kin communication drove the addition of a fully propositional and intentional semantics plausible analytic dual stage results."

<sup>10</sup> Fitch, W. T., *The biology and evolution of music: A comparative perspective*, "Cognition", 100, 2006, p. 198. Cfr: Mithen, S., *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*,

Le diverse prospettive di studio qui delineate forniscono degli scenari di spiegazione estremamente plausibili. Tuttavia, dal momento che il linguaggio verbale non è stato interessato da processi di fossilizzazione in grado di rivelarne l'effettivo *iter* evolutivo, assumiamo che le teorie inerenti il protolinguaggio siano destinate a restare confinate perlopiù al piano speculativo. Cionondimeno, fra tutti, il modello olistico-musicale inaugurato da Darwin è quello che a nostro avviso, oltre ad alimentare la ricerca teorica sull'evoluzione del linguaggio, permette di guardare al linguaggio da una prospettiva funzionalistica che ne amplifica la comprensione, fornendo molteplici quesiti fertili per la ricerca sperimentale, quali i seguenti: in che modo la complessità fonica (eventualmente regolata) dei messaggi olistici ha trovato una propria declinazione nella combinazione strutturata di unità semantiche discrete? Quali meccanismi cognitivi fondano tale processo di costituzione e comprensione intersoggettiva delle unità semantiche discrete entro contesti sintatticamente regolati?

Per questa ragione ci iscriviamo a tale paradigma di spiegazione, ma con la volontà di interpretarne le assunzioni in chiave funzionale. In altri termini, intendiamo il nostro studio di ricerca finalizzato non *sic et simpliciter* al guadagno di una mera teoria sulla evoluzione del linguaggio.

---

Harvard University Press, 2007; Brown, S., *The "Musilanguage" Model of Music Evolution*, in N. L. Wallin, B. Merker, S. Brown, (ed.), "The Origins of Music", Mit Press, Cambridge, 2001, pp. 271-300.

Al contrario, vediamo in tale tipo di analisi la metodologia da adottare, ovvero il mezzo attraverso cui risalire alla enucleazione dei meccanismi operanti nel complesso processo di produzione/comprendimento del linguaggio verbale.

Consapevoli del fatto che parlare di “linguaggio” in riferimento ai sistemi di comunicazione non pienamente proposizionali può essere pericolo di confusione concettuale, teniamo a sottolineare che adottiamo il termine “linguaggio” nel suo senso stretto, che identifichiamo con la nozione di “sintassi”. Come vedremo in maggiore dettaglio nel primo capitolo – intendiamo tale termine nell’accezione greca di “*syntaxis*”, ovvero come *capacità di connettere elementi diversi secondo regole strutturali*. Dunque, lo studio sulla evoluzione del linguaggio coincide, nel nostro caso, con l’analisi dell’evoluzione della capacità-chiave di sintassi, da condurre tanto sul piano filogenetico, quanto su quello ontogenetico. Al fine di individuare quei meccanismi cognitivi che, operando al cuore della nostra stessa attività di linguaggio, ne venano l’essenza – è infatti nostro avviso necessario adottare un approccio comparato fra i diversi sistemi di cognizione e comunicazione non verbale –non soltanto in altre specie animali a noi evolutivamente imparentate, ma anche negli infanti pre-verbali. L’analisi delle menti non (e pre-) verbali, infatti, costituisce uno sfondo efficace per l’indagine volta alla comprensione della natura del

linguaggio umano, in grado di rivelare la natura di quei meccanismi che determinandone l'emergenza, ne rappresentano lo scheletro funzionale. Un approccio in grado, dunque, di spingere in avanti la comprensione del concetto di mente, e dei suoi costitutivi processi linguistici.

In particolare, lo studio dei sistemi cognitivi di specie animali sul piano filogenetico permette di risalire alle similarità che stanno al di là delle differenze specifiche (e viceversa), individuando in tal modo la l'evoluzione genetica lineare e le pressioni selettive che hanno condotto alla attuale configurazione del linguaggio umano. È opportuno, a tal proposito, distinguere due classi di tratti cognitivi cruciali: i tratti detti "omologhi", e quelli "analoghi". L'individuazione dei primi rivelerebbe i nuclei cognitivi che uomini e scimmie hanno ereditato dall'antenato comune (vissuto circa 6 milioni di anni fa), e di conseguenza, gli elementi chiave nella genetica lineare del linguaggio. L'individuazione dei tratti "analoghi", invece, ci permette di guadagnare una comprensione maggiore in merito delle condizioni ambientali o delle pressioni selettive che, a prescindere dalla appartenenza alla medesima linea evolutiva, hanno condotto specie filogeneticamente distanti come gli uccelli e gli esseri umani - allo sviluppo delle medesime caratteristiche. Un esempio di tratto analogo è la forma del corpo dei delfini e delle balene che, a prescindere dall'esistenza di un antenato comune, si è imposta come la forma più adatta

a fronte dei vincoli fisici imposti dal nuoto. Sul piano ontogenetico, d'altra parte, l'analisi delle abilità cognitive presenti nella fase pre-verbale può illustrare i pre-requisiti socio-cognitivi necessari alla emergenza delle abilità di comprensione e produzione di espressioni proposizionali.

In tale direzione, il nostro studio si avvale di contributi teorici provenienti dall'intersezione di più branche, dalla psicologia evolutiva alla archeologia, dalla biologia alle neuroscienze, che operano all'unisono nel tentativo di tracciare una mappatura degli "strumenti mentali" condivisi nel mondo animale (umano e non). In altri termini, gettando luce sui vincoli ecologici o sociali che avrebbero favorito (e favoriscono tutt'oggi) l'evoluzione della mente umana, tale collaborazione interdisciplinare ha come obiettivo finale il guadagno di una comprensione più lucida dei processi cognitivi coinvolti in quelle articolazioni sociali e linguistiche che ci contraddistinguono in quanto umani. Tale è dunque la prospettiva teorica entro il quale si colloca la presente ricerca, in cui filosofia e scienza sperimentale si incontrano in una viva dinamica dialogica fatta di uno scambio di interrogativi, mezzi interpretativi e tentativi di risposta sul concetto di linguaggio.

Il presente lavoro si suddivide in tre sezioni. La prima si incentrerà proprio sulla nozione di sintassi: dopo aver chiarito in che modo l'abilità di



sintassi sia in grado di fondare l'accesso (onto)logico alle specie-specifiche *Umwelten*, prenderemo in esame recenti ricerche inerenti diverse forme dicompetenza sintattica, la cui presenza è stata sperimentalmente attestata(sia sul piano della percezione che su quello della produzione) in specie animali non verbali e negli infanti umani. Sulla base di tali riflessioni, nella seconda sezione metteremo a tema una modalità specifica di strutturazione sintattica, ovvero quella prosodico-musicale. Attraverso un'analisi comparativa delle diverse forme musicali presenti fra le vocalizzazioni degli animali non umani e i discorsi proferiti dagli umani, enucleremo una definizione operativa del concetto di musica, in grado di rivelarci le pre-condizioni linguistiche in grado di avviare l'evoluzione del linguaggio verbale tanto a livello *filogenetico* che *ontogenetico*.

Si tratta, come mostreremo, di processi cognitivi di base che operano sulla dimensione acustica, rivestendo un ruolo fondamentale non soltanto nel processo ontogenetico di acquisizione del linguaggio, ma anche nella determinazione del senso di una frase, ovvero della sua comprensione: quegli stessi processi cognitivi che, *nel* gettare le fondamenta strutturali per l'emergenza del linguaggio verbale presso i primi ominidi, ne hanno favorito la disposizione ad esprimere le proprie emozioni agli altri, di creare legami, situazioni di complicità, o indurre stati di allerta. È così che - ascoltando la propria intransigente inclinazione a

comunicare le proprie passioni, le emozioni, le paure ai propri simili, cementando i legami nel gruppo, e fortificandolo, in tal modo, contro gli attacchi esterni *attraverso* strutture sintattico-musicali - l'animale umano portò fuori il suo ambiente bio-logico, un mondo che si fa guardare *attraverso* pensieri articolati e prassi comuni, un mondo in cui ci riconosciamo a partire dalle parole degli altri: parole note che ancora oggi, combinandosi come i suoni di una melodia, generano pensieri e linguaggi nuovi.

Nella terza sezione, infine, passeremo in esame alcune importanti ricerche che identificano proprio nella dimensione sociale della mutua cooperazione la matrice del linguaggio verbale, che ancora oggi instaura con esso una dinamica co-evolutiva. Si tratta di una dimensione entro cui è possibile conferire un senso compiuto persino alla più vaga delle espressioni linguistiche, trascinando in avanti il processo di costituzione di sensi (e significati).

**SEZIONE I:  
“SINTASSI”. UN APPROCCIO COMPARATIVO**

Noi predichiamo della cosa quello che troviamo nel  
metodo di rappresentarla.

**L. WITTGENSTEIN**

# 1. *Umwelt* esintassi: il mondo specie-specifico

## 1.1 La nozione di *Umwelt*: il mondo non è una scatola

In *Della certezza* Ludvig Wittgenstein scrive: «Il medico dei matti potrebbe forse chiedermi: “Sai cos’è questo?” e io potrei rispondergli: “So che questa è una sedia, la riconosco, è sempre stata nella mia stanza”»<sup>11</sup>. Soltanto a chi mette in dubbio il nostro stato di salute mentale, invero, può venir in mente di chiederci se ri-conosciamo qualcosa nella nostra stanza, se vediamo una sedia in quanto tale. Percepire sedie, finestre, rubinetti, persone è per noi una cosa tanto naturale, al punto che non ci sembra molto sensato porci delle domande su tale fenomeno più che intorno al fatto che mangiamo, camminiamo o beviamo<sup>12</sup>. Ebbene, in questo capitolo, un po’ a

---

<sup>11</sup> Wittgenstein, L., *On certainty*, Basil Blackwell, Oxford, 1969; trad. it. di M. Trinchero, *Della certezza. L’analisi filosofica del senso comune*, Einaudi, Torino, 1999, § 355.

<sup>12</sup> Pertinente a tal proposito Wittgenstein, L., *Philosophische Untersuchungen*, Basil Blackwell, Oxford, 1953; trad. it. di M. Trinchero, *Ricerche filosofiche*, Einaudi, Torino, 1999, § 25: «[...] Il comandare,

modo dei medici per i matti, ci soffermeremo sulla caratteristica naturalezza di tali azioni e percezioni, indagandone lo *status* ontologico, alla luce del confronto con le altre specie animali.

La nostra indagine prende le mosse dal nodo teorico cruciale del programma di ricerca di Uexküll, noto zoologo e filosofo naturalista attivo nella prima metà del '900:

Una delle illusioni su cui è più facile cullarsi, è che il rapporto del soggetto con il mondo si svolga nello stesso spazio e nello stesso tempo in cui si svolgono quelli che ci legano al nostro mondo: e ad alimentare questa illusione concorre la credenza che esista un solo mondo, in cui tutti i viventi sono contenuti come in una scatola; e che perciò spazio e tempi siano uguali per tutti<sup>13</sup>.

L'idea qui espressa è che il Mondo non è una entità unica ed indifferenziata per ogni essere vivente. Esso non è un mero contenitore, né un singolo insieme complesso di oggetti ed esseri animati fra loro in relazione, a cui ogni specie ha accesso in gradi diversi. La differenza che passa tra le modalità di accesso epistemico al mondo da parte delle diverse specie viventi non sta nella quantità di informazioni cui si viene in contatto, ma nelle *specifiche* modalità biocognitive in cui ciò avviene, modalità che

---

l'interrogare, il raccontare, il chiacchierare, fanno parte della nostra storia naturale come il camminare, il mangiare, il bere, il giocare».

<sup>13</sup> Uexküll, J. von e Kriszat, G., *Streifzüge durch die Umwelten von Tieren und Menschen: Ein Bilderbuch unsichtbarer Welten*, 1934; trad. it. di P. Manfredi, *I mondi invisibili*, 1936, rist. *Ambiente e comportamento*, Il Saggiatore, Milano, 1967, p. 103. Vedi Carapezza, M., *Uexküll, la nozione di Umwelt e il parlare di animali*, in *Annali del dipartimento di Filosofia, Storia e Critica dei Saperi* (Università di Palermo), 2006, p. 50.

ne fondano lo specifico statuto ontologico. Il nocciolo duro della teoria di Uexküll è, dunque, che esistono tanti *mondi-ambienti* (*Umwelten*) quante sono le specie viventi.

Ognuna di esse (dall'euglena agli umani), infatti, è dotata di uno specie-specifico apparato biologico/cognitivo che ne fonda l'accesso alla realtà, consentendo in tal modo di disporre delle informazioni pertinenti, o meglio, funzionali alla sopravvivenza della specie data<sup>14</sup>. L'insieme di tali informazioni costituisce una *Umwelt* specie-specifica, in cui rientrano, appunto, tutti quegli oggetti la cui percezione ci sembra naturale e ovvia quanto il mangiare, o il respirare.

## 1.2 Se ho un martello, vedo chiodi

Uexküll afferma che «[La *Umwelt*] è l'insieme di mondo sensibile, tutto ciò che un soggetto percepisce e registra, e di un mondo effettuale,

---

<sup>14</sup> La nozione di *Umwelt* viene altresì tematizzata in Heidegger, M., *Die Grundbegriffe der Metaphysic. Welt – Endlichkeit – Einsamkeit*, Klostermann Verlag, Frankfurt ma Main, 1983; trad. it di P. Coriando, *Concetti fondamentali della metafisica. Mondo – Finitezza – Solitudine*, Il Melangolo, Genova, 1992; vedi pp. 252-53: «Infatti cadiamo subito in un grande imbarazzo, in merito alla questione intorno alla maggiore o minore perfezione dell'accessibilità dell'ente, se ad esempio, paragoniamo la facoltà visiva dell'occhio del falco con quella dell'occhio umano, oppure l'olfatto del cane con quello dell'uomo. [...] È vero che siamo abituati anche qui a parlare di animali superiori e inferiori, eppure è un errore fondamentale ritenere che amebe e infusori siano animali meno perfetti di elefanti e scimmie. Ogni animale e ogni specie è in quanto tale esattamente perfetta come le altre. Da tutto quanto abbiamo detto risulta chiaro: se parliamo di una povertà di mondo e di una formazione di mondo, ciò non deve fin da principio venir inteso nel senso di un ordinamento gerarchico di carattere valutativo. [...] Essere povero non significa semplicemente non possedere nulla o poco o meno dell'altro, bensì essere povero significa *fare a meno*. Questo fare a meno, a sua volta, è possibile in maniere diverse – a seconda di *come* il povero fa a meno, cioè di quale condotta mantiene nel fare a meno, di come si pone nei confronti di esso, di *come* considera il fare a meno, in breve, a seconda di *a che cosa* fa a meno e soprattutto di *come* ne fa a meno, cioè di come si sente nel farlo».

tutto ciò che esso opera, o pone in atto o effetto. L'insieme di questi universi forma un'entità chiusa»<sup>15</sup>. In altri termini, la percezione del mondo da parte di ogni specie vivente deriva da uno strettissimo intreccio fra sensorialità e cognitività, in quanto costitutivamente connessa ai bisogni vitali della specie; così, la percezione meramente sensibile di un oggetto è venata dalla possibilità di *riconoscere* in esso qualcosa di essenziale alla vita. Nell'ottica dell'autore, infatti, «non esistono oggetti neutrali. Ogni cosa che cade all'interno di una *Umwelt* è modificata e riformata fino a che essa divenga *significativa*»<sup>16</sup>. In tale direzione, è importante sottolineare come la significatività di una caratteristica ambientale sia immediatamente connessa con la possibilità di attribuirvi una qualità funzionale, una disposizione all'uso. Ciò significa che qualcosa può rientrare nel campo percettivo/cognitivo di un essere vivente, e dunque far parte della sua specifica *Umwelt* se la sua percezione sensoriale comprende un intrinseco ed immediato riconoscimento della funzionalità effettuale, ovvero di ciò che per mezzo di esso si può fare:

Possiamo dire che ogni soggetto animale prende contatto con il suo oggetto mediante una tenaglia a due branche: una percettiva ed un attiva: la prima branca conferisce all'oggetto un carattere distintivo, la seconda vi imprime un carattere determinato dall'atto, vale a dire effetto di questo, e che perciò

---

<sup>15</sup> Uexküll, J. von e Kriszat, G., Op. cit., p. 79.

<sup>16</sup> Uexküll, J. von, *Bedeutungslehre*, Leipzig, 1940; trad. ing. *Theory of Meaning, Semiotica*, 42, Mouton de Gruyter, Berlin-New York, 1982, pp. 25-80 e 31, cit. in Carapezza M., cit., p. 51 (corsivo mio).

chiameremo effettuale. [...] Si dirà dunque che il carattere effettuale cancella il carattere percepito o, più brevemente: *l'effetto annulla la percezione*<sup>17</sup>.

L'implicito riconoscimento della funzione di un oggetto, dunque del suo effetto all'interno della *Umwelt*, ne colora in modo precipuo la percezione, riorganizzando il dato sensoriale all'interno di una specifica connessità cognitivo-funzionale. Per dirla con Abraham Maslow, se l'unico utensile che abbiamo a disposizione è un martello, tenderemo a vedere ogni problema come un chiodo<sup>18</sup>.

Contraddistinto in tale maniera, un oggetto cessa di essere semplicemente un elemento confinato all'ambito sensoriale, per farsi *portatore di significato*, cioè per esser percepito *in quanto* oggetto recante tratti pertinenti (riconoscibili come tali dagli organi biocognitivi della data specie) all'interno della sua *Umwelt*, ovvero in quanto dotato delle possibilità di espletare determinate funzioni. Ad esempio, un essere umano può percepire uno stesso bastone come un componente costitutivo di una scopa, o come un mezzo per agganciare un oggetto posto ad una certa distanza. Allo stesso modo, un paguro può percepire le attinie come animaletti in grado di pulire e proteggere il suo guscio, o nel caso in cui

---

<sup>17</sup> Uexküll, J. von e Kriszat G., Op. cit., pp. 92-93.

<sup>18</sup> Maslow, A. H., *The Psychology of Science*, Henry Regnery Co, 1970, pp. 15-16: "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail".



fosse privo di un guscio da far mantenere, essi verranno percepiti come un potenziale pasto:

Per comprendere questi diversi comportamenti del crostaceo non basta considerare le diverse figure percepite, che gli forniscono i suoi organi di senso; ma occorre ammettere l'intervento di una qualità o tonalità effettuale [*Wirkungton*], che integra la figura percepita in figura effettuale. A seconda della tonalità effettuale che si sovrappone alla figura percepita, il medesimo oggetto può dar luogo a diverse figure effettuali [*Wirkbild*].<sup>19</sup>

Un po' come accade nel caso di un riorientamento gestaltico, uno stesso oggetto può dunque esser visto come cose diverse in base alla propria effettualità<sup>20</sup>. E, in base allo stesso principio gestaltico, uno stesso insieme di oggetti e stati di cose, dato come la somma di tutte le *Umwelten* delle specie viventi (o, più semplicemente una stessa porzione di dati accessibili allo stesso tempo da più specie) possono esser visti da ciascuna come mondi totalmente diversi. E ciò avviene, come abbiamo visto, in funzione dei meccanismi percettivo-cognitivi che fanno parte del corredo biologico di ciascuna specie.

La chiarezza di tale concetto può emergere meglio alla luce di un confronto con una nozione parallela del dominio matematico. Per dirla con Lo Piparo:

---

<sup>19</sup> Uexküll, J. von e Kriszat, G., Op. cit., pp. 172-73.

<sup>20</sup> Vedi Wittgenstein, L., *Philosophische Untersuchungen*, cit., parte II, § XI, pp. 255-71.

La capacità che ciascuna specie ha di accedere al mondo è rappresentabile come una complicata funzione matematica che genera coppie formate da “fatti del mondo (insieme X degli argomenti) e stati mentali (insieme delle immagini Y dell’insieme X)”. [...] Al cambiare delle funzioni zoocognitive cambiano anche le coppie “fatti del mondo-immagini mentali” esattamente come il numero 6 fa coppia col (ha come propria immagine il) numero 3 rispetto alla funzione  $x/2$ , fa coppia col (ha come propria immagine il) numero 2 rispetto alla funzione  $x/3$  ma non è accoppiabile con (non ha come immagine) nessun numero razionale rispetto alla funzione  $\sqrt{x}$ . [...] Nella descrizione e spiegazione delle zoocognitività il dato primario non è né il mondo né la mente (uso questo termine in un’accezione così lata da includere ogni forma di sensibilità) ma la funzione che governa la coppia *Umwelt*-mente.<sup>21</sup>

In tal senso, i meccanismi che strutturano l’apparato biocognitivo di ogni singola specie, rappresentano lo specifico algoritmo implementabile sull’insieme dei fatti/oggetti del mondo (insieme X degli argomenti). Il valore Y della peculiare funzione presa in esame è l’insieme delle immagini-percepite dal corredo biologico specie-specifico. È quest’ultima, infatti a determinare cosa è immagine di che cosa. Ad esempio, un caso interessante, oltre che curioso, è quello della “percezione della morte” da parte delle formiche. Quando queste muoiono emanano una sostanza chimica chiamata acido oleico. Il sociobiologo Ed O. Wilson notò che quando le formiche scoprono un membro della colonia morto, lo portano fuori dalla colonia, e lo abbandonano. Ciò che indica la morte presso tale specie non è la totale immobilità, e la cessazione dell’attività cardiorespiratoria. Nella loro *Umwelt*, la morte viene “significata” dal

---

<sup>21</sup> Lo Piparo, F, *Il mondo, le specie animali e il linguaggio. la teoria zoocognitiva di Wittgenstein*, in M. Carenini e M. Matteuzzi (a cura di), *Percezione, Linguaggio, Coscienza. Saggi di filosofia della mente*, Quodlibet, Macerata, 1998, p.188.

particolare odore provocato dall'acido oleico. Infatti, se si cosparge una formica in grado di respirare e compiere le proprie attività (ovvero ciò che gli esseri umani considererebbero una formica viva, in base ai propri parametri bio-cognitivi, e culturali) con tale acido, le sue con-specifiche la trascinano fuori dal formicaio, segno del fatto che la percepiscono morta. Volendo rendere tale caso in termini matematici, possiamo dire che nel loro mondo il valore  $y$  (formica morta - a prescindere dalla considerazione del fatto che respiri, o meno) è dato dalla funzione-odore dell'acido oleico applicata su  $x$  (corpo della formica). Un ulteriore esempio è fornito dalla percezione dei suoni da parte degli esseri umani. Essi, a differenza di altri animali, sono in grado di percepire esclusivamente quei suoni compresi solo entro una determinata lunghezza d'onda. La loro *Umwelt*, in questo senso, comprende ciò che è accessibile al loro apparato uditivo, differente da quello dei delfini, o dei cani, che invece è in grado di percepire gli ultrasuoni. Lo stesso vale per lo spettro della luce, di cui può vedere (solo) una specifica porzione.

In tale prospettiva, si comprende bene che l'uomo non sta al vertice di una piramide cognitivo-percettiva in grado di rendergli un accesso privilegiato (in termini sia quantitativi che qualitativi) al mondo, inteso dunque come unica entità data, ed indifferenziata per tutti gli organismi. Al contrario, egli ha accesso ad un mondo, o meglio una *Umwelt*, il cui statuto

ontologico, fondato appunto dalla sua specie-specifica dotazione biocognitiva, si può considerare a buon diritto pari (seppur diverso) a quello di tutte le altre specie viventi.

Ora, il *punctum saliens* del presente capitolo è che ciò che rende alla specie umana l'accesso epistemico al mondo è il linguaggio proposizionale. Esso, per esprimerci alla Uexküll, è l'organo cognitivo-percettivo specifico degli uomini, che permette la pertinentizzazione di quegli aspetti della realtà significativi ed effettuali. In altri termini, attraverso il linguaggio, passando per le sue articolazioni, la specie umana percepisce oggetti (o meglio, come vedremo più avanti, connessioni), laddove in tale approccio è immediatamente data la conoscenza di ciò che ne possiamo fare. Esso fonda l'ontologia della Umwelt umana, e vi cala l'atmosfera di una specifica forma di vita (*Lebensform*).

A questo punto della nostra trattazione, una volta guadagnata la nozione di *Umwelt*, e chiaritone la matrice fondante nel caso della specie umana, mi appresto a gettar luce su un ultimo nodo tematico essenziale, inerente il linguaggio animale, e in modo particolare quello umano; si tratta del tema delle connessioni cosali.

### 1.3 Connessioni: la specificità delle *Umwelten*

Il nodo concettuale del presente paragrafo riguarda la modalità epistemica universale di accesso al mondo. Evidentemente, si tratta di una nozione assai controversa. Per questo lo introduco con le parole di uno dei massimi studiosi del linguaggio del secolo scorso, Ludwig Wittgenstein:

Il mondo è la totalità dei fatti, non delle cose.

Ciò che accade, il fatto, è il sussistere di connessioni cosali (*Sachverhalten*).

Lo stato di cose è un nesso d'oggetti (entità, cose)

[...] Come noi non possiamo affatto pensare oggetti spaziali che siano fuori del tempo, così noi *non* possiamo pensare un oggetto che sia fuori della possibilità della sua relazione con altri.<sup>22</sup>

Secondo la prospettiva wittgensteiniana, dunque, gli oggetti sono come i punti nodali di un'ampia ragnatela, e ciò che forma le immagini specie-specifiche sono i *nessi* fra gli oggetti del mondo. Rifacendoci ancora una volta al linguaggio matematico, possiamo dire che l'elemento universale, comune a tutti gli algoritmi in grado di generare le *Umwelten* di tutti gli animali è la capacità di cogliere le relazioni, le connessioni fra le parti. Lo Piparo illustra in modo lucido tale concetto, spiegando la posizione wittgensteiniana:

---

<sup>22</sup> Wittgenstein, L., *Tractatus logico-philosophicus*, Routledge and Kegan Paul, Oxford, 1961, trad. it. di A. G. Conte, *Tractatus logico-philosophicus*, Einaudi, Torino, 1998, proposizioni 1.1, 2 (qui mantengo la traduzione di *Sachverhalten* proposta da Lo Piparo in *Il mondo, le specie animali e il linguaggio. la teoria zoocognitiva di Wittgenstein*, cit., p. 199), 2.01, 2.0121.

La tesi che qui Wittgenstein sta sostenendo è abbastanza semplice: in tutte le *Umwelten* gli oggetti [*Gegenstände*], le cose [*Dingen, Sagen*] non possono esistere se non come parti di relazioni [*Verbindungen*] tra oggetti o cose. [...] *Il mondo a cui le specie animali accedono non è formato da cose o oggetti ma da nessi o articolazioni di cose-oggetti.*<sup>23</sup>

Ciò vuol dire che del mondo (inteso come la totalità delle *Umwelten*) vengono percepite in prima istanza le relazioni fra le parti, le loro articolazioni interne<sup>24</sup>. Dunque, la facoltà di percepire tali *connessioni cosali* è presente in ogni specie vivente, sebbene negli umani spicchi in complessità. Secondo il quadro teorico di Wittgenstein, negli uomini le “antenne” cognitive in grado di configurare quella che percepiscono come il proprio mondo (ovvero la propria *Umwelt*) sono le proposizioni linguistiche, le cui coordinazioni logico-sintattiche captano i nessi nel mondo. E sono proprio tali coordinazioni lo schema di traduzione, o forse di mascheramento<sup>25</sup> che foggiano l’ambiente dell’uomo secondo quelle radicali, specie-specifiche articolazioni sociali o linguistiche che ne determinano *la* forma di vita. Ed è proprio in virtù di tali “antenne” biologiche, che ogni individuo, a prescindere dalla cultura o gruppo sociale o linguistico di appartenenza, è in linea di principio in grado di comprendere *l’altro*, di iniziare con esso dei processi di negoziazione di senso, o di

---

<sup>23</sup> Lo Piparo F., Op. cit., p. 200 (corsivo mio).

<sup>24</sup> Vedi, a tal proposito Wittgenstein L., *Philosophische Untersuchungen*, cit. § 104: « Si predica della cosa ciò che insito nel modo di rappresentarla. Scambiamo la possibilità del confronto, che ci ha colpiti, per la percezione di uno stato di cose estremamente generale».

<sup>25</sup> Vedi Carapezza, M., *Segno e Simbolo in Wittgenstein*, Bonanno, Catania, 2006.

condividerne le emozioni e la loro elaborazione. Di contro, tali antenne, sono paradossalmente anche ciò che impedisce la perfetta comprensione della mente degli altri animali, o della loro effettiva *Umwelt*, spesso non “captabile” dai, né traducibile *tout court* nelle articolazioni biocognitive dell’uomo.

A rigore di chiarezza, è opportuno, a questo punto, ricapitolare i punti concettuali fin qui esposti. In base alla teoria di Uexküll ogni organismo percepisce il mondo *in funzione* del proprio corredo biocognitivo. Se innestiamo tale assunto alla tesi wittgensteiniana ne segue il guadagno della seguente conclusione: del mondo, ciò che ogni organismo “capta”, in base allo specie-specifico apparato bio-percettivo, sono le relazioni fra gli oggetti. La percezione del mondo *in* (attraverso) relazioni fra le parti sembra, in ultima analisi, un elemento chiave, comune alle diverse specie viventi, e nel caso specifico degli uomini essa si incarna nei nessi sintattico-proposizionali, determinando al contempo la specie-specificità della *Lebenswelt* umana, e la impossibilità cognitiva di accedere pienamente alla mente (quindi alla *Umwelt*) degli altri animali.

Quanto dichiarato a proposito della specificità dell’accesso epistemico al mondo da parte degli umani, dunque, vale *mutatis mutandis* per tutte le specie animali. È a nostro avviso interessante leggere sotto tale luce le seguenti proposizioni del *Tractatus*:

Che gli elementi dell'immagine siano in una determinata relazione l'uno con l'altro rappresenta che le cose sono in questa relazione l'una con l'altra. Questa connessione degli elementi dell'immagine io la chiamo struttura dell'immagine; la possibilità di questa struttura io la chiamo forma di raffigurazione dell'immagine.

La forma di raffigurazione è la possibilità che le cose siano l'una con l'altra nella stessa relazione che gli elementi dell'immagine.

È *così* che l'immagine è connessa con la realtà; giunge ad essa.

Queste coordinazioni sono quasi le antenne degli elementi dell'immagine, con le quali l'immagine tocca la realtà.<sup>26</sup>

In tale direzione tematica si colloca la presente trattazione, il cui obiettivo è quello di approfondire l'indagine intorno all'esistenza di un nucleo essenziale della facoltà di linguaggio, che, condiviso dalle diverse specie animali, va al di là delle loro differenze specifiche.

Definita, dunque, la sintassi come la capacità di creare e percepire connessioni cosali *attraverso* il linguaggio tipico della propria specie di appartenenza, resta da capire come queste abilità vengano tradotte nella effettiva percezione bio-cognitiva della *Umwelt*. Verosimilmente, non saremo mai in grado di approdare a tale conoscenza. Tuttavia, pur non abilitati a sapere come le altre specie *vedono* esattamente la realtà, potremmo risalire - attraverso quello che poi è l'unico paradigma possibile da cui ci è dato di indagare la mente degli animali, ovvero *attraverso* il nostro stesso linguaggio – alla facoltà delle specie di animali qui prese in esame di produrre e riconoscere taluni tipi di connessioni sintattico-logiche,

---

<sup>26</sup> Wittgenstein, L., *Tractatus logico-philosophicus*, cit., proposizioni. 2.15, 2.151, 2.1511, 2.1515.



e del modo in cui queste si riversano nel dominio logico, sociale e pragmatico. Troviamo, infatti, che l'analisi della "capacità di sintassi" nelle altre specie sia il luogo più appropriato in cui guardare alle dinamiche evolutive del linguaggio proposizionale, in grado di gettar luce sui tratti cognitivi evolutivamente condivisi con le altre specie e, al contempo, sul *quid bio-cognitivo* che determina la specie-specificità dell'uomo, ovvero del suo linguaggio. Una specificità di cui tenterò di enucleare taluni tratti costitutivi, che emergerà dall'analisi comparativa delle competenze sintattiche nelle diverse specie.

## **2. Canti, vocalizzi, e proposizioni: similarità oltre le differenze**

Gli umani sono l'unica specie animale dotata di linguaggio proposizionale? La risposta, quasi immediata ed intuitiva che crediamo chiunque sia incline a dare sia: sì. Eppure, affermare che gli esseri umani siano gli *unici* animali a vantare la forma proposizionale del linguaggio presuppone la conoscenza delle abilità cognitive che caratterizzano la mente degli animali *non* umani. Crediamo, infatti, che al fine di guadagnare chiarezza sui tratti che differenziano gli esseri umani dalle altre specie (o, in altri termini, la specificità al di là delle differenze), sia necessario effettuare una dettagliata analisi comparativa in grado di gettar luce sui tratti distintivi dei sistemi cognitivi di più specie animali.

In particolare, scandaglierò la questione dell'unicità del linguaggio umano alla luce di un suo tratto intrinsecamente costitutivo, la "sintassi", che, come vedremo, seppure in forme meno complesse, caratterizza il sistema cognitivo tanto delle scimmie, quanto degli uccelli canori. Assumiamo, infatti, che alcune caratteristiche che venano il sistema cognitivo dell'uomo nel suo complesso, siano isolatamente presenti in altri animali, laddove si sarebbero evolute in forme semplici, e per funzioni diverse da quelle assolve dall'articolazione verbale del linguaggio. Aspetti che, come vedremo, possono chiarire le origini evolutive del linguaggio umano.

Parlare di sintassi in riferimento ai sistemi di comunicazione animale, può generare, però, fraintendimenti e confusione concettuale. Per tale ragione, vorrei procedere con cautela metodologica, specificando la definizione operativa del termine. Adotto il termine "sintassi" in riferimento all'accezione greca di "*syntaxis*", composto da *syn* (con, insieme) – e *taxis* (ordine, connessione, in base a specifiche regole strutturali). Alla luce di tale specifico significato del termine, obiettivo del presente capitolo sarà di intercettare fra le espressioni vocali dei primati non umani e degli uccelli canori, meccanismi sintattici operanti a livello fonico, o sul piano del significato. La domanda di fondo è se sia possibile

identificare - in tali forme di sintassi - analogie (o omologie) evolutive della sintassi del linguaggio umano.

Invero, lo studio della sintassi è oggetto di un dibattito che specialmente negli ultimi decenni stimola la realizzazione di diversi studi sperimentali applicati in più domini sensoriali (auditivo, visivo, visuo-motorio). Ai fini del nostro studio, sono assai rilevanti le ricerche condotte sul dominio auditivo, in grado di rivelarci specificatamente come le abilità combinatorie delle specie non umane in questione si traducono e trovano applicazione nella dimensione fonico-vocale<sup>27</sup>. È bene precisare che tale obiettivo teorico trova una duplice presupposizione implicita che è necessario tematizzare, prima di mettere a fuoco lo studio comparativo dei sistemi linguistico-cognitivi degli infanti umani (ovvero di individui *pre-verbali*), e di animali *non-verbali*. La presupposizione in questione inerisce - da una parte, la caratterizzazione geneticamente determinata delle abilità di sintassi negli umani, ereditata filogeneticamente – dall'altra l'idea che essa si sia evoluta a fronte di pressioni ambientali specifiche. Tengo a sottolineare, per amor di chiarezza espositiva, che, nonostante riconosca la presenza di determinati meccanismi e processi senz'altro ereditati per via

---

<sup>27</sup> È opportuno precisare che la messa a fuoco della nostra analisi su tale modalità sensoriale deriva dalla osservazione del fatto che questa sia il *modus* principale in cui il linguaggio verbale viene espresso; tuttavia, tale considerazione non esclude la possibilità che le abilità sintattiche si siano evolute su un piano dominio-generale, in cui probabilmente si riversano tuttora, caratterizzando la precipua *Gestalt* degli specie-specifici sistemi di cognizione e azione nel mondo - né il riconoscimento della esistenza di un cardinale rapporto di interdipendenza funzionale tra i domini vocale e gestuale nel proferimento delle espressioni verbali umane.

genetica, troviamo che l'apporto dell'ambiente esterno sia altrettanto (se non maggiormente) determinante nello specificare la diversità specie-specifica - oltre che individuale - delle espressioni.

## 2.1 L'ipotesi innatista

The grammar of a language can be viewed as a theory of the structure of this language. Any scientific theory is based on certain finite set of observations, and, by establishing general laws stated in terms of certain hypothetical constructs, it attempts to account for these observations, to show how they are interrelated, and to predict an indefinite number of new phenomena.<sup>28</sup>

In tale passo, tratto da un articolo chiave del lavoro di ricerca condotto dal linguista Noam Chomsky, lo studioso presenta una sorta di metodologia di studio del linguaggio, affermando che alla stregua di ogni altra indagine scientifica, esso deve avvalersi di un insieme finito di leggi o assiomi, dalle quali poter derivare un insieme più ampio di osservazioni tra loro interconnesse. Al fine di enuclearne le leggi di funzionamento, è necessario, dunque, analizzare il linguaggio alla stregua di ogni altro oggetto di analisi delle indagini scientifiche, ovvero a partire dai suoi “atomi” costitutivi - i sintagmi (dal greco *syn tassein*) - e dalle regole che ne governano la combinazione entro le proposizioni. Vediamo, dunque, da

---

<sup>28</sup> Chomsky, N., “Three models for the description of language”, *Information Theory*, 1956, p. 113.

vicino la struttura interna di un sintagma. Esso si compone dei seguenti elementi:

- una “testa” (T), che è ciò che dà il nome al sintagma; ad esempio, nel caso in cui la testa è un nome, il sintagma sarà di tipo nominale.
- uno specificatore (S),
- un complemento (C)<sup>29</sup>.

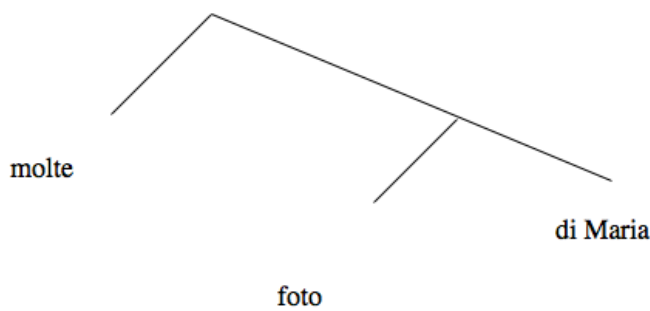
Una delle scoperte più intriganti della linguistica moderna è che i sintagmi, in tutte le lingue del mondo, presentano effettivamente un ordine interno, ovvero una struttura *asimmetrica*, in quanto le sottounità che lo costituiscono stanno su ben determinati livelli gerarchici<sup>30</sup>. È possibile dare una rappresentazione visiva, geometrica dell’asimmetria dei sintagmi, per mezzo di un diagramma ad alberi. Andrea Moro ne spiega molto bene la struttura:

---

<sup>29</sup> Vedi Chomsky, N., *Rules and Representations*, Columbia University Press, 1980; trad. it. di G. Gallo, *Regole e rappresentazioni*, Baldini Castoldi Dalai, Milano, 2008.

<sup>30</sup> Ciò si traduce nella impossibilità di sostituire alcuni elementi con un pronome, o di spostare a caso i termini all’interno di una frase quando, ad esempio, la si deve rendere in forma interrogativa (movimento sintattico).

Il diagramma ad alberi non è altro che una rappresentazione «bidimensionale» dei legami che sono contenuti nella linea continua della sequenza delle parole, dove per «bidimensionale» s'intende la rappresentazione gerarchica degli elementi. Per esempio, per il caso di molte foto di Maria possiamo avere una rappresentazione bidimensionale come nel grafico che segue, dove i rapporti tra specificatore [molte – specifica la quantità espressa dalla sottounità sintagmatica foto di Maria], testa e complemento sono resi espliciti in modo graficamente evidente, nel senso che la testa foto e il complemento di Maria stanno allo stesso livello (tecnicamente si dice che sono «adiacenti») e sono direttamente congiunti da due «rami» dell'albero sintattico (e costituiscono una sottounità), mentre lo specificatore molte è adiacente al «nodo» dal quale si dipartono i due rami<sup>31</sup>:

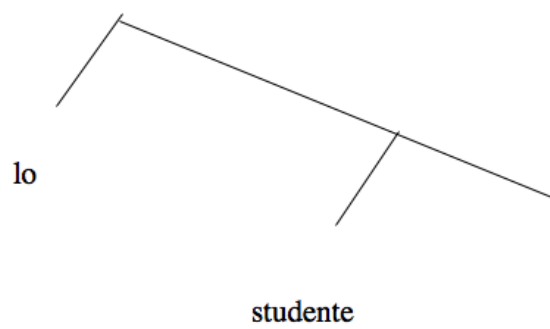


In altri termini, lo specificatore “molte” si trova in una posizione gerarchicamente superiore rispetto alla sottounità “foto di Maria”. In questo senso possiamo affermare che è *prominente* rispetto ad essa<sup>32</sup>. Ora, il *punctum saliens* è che all’interno dei sintagmi, è altresì possibile individuare delle sottounità che presentano la medesima struttura

<sup>31</sup> Moro, A., *I confini di Babele*, Longanesi, Milano, 2006, p. 87.

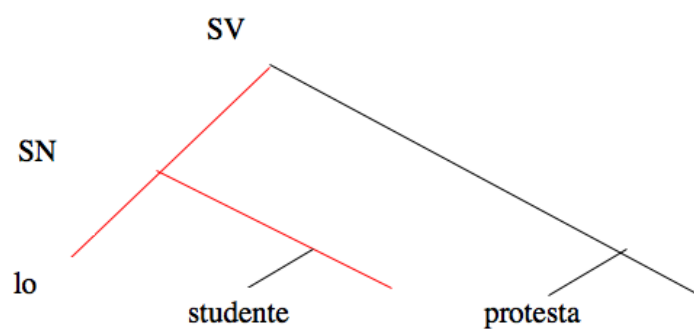
<sup>32</sup> La geometria interna dei sintagmi può esser individuata per mezzo di test specifici, detti “di costituenza”. Un esempio ne è la possibilità di sostituire le sottounità con un elemento “proforma” come il pronome complemento “ne”. Infatti solo la testa e ciò che la segue possono essere sostituiti da tale pronome, e non la testa assieme allo specificatore. Ad esempio, data la frase Paolo ha visto molte foto di Maria, è corretto dire “Paolo *ne* ha visto molte” (dove il *ne* sta per la testa + il complemento *foto di Maria*), piuttosto che “Paolo *ne* ha visto di Maria” (dove *ne* sta per lo specificatore + la testa *molte foto*). In altri termini, ciò vuol dire che la coppia testa + complemento stanno ad un livello gerarchico inferiore rispetto allo specificatore “molte”, il quale non è sostituibile da alcun elemento proforma.

asimmetrica, la quale occorre ricorsivamente. Ad esempio, il sintagma verbale (SV) “Lo studente protesta”, presenta al suo interno il sintagma nominale (SN) “lo studente”, la cui struttura è così rappresentabile:



Ne segue che la struttura del sintagma verbale “Lo studente protesta”

sarà:



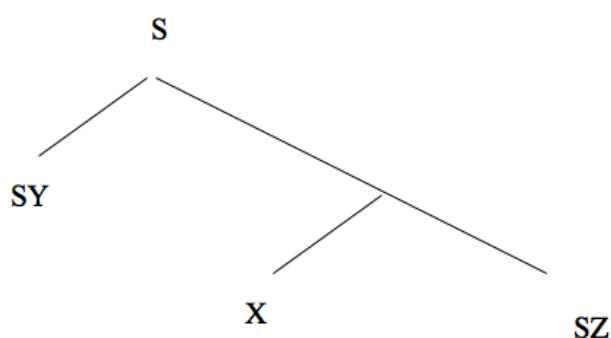


Avremmo potuto saturare ricorsivamente la posizione del C del sintagma nominale “Lo studente”, con un aggettivo, ad esempio “indignato”, e a sua volta, integrare a questo (considerato come la testa di un nuovo sintagma) un nuovo C, ad esempio, il motivo della sua indignazione. Ma per non complicare molto l’esposizione della teoria chomskiana, fermiamoci alla proposizione semplice “Lo studente protesta”. Come si può osservare abbiamo qui un caso di annidamento ricorsivo, in quanto una medesima struttura gerarchia e asimmetrica, ovvero quella del SN, si trova all’interno di una struttura (SV) del medesimo tipo. Moro ne parla così:

Generalizzando, indicando cioè con X una testa qualsiasi (N [nome], V [verbo], P [preposizione] o A [aggettivo], varrà per tutti il seguente schema asimmetrico, dove lo specificatore e il complemento, che qui sono indicati rispettivamente con «SY» e «SZ» possono al loro volta avere la struttura asimmetrica del sintagma nel quale sono contenuti [...]. *In questo senso si può concludere che i sintagmi realizzano una struttura ricorsiva e omogenea. [...]* Fin qui abbiamo osservato la struttura dei «mattoni» che compongono una frase. *Una delle scoperte più importanti della teoria della sintassi è che anche una frase intera ha una struttura sintagmatica dello stesso tipo di quella illustrata qua.*<sup>33</sup>

---

<sup>33</sup> Moro, A., Op. cit., pp. 89-90 (corsivo mio). Vedi Chomsky, N., *Language and problems of knowledge. The managua lectures*, The MIT press, Cambridge, 1988, trad. it. di C. Donati e A. Moro, *Linguaggio e problemi della conoscenza*, Il Mulino, Bologna, 1998, pp. 59-61.



In altri termini, così come nel caso della funzione successore, anche in linguistica occorre il fenomeno della ricorsività che, a partire dalle sottounità di un sintagma dà luogo alla configurazione degli enunciati per mezzo della iterazione (ricorsiva) dello schema asimmetrico “specificatore-testa-complemento” (da qui in poi S-TC), stabilendo diversi livelli di prominenza. Come afferma lo stesso Chomsky, «le regole del linguaggio non prendono in esame il semplice ordine lineare ma sono *dipendenti dalla struttura* [...]. Le regole operano su espressioni alle quali viene assegnata una certa struttura nei termini di gerarchia di sintagmi di vario tipo»<sup>34</sup>. Per comprendere meglio tale affermazione, è opportuno far capo alla distinzione fra la struttura profonda e quella superficiale di un enunciato, di cui, in questa sede mi limito a fornire una spiegazione estremamente essenziale. La prima consiste nelle regole ricorsive di combinazione

---

<sup>34</sup> Chomsky, N., *Language and problems of knowledge. The managua lectures*, cit., p.41.

sintattica degli elementi all'interno della frase, regole che attribuiscono una posizione ben specificata ad ogni unità che vi occorre, dunque un ruolo logico a ciascuna di essa, che permette al parlante/fruitore di sapere, in base alla posizione del sintagma, “chi ha fatto cosa a chi”. È in base alle regole della struttura profonda, dunque, che i parlanti si orientano nella comprensione/produzione di enunciati, assegnando *automaticamente* ad ogni struttura sintattica una struttura ad alberi sempre basata sulla stessa organizzazione gerarchica asimmetrica del tipo S-TC.

La struttura superficiale (s), invece, consiste in un albero diverso, che consente di spostare i sintagmi, mantenendone tuttavia intatti i ruoli “profondi”, che a mo' di bussola orientano nella comprensione. Ad esempio, anziché dire: “Giovanni (S) diede alla persona che hai incontrato l'altra sera prima di andare al ristorante vicino al fiume (T) quel libro (C)”, rispettando l'ordine dello schema S-TC della grammatica generativa – in base alla struttura-s possiamo spostare il C (nel caso specifico, “quel libro”) nella posizione accanto allo S, rendendo in tal modo più immediatamente fruibile il senso della frase (Giovanni diede quel libro alla persona che hai incontrato l'altra sera prima di andare al ristorante vicino al fiume), o consentendo di agganciare all'elemento spostato all'inizio dell'enunciato dati sopraggiunti sul momento nella mente del parlante (Giovanni diede quel libro *di cui parlavano in televisione*, alla persona che hai incontrato

l'altra sera prima di andare al ristorante vicino al fiume)<sup>35</sup>. Come si può notare, a discapito del movimento sintattico, la comprensione del significato profondo dell'enunciato è agevolata dagli accordi, di genere, numero, o pronominali che occorrono al suo interno.

La cosiddetta “ipotesi genetica” di Chomsky, secondo cui la struttura computazionale del linguaggio è parte della dotazione biologica innata degli uomini è strettamente connessa all'idea dell'esistenza di una “grammatica universale” sottesa a tutte le lingue storico-naturali. Al fine di chiarire in cosa consista quest'ultima nozione, proviamo a vedere cosa accade se invertiamo l'ordine della testa e del complemento nel diagramma ad alberi. Da una frase dalla struttura S-TC come (1) “Maria ha detto che Gianni ha visto quella foto” verrebbe fuori (2) “Maria Gianni quella foto visto ha che detto ha”<sup>36</sup>, una frase che in italiano suona come un mero miscuglio insensato di parole. È divertente notare che, in realtà la frase (2) è perfettamente ordinata secondo lo schema sintattico delle frasi in giapponese o in miskito, che appunto è S-CT. Ciò dimostra che variazioni anche minime, come è il caso dell'ordine della testa e del complemento, hanno degli effetti molto evidenti. Tecnicamente, le parti variabili tra le sintassi di diverse lingue si chiamano “parametri”, mentre le parti invarianti del sistema delle frasi si definiscono “principi”. E su di essi si basa il

---

<sup>35</sup> Vedi *Ivi*, pp. 112-13.

<sup>36</sup> Per visualizzare i diagrammi ad alberi delle frasi date, vedi Moro A., *Op. cit.*, pp. 123-24.

modello detto, appunto, “modello a principi e parametri”<sup>37</sup> che domina il panorama degli attuali studi della linguistica moderna.

Oltre al parametro dato dall'ordine della testa e del complemento ne esistono altri. Questi dunque sono gli elementi che determinano la diversità fra le diverse lingue storico naturali, e per questo, diversamente da quanto accade per i principi, vengono *appresi* per addestramento. L'aspetto universale (non appreso, bensì, come vedremo biologicamente determinato) è il fatto che ogni sintagma è composto da specificatore, testa e complemento, e che lo specificatore è prominente sulla sequenza testa-complemento. In altri termini, lo schema S-TC/CT, passato alla letteratura con il nome di “principio di dipendenza dalla struttura” è costitutivo di tutte le lingue storico-naturali, e per dirla con Moro, costituisce una sorta di “trama nascosta” dietro la loro più immediata veste monodimensionale, data dalla sequenza lineare (visiva o sonora) delle parole. Secondo Chomsky,

Gli studi più affascinanti sulla struttura del linguaggio sono quelli che riguardano gli universali linguistici: cioè, i principi che sottostanno al linguaggio in generale, inteso come *necessità biologica* (e non logica). Data la ricchezza e la complessità del sistema grammaticale di un linguaggio umano e data l'uniformità della sua acquisizione sulla base di evidenze limitate e spesso corrotte, *non c'è dubbio che devono esistere principi universali altamente*

---

<sup>37</sup> Vedi Pinker, S., *The Language Instinct*, 1994, trad. it di G. Origgi, *L'istinto del linguaggio*, Arnoldo Mondadori, Milano, 1997, p. 102.

*restrittivi, che determinano la struttura generale di tutti i linguaggi umani e forse anche gran parte della sua struttura specifica.*<sup>38</sup>

Tutti gli individui umani dunque nel produrre le frasi, assegnano a ciascuna di esse la struttura gerarchica che abbiamo appena visto. Ciò avviene inconsciamente, ovvero “per necessità biologica”, in quanto esiste una sorta di schema, un programma geneticamente determinato che permette di derivare in linea di principio tutte le strutture delle lingue storico naturali possibili. Chomsky chiama tale schema “grammatica universale”:

La grammatica universale è un sistema che è geneticamente determinato allo stato iniziale, ed è specificato, irrobustito, articolato e raffinato nelle condizioni predisposte dall’esperienza, per produrre le particolari grammatiche rappresentate negli stati stabili raggiunti<sup>39</sup>

## **2.2 Un povero stimolo**

Un tema importante che converge nella direzione dell’ipotesi innatista, secondo cui la mente presenta dei tratti innati, ovvero delle articolazioni interne che guidano l’acquisizione di una lingua e la

---

<sup>38</sup> Chomsky, N., *Rules and Representations*, cit., p. 401 (corsivo mio).

<sup>39</sup> *Ivi*, p. 403.

produzione di enunciati, è il cosiddetto “argomento della povertà dello stimolo”. Pinker ne spiega bene l’origine:

Negli anni Cinquanta le scienze sociali erano dominate dal comportamentismo, la scuola di pensiero resa famosa da John Watson e B.F. Skinner. [...] Il comportamento era spiegato mediante poche leggi di apprendimento per stimolo-risposta che potevano essere studiate su topi che premono pulsanti e cani che sbavano a comando. Ma Chomsky portò l’attenzione su due fatti fondamentali del linguaggio. In primo luogo, ogni enunciato che una persona formula o comprende è una combinazione del tutto nuova di parole, che appare per la prima volta nella storia dell’universo. Di conseguenza, il linguaggio non può essere un repertorio di risposte; il cervello deve contenere una ricetta o un programma che può costruire un insieme illimitato di enunciati a partire da un insieme finito di parole. Questo programma deve essere definito una grammatica mentale.<sup>40</sup>

Il secondo punto su cui Chomsky indirizzò gli sguardi consiste nel fatto, evidente, che i bambini sviluppano tale grammatica con una naturalezza che non può non far risalire alla dotazione biologica di uno schema comune a tutte le lingue, che guida il neo-parlante nella produzione e nella interpretazione coerente di enunciati. Ciò è attestato dal fenomeno detto “ipercorrettismo”<sup>41</sup>, che si realizza in quei casi in cui i bambini applicano le regole della lingua anche nei casi in cui sono previste le eccezioni: “ho romputo la tazza”, “ho aprito la porta”, “uovi” e così via.

---

<sup>40</sup> Pinker, S., *The language instinct*, cit., p. 14.

<sup>41</sup> Vedi Ivi, pp. 263-69.

Nel quadro del paradigma innatista si colloca coerentemente anche l'esistenza di un periodo critico nei bambini (che va dalla nascita fino alla pubertà) oltre il quale diviene difficoltosa l'acquisizione di una lingua. Ciò attesta che esiste un dispositivo innato ed universale alla base della competenza linguistica dei parlanti, che, se non opportunamente stimolato entro un determinato periodo, incorre in una sorta di "atrofizzazione" irreversibile. Essa emerge, altresì, con evidenza in quegli adulti che si apprestano ad imparare una lingua diversa dalla propria<sup>42</sup>. In breve, possiamo trovare una sintesi essenziale ed incisiva della tesi di Chomsky nel seguente passo chiave:

Sembra che il bambino si accosti al compito di acquisire il linguaggio con un ricco schema concettuale già instaurato ed anche con un ricco sistema di assunti circa la struttura del suono e la struttura di enunciati più complessi. Questi costituiscono la parte della nostra conoscenza che «proviene dalla mano originale della natura», per usare i termini di Hume. Essi costituiscono una parte della dotazione biologica umana, che dev'essere risvegliata dall'esperienza ed affinata ed arricchita nel corso delle interazioni del bambino con il mondo umano e materiale.<sup>43</sup>

---

<sup>42</sup> Anche la fenomenologia della visione offre diversi elementi che validano l'ipotesi che la mente, lungi dall'essere una mera *tabula rasa* va sempre oltre l'informazione percettiva data. Ad esempio, l'ipotesi innatista è supportata in modo eclatante da alcuni disegni effettuati da ciechi congeniti, che dimostrano come, in effetti, la capacità di percezione/produzione di una rappresentazione, sia essa grafica o frasale, dipende da un insieme di strutture che va oltre i dispositivi deputati all'accesso sensoriale. Vedi, a tal proposito, Ferretti F., *Perché non siamo speciali*, Laterza, Roma-Bari, 2007, pp. 33-44 e Pinker S., *Op. cit.*, p. 189 e segg.

<sup>43</sup> Chomsky, N., *Language and problems of knowledge*, cit., p.30. È a mio avviso, essenziale notare come in questo passo Chomsky attribuisca un ruolo ben delineato e cruciale nel processo di acquisizione di una lingua. A tal proposito, vedi. *Ivi*, p. 42: «Il bambino che impara l'italiano o qualsiasi altra lingua umana sa, prima di ogni esperienza, che le regole dipenderanno dalla struttura. [...] Questa conoscenza fa parte della dotazione biologica del bambino, fa parte della facoltà del linguaggio. Essa costituisce una parte dell'equipaggiamento mentale con il quale il bambino affronta il mondo dell'esperienza» (corsivo mio).



Parte della nostra competenza linguistica deriva, dunque dalla “mano originale della natura”, in quanto trasmessa come dotazione biologica. Troviamo tale assunto molto vero, ma crediamo vada inserito in una visione più ampia che estenda il significato di “dotazione biologica” alla necessità incontrovertibile delle interazioni sociali e delle esperienze con e nell’ambiente circostante.

### **2.3 Grammatiche per lo studio della sintassi**

L’approccio comparativo allo studio della sintassi ha interessato accesi dibattiti teorici nelle ultime decadi, e trova tuttora ampia risonanza nell’ambito dello studio dello sviluppo del linguaggio tanto sul piano ontogenetico, che su quello filogenetico. Al fine di esplorare i vari livelli di competenza sintattica in diverse specie animali - e dunque, trovare quei tratti condivisi in grado di gettare luce sulle tappe e sulle pressioni evolutive che hanno guidato l’evoluzione del linguaggio nella sua odierna forma proposizionale – diversi studiosi hanno aderito al cosiddetto paradigma dell’*artificial grammar learning* (apprendimento di una grammatica artificiale). Questo consiste nel derivare un linguaggio da regole logico-formali, ovvero dalle regole di una grammatica formale sulla cui base costruire dei mini-linguaggi artificiali adatti ad essere

implementati in test sperimentali. Nello specifico, l'idea è di esporre ad una medesima classe di stimoli - con diversi livelli di complessità interna - soggetti appartenenti a diverse specie animali, al fine di confrontarne le abilità di apprendimento. Così, avvalendosi di uno studio comparativo sulla capacità di acquisizione di linguaggi artificiali (ovvero di linguaggi che presentano delle proprietà particolari dalla cui formalizzazione in termini matematici vengono derivati degli insiemi di frasi che presentano specifiche proprietà *sintattiche*<sup>44</sup>), è possibile individuare quei tratti sintattici attraverso, e da cui ha preso piede l'evoluzione del linguaggio proposizionale. Tali studi si suddividono generalmente in una fase di familiarizzazione, in cui i soggetti vengono esposti (abituati) ad un certo tipo di stimoli che presentano dei pattern strutturali, o delle regolarità statistiche - e una fase di test, in cui si esamina la capacità dei soggetti di riconoscere i pattern strutturali di cui tali stimoli si compongono, o di applicarli su stimoli diversi. La procedura metodologica adottata viene descritta nei seguenti termini in un articolo pubblicato di recente da O'Donnell, Fitch e Hauser:

In recent years, there has been much renewed interest in experimentation with a variety of methods using artificial languages [...]. An artificial language is a set of sentences, often constructed from nonsense syllables, which has some property of interest to researchers. There have been several different

---

<sup>44</sup> O'Donnell, T. J., Hauser, M. D., & Fitch, W. T., *Using mathematical models of language experimentally*, "Trends in Cognitive Sciences", 9, 2005, pp. 284-289.

experimental paradigms employed using artificial languages, but in general they use the following procedure. First subjects are exposed to some artificial language that is derived from some rule or principle. Next, subjects are given some task that requires that they discriminate between exemplars that are either consistent or inconsistent with the language they were exposed to. The use of artificial languages has the advantage of allowing the researcher to control carefully for the structure and information present in the language and focus specifically on phenomena of interest. These advantages are essential for testing human infants and non-human animals. Formal languages can be used as models of the artificial languages in our experiments. Because formal languages can also be used as models of natural language this allows us to connect experiments with linguistic theory<sup>45</sup>.

Al fine di isolare l'aspetto meramente combinatorio della struttura da qualsiasi altra caratteristica in grado di veicolare informazioni, spesso le frasi utilizzate negli esperimenti come stimoli su cui testare le abilità cognitive sono costituite da stringhe di suoni prive di senso, o proferite da una voce artificiale per isolare ogni colore prosodico ed emozionale. Di fatti, lo scopo cardine alla base di tale paradigma è l'individuazione delle competenze di natura esclusivamente *sintattico-formale*, deputata ad indicare il confine cognitivo tra i sistemi cognitivi delle specie animali non umani e gli esseri umani. In particolare, in tale prospettiva di studio si è posta l'attenzione sulla capacità di apprendere due specifici pattern in grado di rivelarci il livello di competenza sintattica delle varie specie testate: gli schemi che presentano proprietà distribuzionali statisticamente regolate, e due grammatiche della cosiddetta gerarchia di Chomsky. Nei prossimi paragrafi prenderemo in esame la natura di tali "grammatiche

---

<sup>45</sup> *Ivi*, p. 287.

artificiali”, e le implicazioni teoriche dei risultati degli studi sperimentali su cui sono stati implementati, tenendo presente che una risposta esaustiva alla domanda di fondo di tali ricerche non può prescindere dalla distinzione dei seguenti livelli di competenza sintattica tanto negli uccelli canori, quanto nelle scimmie:

- l’abilità di *percepire* e riconoscere la complessità sintattica di una data connessione di suoni vocali,
- l’abilità di *comprendere* il significato di proferimenti sintatticamente articolati (abilità – è bene precisarlo – da distinguere rispetto alla mera familiarità con gli effetti comportamentali derivanti del proferimento di vocalizzi o canti complessi),
- l’abilità di *produrre* spontaneamente delle espressioni vocali sintatticamente articolate.

### **3. Percezione di pattern fono-sintattici: apprendimento statistico e acquisizione di regole**

#### **3.1 L'apprendimento statistico negli infanti pre-verbali**

L'ambiente circostante ci circonda di una quantità elevatissima di stimoli. La velocità e l'accuratezza con cui una specie discerne quelle informazioni rilevanti ai fini della propria sopravvivenza possono essere il fine implicito dello sviluppo (tanto in chiave ontogenetica, quanto filogenetica) dei meccanismi cognitivi dedicati alla percezione del mondo esterno. In alcune specie, l'approccio all'ambiente esterno è totalmente vincolato ad abilità e meccanismi determinati geneticamente. In altre specie, l'esperienza sensibile dell'ambiente circostante stesso gioca un ruolo fondamentale nello sviluppo della fisionomia di quei meccanismi e connessioni neurali geneticamente favoriti. In altri termini ha luogo una

sorta di interazione bio-logica che vede ciascun termine della relazione retroagire sull'altro. Tale questione è centrale nell'ambito della tematizzazione dell'acquisizione ontogenetica del linguaggio. In effetti, la velocità con cui gli infanti imparano ad “orientarsi nella lingua”, e l'effettiva discrepanza tra gli stimoli linguistici cui sono esposti e quelli che poi riprodurranno in maniera del tutto creativa – lasciano pochi dubbi sulla natura innata dei processi di acquisizione, comprensione, e produzione del linguaggio. Tuttavia, sebbene molti studiosi enfatizzano tale aspetto, altri evidenziano il ruolo della esperienza e dell'esposizione sensoriale nel modulare l'attività neuronale. Fra questi ultimi, è doveroso menzionare Jenny Saffran, ed in particolare lo studio che il suo *team* di ricerca ha condotto nel 1996<sup>46</sup> su un gruppo di bambini dell'età di otto mesi. L'assunto di fondo di tale ricerca è che le informazioni sensoriali cui gli infanti sono esposti - e nello specifico le sequenze di suoni - presentano delle regolarità combinatorie quantificabili in termini di distribuzioni statistiche, e che la probabilità che una sillaba ne segua un'altra è più alta laddove queste occorrono all'interno di una stessa parola piuttosto che in due parole distinte. Date, ad esempio, le parole “buona” e “torta”, la probabilità che la sillaba “tor-” sia seguita da “ta” sarà più elevata rispetto alla probabilità che “na” sia seguita da “tor”. Secondo il paradigma detto

---

<sup>46</sup> Saffran, J. R., Aslin, R. N., & Newport, E. L., *Statistical learning by 8-month-old infants*, “Science”, 274, 1996, pp. 1926–1928.

dell'apprendimento statistico del linguaggio (in cui tale ricerca si iscrive), ciò sarebbe un tipo di informazione essenziale nel segmentare le parole all'interno di un flusso di discorso che, come spesso avviene nelle conversazioni ordinarie (o come si può notare ascoltando un discorso proferito in una lingua che non conosciamo), spesso non presenta pause indicative tra un termine e l'altro:

One important source of information that can, in principle, define word boundaries in any natural language is the statistical information contained in sequences of sounds. Over a corpus of speech there are measurable statistical regularities that distinguish recurring sound sequences that comprise words from the more accidental sound sequences that occur across word boundaries.[...] We asked whether 8-month-old infants can extract information about word boundaries solely on the basis of the sequential statistics of concatenated speech.<sup>47</sup>

Per testare tale ipotesi, Saffran *et al.* hanno inventato un mini-linguaggio del tutto privo di senso, le cui unità costitutive (quattro parole trisillabiche) vennero costruite concatenando un insieme di sillabe secondo specifiche regole di sequenziamento statistico. È opportuno sottolineare che secondo gli studiosi, tale abilità è centrale nell'acquisizione ontogenetica del linguaggio negli umani, e specificamente, nella *pertinentizzazione* delle parole, ovvero il processo che permette agli infanti di determinare quali sequenze di suoni - all'interno di un flusso continuo di discorso - costituiscano parole ben formate. Ad esempio, un modo per

---

<sup>47</sup> Ivi, p. 1927.

identificare le parole occorrenti in un discorso pronunciato in una lingua nuova è *calcolare* le probabilità condizionali tra le sequenze di sillabe, altrimenti dette “probabilità di transizione” (frequenza di XY/ frequenza di X, dove X e Y sono sillabe). La loro ipotesi di ricerca è che quelle sequenze di suoni che occorrono con maggiore frequenza saranno classificate come parole. Immaginiamo di avere un mini linguaggio di 4 parole, una delle quali è “nice” e le altre tre: “paper”, “children”, e “guitar”. La probabilità che la sillaba *ce* sia seguita da *pa*, *chi*, o *gui* è di 1 su 4, mentre la probabilità che la sillaba *ni* sia seguita da *ce*, o che *chil* sia seguito da *dren* - è di 1 su 1. Gli autori hanno applicato tale formula computazionale per la creazione di un mini linguaggio artificiale di cui alcune unità costitutive erano ad esempio, *pabiku*, *tibudo*, *golatu*, *daropi* – che come si può notare, non presentavo alcun valore semantico. La fase di familiarizzazione a tale linguaggio prevedeva l’esposizione ad un flusso vocale continuo ed ininterrotto per due minuti, che potrebbe essere rappresentato graficamente nel seguente modo: *tibudogolatupabikudaropi golatutibudo...* In base alle proprietà transizionali delle unità (1.0 tra le sillabe di una medesima unità, 0.3 tra sillabe di unità diverse), è possibile segmentare - nel flusso fonico continuo di familiarizzazione - le parole ben formate, discriminandole da quelle che gli autori chiamano *part-words* (“parole parziali”). Queste ultime consistono in unità costituite da due



sillabe proprie di una determinata parola e la terza appartenente ad un'altra parola del mini-linguaggio, e collocate in una posizione adiacente nel flusso del discorso. Ne sono un esempio: *tudaro, pigola, bikuti, etc.*

L'esperimento prevedeva una fase di familiarizzazione, in cui i bambini vennero esposti auditivamente ad un flusso uditivo continuo che consisteva nella ripetizione random delle quattro parole per un tempo totale di due minuti; in seguito, utilizzando il paradigma della violazione delle aspettative, i ricercatori hanno poi sottoposto i soggetti a due tipi di test diversi: uno in cui se ne controllava l'abilità di distinguere le parole del linguaggio con cui avevano familiarizzato da parole che non ne facevano parte – ed uno in cui si testava, invece, la loro abilità di distinguere tra parole “ben formate” e “part-words”. Da tale studio è emerso che i neonati di otto mesi sono effettivamente in grado di sfruttare le informazioni relative alla distribuzione statistica delle sillabe per individuare le parole all'interno di un flusso continuo di discorso. A tal proposito, è importante sottolineare che il linguaggio utilizzato nella fase di familiarizzazione era stato realizzato artificialmente, al fine di poter controllare l'assenza di ogni indizio (diverso dalle regolarità statistiche) potenzialmente utile alla segmentazione delle parole. Infatti il flusso del mini linguaggio era: a) riprodotto da una voce artificiale, b) continuo, ovvero senza pause di silenzio tra una parola e l'altra, c) e mancava di ogni inflessione prosodica

in grado di segnalare l'inizio e la fine delle parole all'interno del flusso stesso. Così i ricercatori commentano i risultati finali della loro ricerca:

Our results raise the intriguing possibility that infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language.<sup>48</sup>

Gli autori di tale studio - che trova ancora oggi ampia risonanza nel dibattito sulla acquisizione dell'abilità del linguaggio su scala tanto ontogenetica, quanto filogenetica – vedono i risultati del loro studio come una conferma evidente della priorità della dimensione sensoriale nell'acquisizione del linguaggio.

A tale proposito, troviamo condivisibile la tesi del ruolo imprescindibile dell'esperienza nello sviluppo delle abilità cognitive in generale, e linguistiche in particolare. Tuttavia, crediamo che la durata minima della fase di familiarizzazione, e la neutralità emozionale del segnale siano da annoverare fra gli argomenti a favore della tesi opposta, in quanto stimoli “poveri”<sup>49</sup>. Dal mio punto di vista, lo studio di Saffran *et al.* conferma la presenza di un meccanismo cognitivo adatto a percepire, a mo' di antenna, delle regolarità distribuzionali effettivamente presenti nel flusso

---

<sup>48</sup> *Ivi*, p. 1928.

<sup>49</sup> Cfr. Yang, C., *Universal Grammar, statistics or both?*, “Trends in Cognitive Sciences”, 8, 2004, pp. 451-456.

continuo di stringhe, e ciò dimostra come l'acquisizione di un linguaggio sia realmente funzionale anche alla natura strutturale degli stimoli; in tale studio, infatti, viene taciuto un aspetto a nostro avviso nettamente fondamentale dell'esperienza, ovvero la dimensione emozionale e di reciprocità comunicativa. Quest'ultima permea di sé ogni interazione sociale (sede naturale della acquisizione dei proferimenti linguistici), assumendo un ruolo cruciale ed impenscindibile nella decodificazione e nella costruzione del senso di un qualsiasi messaggio.

A discapito di tale importante omissione, lo studio di Saffran *et al.* rimane una ricerca chiave in quest'ambito di indagini, gettando luce sulla sorprendente capacità biologica degli infanti umani di discernere gli aspetti in qualche modo "regolari" dei dati sensoriali stessi, anche dopo un'esposizione dalla durata brevissima. Date tali considerazioni, è opportuno a questo punto chiedersi se l'abilità di individuare delle regolarità distribuzionali nel dato sensoriale sia meramente connesso alla percezione di suoni linguistici (ovvero unità che suonano come parole prive di senso), o non si estenda invece a stimoli di natura diversa, rappresentando dunque un'abilità dominio-generale che si espleta in differenti sfere sensori-motorie.

In tale direzione, diverse ricerche hanno dimostrato che l'abilità di apprendimento statistico si estende anche a stimoli che non suonano come

appartenenti ad una lingua storico-naturale. Ad esempio, la stessa Jenny Saffran, con un gruppo di collaboratori, ha condotto un ulteriore studio<sup>50</sup> sulla scia della ricerca appena descritta. Il fine era di scoprire se la capacità di utilizzare le regolarità statistiche delle sequenze di suoni per individuare unità discrete (simili a parole) all'interno di un flusso continuo - possa sottendere anche l'abilità di segmentare sequenze discrete costituite da note musicali. L'esperimento condotto in tale ricerca era identico a quello della ricerca sui segmenti linguistici, ma con la sola differenza che ogni sillaba del mini-linguaggio venne sostituita con una nota. Gli studiosi ne descrivono i risultati nei seguenti termini:

In short, our results suggest that the same statistical learning mechanism can operate on both linguistic and non-linguistic stimuli. Thus at least part of the machinery involved in natural language learning may be shared with other pattern learning processes.<sup>51</sup>

Ulteriori studi condotti sul dominio visuale giungono alle medesime conclusioni. Ad esempio, da una ricerca condotta da Fiser e Aslin<sup>52</sup> sull'apprendimento statistico di combinazioni di elementi visivi, emerge che i bambini di nove mesi sono in grado di estrarre regolarità dalla presentazione di stimoli consistenti in scene visive:

---

<sup>50</sup> Saffran, J. R., Johnson, E., Aslin, R. N., *Statistical learning of tone sequences by human infants and adults*, "Cognition", 70, pp. 27-52.

<sup>51</sup> *Ivi*, p. 47.

<sup>52</sup> Fiser, J., Aslin, R.N., *Statistical learning of new visual feature combinations by infants*, "Proceedings of the National Academy of Sciences", 99, 2002, pp. 15822-15826.

The ability of humans to recognize a nearly unlimited number of unique visual objects must be based on a robust and efficient learning mechanism that extracts complex visual features from the environment. To determine whether statistically optimal representations of scenes are formed during early development, we used a habituation paradigm with 9-month-old infants and found that, by mere observation of multielement scenes, they become sensitive to the underlying statistical structure of those scenes. After exposure to a large number of scenes, infants paid more attention not only to element pairs that cooccurred more often as embedded elements in the scenes than other pairs, but also to pairs that had higher predictability (conditional probability) between the elements of the pair.<sup>53</sup>

Posta dunque la dominio-generalità dei meccanismi di implicita analisi sintattica che operano nella categorizzazione degli input sensoriali in unità discrete, vale la pena di chiedersi se - oltre a quelli adottati da Saffran *et al.* (1996) - vi sono altri tipi di regolarità combinatorie degli stimoli stessi che ne rendono possibile la segmentazione in unità discrete.

A tal proposito, Newport e Aslin<sup>54</sup> hanno condotto uno studio su adulti umani, che hanno poi replicato - come vedremo nel prossimo paragrafo - su una specie di primati non umani. La domanda teorica che fonda tale studio è se la capacità di riconoscere regolarità sintattiche (statisticamente regolate) ha luogo anche laddove gli elementi costitutivi di un'unità discreta - isolabile per mezzo di statistiche distribuzionali - non sono immediatamente adiacenti fra loro. In altre parole, gli autori si

---

<sup>53</sup> Ivi, p. 15822. Cfr. Brady, T., Oliva, A., *Statistical learning using real-world scenes*, "Psychological science", 19, 2008, pp. 678- 685.

<sup>54</sup> Newport, E. L., Aslin, R. N., *Learning at a distance I. Statistical learning of non-adjacent dependencies*, "Cognitive psychology", 48, 2004, pp. 127-162. Cfr. Van Heugten, M., Johnson, E. K., *Linking infants' distributional learning abilities to natural language acquisition*, "Journal of Memory and Language", 63, 2010, pp. 197-209.

chiedono se i meccanismi di apprendimento statistico operano anche su stimoli temporalmente non contigui. Al fine di trovare una risposta a tale quesito, gli autori hanno generato due mini-linguaggi artificiali: uno in cui le probabilità transitive operavano tra sillabe, ed uno in cui operavano al livello dei segmenti fonetici; in entrambi i linguaggi, le unità discrete erano costituite da tre elementi, ed erano costruiti in modo che la probabilità di transizione tra il primo elemento (la sillaba in un linguaggio, il segmento nell'altro) e l'ultimo fosse 1.0, ovvero pari al 100%, mentre tra il primo e il secondo, il secondo e il terzo, e il terzo e l'elemento della unità successiva le probabilità di transizione erano inferiori (tra .20 e .25). Curiosamente, i due tipi di linguaggi hanno sortito risultati diversi:

Across experiments and experimental conditions, different types of non-adjacent regularities are compared. Our results across these studies clearly show that adult learners are highly selective in the types of non-adjacent regularities they are readily able to compute. Non-adjacent *syllable* regularities are very difficult to acquire; while they can be performed in other studies employing simpler learning paradigms, they are not acquired in our studies. In contrast, non-adjacent *segment* regularities—patterns among consonants, skipping over vowels, and also patterns among vowels, skipping over consonants—are extremely easy to acquire and are readily learned in our paradigm.<sup>55</sup>

---

<sup>55</sup> *Ivi*, p. 154 (corsivi miei).

Table 1  
Design of non-adjacent syllable language used in Experiment 1

1st-3rd Syllable word-frames	2nd Syllables	Words
ba_te	di	ba di te ba ku te
gu_do	ku	ba to te ba pa te
pi_ra	to	gu di do gu ku do
ke_du	pa	gu to do gu pa do
lo_ki		pi di ra pi ku ra pi to ra pi pa ra
		ke di du ke ku du ke to du ke pa du
		lo di ki lo ku ki lo to ki lo pa ki

Table 3  
Design of two non-adjacent segment (consonant) languages used in Experiment 2

Consonant-frames	Vowel-fillers
<i>Language A</i>	
p_g_t_	[_a] [_i] [_ae]
d_k_b_	[_o] [_u] [_e]
<i>Language B</i>	
t_d_k_	[_ae] [_a] [_i]
b_p_g_	[_e] [_o] [_u]

Table 4  
Test words and test partwords used in Experiment 2

	Words	Partwords
<i>Non-adjacent syllable languages</i>		
Language A	di ki tae di gu tae po ki ga po gu ga ke ki bu ke gu bu	tae po ki ga ke ki bu di gu gu tae di ki ga ke gu bu po
Language B	bae pa ku bae be ku te pa da te be da go pa pi go be pi	ku te be da go be pi bae pa be ku go pa da bae pa pi te
<i>Non-adjacent segment languages</i>		
Language A	do ki bae da ku be po gu tae pa gi te	be po gu tae da ki ku bae pa gi te do
Language B	bae po gu be pa gi tae da ku te do ki	gi tae da ku be po pa gu te do ki bae

Table 6  
Design of two non-adjacent segment (vowel) languages used in Experiment 3

[c <sub>1</sub> ]V <sub>1</sub>	[c <sub>3</sub> ]V <sub>2</sub>	[c <sub>5</sub> ]V <sub>3</sub>
[c <sub>2</sub> ]	[c <sub>4</sub> ]	[c <sub>6</sub> ]
[c <sub>1</sub> ]V <sub>4</sub>	[c <sub>3</sub> ]V <sub>5</sub>	[c <sub>5</sub> ]V <sub>6</sub>
[c <sub>2</sub> ]	[c <sub>4</sub> ]	[c <sub>6</sub> ]
Vowel-frames	Consonant-fillers	
<i>Language A</i>		
_a_u_e	[p_] [g_] [t_]	
_o_i_ae	[d_] [k_] [b_]	
<i>Language B</i>		
_ae_u_u	[t_] [d_] [k_]	
_e_o_i_	[b_] [p_] [g_]	



Table 7  
Test words and test partwords used in Experiment 3

	Words	Partwords
<i>Non-adjacent segment languages</i>		
Language A	pa ku te da gu be po gi tae do ki bae	te do ki bae pa gu ku be po gi tae da
Language B	tae pa ku bae da gu te do ki be po gi	ku be po gi tae da pa gu te do ki bae

**Fig. 1.** Immagini modificate da Newport *et al.* (2004). Le regolarità distribuzionali implementate su segmenti fonetici (consonanti/vocali), ma non su sillabe - sono facilmente acquisibili da soggetti adulti.

Ciò vuol dire che i meccanismi di apprendimento sintattico operano anche su elementi distanti, ma soltanto se si utilizzano i segmenti fonetici - ovvero consonanti e vocali, ma non sillabe costituite dalla combinazione di una consonante con una vocale - come unità minime su cui si implementare le regolarità distribuzionali (figura 1). La causa di tale differenza nel processamento dei suoni necessita tuttora di ulteriori indagini sperimentali. Tuttavia, è noto che i due tipi di segmenti, ovvero le consonanti e le vocali assumono dei ruoli diversi nel modulare i meccanismi computazionali nella categorizzazione del flusso di suoni. In particolare, le consonanti sono più “adatte” ad astrarre computazioni in grado di individuare unità discrete,

mentre le vocali, per via della loro possibilità di veicolare informazioni prosodiche, come la variabilità del tono o l'estensione del suono stesso, si prestano maggiormente alla conduzione di informazioni legate alla struttura *sintattica* del flusso, ovvero intorno ai pattern strutturali che ne caratterizzano la distribuzione dei suoni. Tale conclusione è quanto emerge da uno studio condotto da Toro *et al.* (figura 2)<sup>56</sup>:

We have proposed that consonants give cues primarily about the lexicon, whereas vowels carry cues about syntax. In a study supporting this hypothesis, we showed that when segmenting words from an artificial continuous stream, participants compute statistical relations over consonants, but not over vowels. In the study reported here, we tested the symmetrical hypothesis that when participants listen to words in a speech stream, they tend to exploit relations among vowels to extract generalizations, but tend to disregard the same relations among consonants. In our streams, participants could segment words on the basis of transitional probabilities in one tier and could extract a structural regularity in the other tier. *Participants used consonants to extract words, but vowels to extract a structural generalization.* They were unable to extract the same generalization using consonants, even when word segmentation was facilitated and the generalization made simpler. Our results suggest that different signal-driven computations prime lexical and grammatical processing.<sup>57</sup>

---

<sup>56</sup> Toro, J. M., Nespors, M., Mehler, J., Bonatti, L. L., *Finding Words and Rules in a Speech Stream*, "Psychological science", 19, 2008, pp. 137-144.

<sup>57</sup> *Ivi*, p. 137 (corsivo mio).

**TABLE 1**  
*Experimental Stimuli in Experiment 1*

Familiarization words	Segmentation test		Generalization test	
	Words	Part-words	Rule-words	Nonrule-words
tapena	tepane	penabe	biduki	buduki
tapona	tapona	dokato	budiku	biduku
tepane	topano	paneba	tipani	tupuni
tepona	topeno	dakoba	tupinu	tipunu
topano	badeka	naboda		
topeno	bedake	katepa		
badeka	bedoke	nebade		
badoka	bodeko	kobado		
bedake				
bedoke				
bodako				
bodeko				

**Fig. 2** Immagine modificata da Toro *et al.* (2008). I soggetti che hanno preso parte all'esperimento tendevano a generalizzare il pattern statistico inerente la distribuzione delle consonanti per estrarre le parole dal flusso fonico, e quello inerente le vocali per risalire ad una generalizzazione strutturale del tipo ABA.

Tali risultati, sebbene inerenti soggetti adulti, possono considerarsi un'importante estensione dello studio di Saffran *et al.* (1996)<sup>58</sup>, in quanto ci mostrano, non soltanto che la presenza di relazioni statisticamente regolate all'interno del segnale sono in grado di modulare l'azione dei meccanismi computazionali, ma gettano anche luce sulle differenti proprietà modulatorie degli elementi *sintatticamente* correlati all'interno del segnale stessa. Questa idea solleva una importante considerazione in merito ai

<sup>58</sup> Saffran, J. R., Aslin, R. N., Newport, E. L., *Statistical learning by 8-month-old infants*, cit.

meccanismi di acquisizione del linguaggio. In effetti, al fine di orientarsi in una lingua, non è sufficiente essere in grado di discriminare le diverse unità costitutive, ma crucialmente, bisogna essere in grado di determinare in che modo le diverse unità si relazionano l'una all'altra, ovvero come includere i suoni linguistici sistematicamente correlati in un'unica unità categoriale discreta, e riconoscerne le relazioni logico-sintattiche. Nel prossimo paragrafo metteremo a tema i processi cognitivi coinvolti in tale operazione sintattica.

### **3.2 L'apprendimento di regole negli infanti pre-verbali**

La *naturalzza* con cui gli infanti acquisiscono una (o più) lingue nasconde il complicato intreccio di operazioni che stanno alla base di tale fenomeno. Il processo di acquisizione ontogenetica del linguaggio, come abbiamo visto, fa capo (almeno) alle proprietà statistiche distribuzionali del segnale sulla cui base ha luogo la segmentazione delle parole. Tuttavia, è possibile che tale abilità sia corredata da altre competenze, che si intrecciano ad essa, a formare un fascio di operazioni che all'unisono permettono al bambino di orientarsi nella lingua, ovvero di percepirne le unità costitutive, di comprenderne i significati, e di parlarla. A tal proposito

sono interessanti le osservazioni di Gary Marcus *et al.*<sup>59</sup>, la cui ipotesi di ricerca è che gli infanti siano in grado non soltanto di percepire regolarità distribuzionali, ma di discriminare veri e propri pattern strutturali assimilabili a regole algebriche in una sequenza di suoni non-senso. In tale prospettiva, Marcus assume che la capacità di discriminare tali “grammatiche” categoriali starebbe alla base della abilità di discriminare relazioni logiche tanto fra le variabili matematiche, quanto fra le unità lessemiche di una frase.

Nello specifico, al fine di testare l’abilità degli infanti di “captare” pattern logico-sintattici fra unità appartenenti ad insiemi distinti, Marcus ha implementato nel suo studio sperimentale dei pattern molto semplici, del tipo “ABA” e “ABB”, a cui corrispondevano, rispettivamente stringhe del tipo “*ga ti ga*”, o “*wo fe wo*” e “*ga ti ti*” e “*wo fe fe*”. Da tale ricerca emerge che i neonati di appena sette mesi d’età sono in grado di discriminare questi due tipi di “grammatiche”, sottese tanto al linguaggio verbale quanto a quello matematico:

We propose that a system that could account for our results is one in which infants extract abstract algebra-like rules that represent relationships between placeholders (variables), such as “the first item X is the same as the third item Y,” or more generally, that “item I is the same as item J.” In addition to having the capacity to represent such rules, our results appear to show that infants have the ability to extract those rules rapidly from small amounts of input and to

---

<sup>59</sup> Marcus, G., Vijayan, S., Rao, S. B., Vishton, P. M., *Rule Learning by Seven-Month-Old Infants*, “Science”, 283, 1999, pp. 77-79.

generalize those rules to novel instances. If our position is correct, then infants possess at least two distinct tools for learning about the world and attacking the problem of learning language: one device that tracks statistical relationships such as transitional probabilities and another that manipulates variables, allowing children to learn rules. Even taken together, these tools are unlikely to be sufficient for learning language, but both may be necessary prerequisites.<sup>60</sup>

L'idea espressa, dunque, presenta due assunti di fondo. Il primo è che la competenza “algebraica” di riconoscere un algoritmo implementabile sulle variabili numeriche possa trovare un analogo sul piano delle variabili di natura simil-linguistica, qual è il caso delle sillabe che utilizza nel suo studio. In secondo luogo, l'autore assume che la relazione fra le unità sillabiche possa essere governata oltre che da regole di tipo statistico-distribuzionali, anche da principi di natura logico-algebraica; così un algoritmo molto semplice del tipo “il primo elemento è uguale all'ultimo” è in grado di fondare la connessione logica fra due elementi spazialmente o temporalmente non consecutivi, tanto sul piano delle variabili numeriche, quanto su quello dei suoni vocali<sup>61</sup>.

In tale studio viene enfatizzato un aspetto molto importante, ovvero che lo studio delle abilità di individuare unità minime per mezzo di

---

<sup>60</sup> Ivi, p. 79. Cfr. Pena, M., Bonatti, L. L., Nespor, M., Mehler, J. *Signal-Driven Computations in Speech Processing*, “Science”, 298, 2002, pp. 604–607.

<sup>61</sup> L'ipotesi secondo la quale i soggetti di tale studio astraggano delle regole grammaticali, piuttosto che discriminare probabilità di transizioni tra sillabe è stata oggetto di un acceso dibattito. Vedi, a tal proposito: Seidenberg, M. S., *Do infants learn grammar with algebra or statistics?*, “Science”, 284, 1999, p. 433; Christiansen, M. H., Curtin, S., *Transfer of learning: Rule acquisition or statistical learning?*, “Trends in Cognitive Science”, 3, 1999, pp. 289–290; Marcus, G. F., *Reply to Christiansen and Curtin*, “Trends in Cognitive Sciences”, 3, 1999, pp. 290–291.

statistiche distribuzionali debba essere affiancato dall'analisi della cruciale competenza cognitiva di connettere tali unità (siano esse sillabe e parole non senso nei linguaggi artificiali, o lessemi nel caso delle lingue naturali) entro pattern strutturali. Tuttavia, troviamo che l'indagine su tale tipo di connessioni strutturali necessita di alcune fondamentali chiarificazioni terminologiche e concettuali inerenti, in particolare, la natura effettiva delle relazioni intercorrenti fra i diversi stimoli, argomento cruciale nella comprensione delle differenze costitutive fra le competenze sintattiche delle diverse specie animali. Nel prossimo capitolo cercheremo di delineare con maggiore chiarezza la fisionimia di tale tema. Prima, però, vale la pena di passare in rassegna alcuni studi comparativi sulla presenza di analoghi meccanismi dedicati all'apprendimento di regole sintattiche negli animali non umani. Sarà questo l'oggetto di studio del prossimo paragrafo.

### **3.3 Acquisizione di regole e apprendimento statistico nelle scimmie**

Le ricerche sperimentali condotte sugli infanti confermano, come abbiamo visto, la presenza funzionale di meccanismi dominio-generalisti dedicati alla percezione strutturale di regolarità *sintattiche* fra i dati sensoriali, siano essi di natura linguistica, musicale, o visuale. Alla luce di

tali dati, obiettivo del presente paragrafo sarà di esplorare la possibilità che tali capacità, proprie degli esseri umani pre-verbali possano estendersi agli animali non umani. In altri termini, cercherò di esplorare la tesi secondo la qua cui gli animali non umani condividono con altre specie animali l'abilità di *sintassi*, intesa, in questo caso come la capacità di percepire delle regolarità strutturali descrivibili in termini statistici.

In un studio pubblicato nel 2001, Marc Hauser *et al.*<sup>62</sup> hanno replicato sui tamarini dal ciuffo bianco lo stesso esperimento condotto da Marcus relativamente alle competenze algebriche degli infanti. Nello specifico, gli studiosi hanno testato le capacità di tale specie di scimmie di: a) discriminare statistiche sequenziali a partire da un flusso di sillabe concatenate, e b) usare tali statistiche per determinare quali sequenze di sillabe formano parole ben formate. In altri termini, gli studiosi si sono chiesti se tale specie vanta il meccanismo cognitivo necessario ad acquisire determinati pattern sintattici, e a riconoscere tali regole in quanto corrette. Da tale studio emerge che i tamarini, in effetti, vantano tale competenza cognitiva. Come osservano gli autori dello studio:

These results suggest that humans and tamarins may share a common mechanism for rapidly processing and retaining these aspects of serial order information. However, even if humans and some non-human primates share the ability to

---

<sup>62</sup> Hauser, M. D., Newport, E. L., Aslin, R. N., *Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins*, "Cognition", 78, 2001, pp. B53 - B64.



compute certain elementary patterns of serial order, these abilities in humans (at least for the domain of language) have been extended and elaborated well beyond what we have demonstrated in the present tasks, into complexities of serial order processing that we do not expect non-human primates to share. The type of statistical computation examined here (computing co-occurrence relations between immediately adjacent elements) is a fundamental aspect of serial learning, potentially important to providing learners with an initial segmentation of speech or many other types of patterned domains [...]. However, such an analysis would not be adequate for acquiring higher level patterns in language; many structures in natural languages involve relationships between classes of elements, or relations between elements not immediately adjacent to each other [...]<sup>63</sup>.

Tale ricerca ci rivela dunque che l'abilità di discriminare e computare regolarità statistiche, comune tanto agli esseri umani quanto ai tamarini non è il tratto cognitivo in grado di demarcare la specificità delle abilità sintattiche dei tamarini da quelle umane, i cui pattern sono caratterizzati da relazioni infrastrutturali ben più complesse, ed i cui elementi possono non essere acusticamente adiacenti o vicini. In tale prospettiva di indagine si collocano gli studi condotti dallo stesso Hauser e collaboratori<sup>64</sup> sulla abilità dei tamarini di astrarre regole combinatorie a partire da una stringa di sillabe, e di generalizzarne l'applicazione su stringhe differenti. Infatti, se nella fase di abituazione, i tamarini sono esposti, ad esempio, alle stringhe costruite sulla regola ABB (i.e. "ga ti ti", "ga na na", "li ti ti"), essi saranno in grado di riconoscere in quanto non ben formate la stringhe "ga ga ti", o "li li ti", che vanno sotto la regola *sintattica* AAB

---

<sup>63</sup> Ivi, p. B62.

<sup>64</sup> Hauser, M. D., Weiss, D., & Marcus, G., *Rule learning by cotton-top tamarins*, *Cognition*, 86, 2002, pp. B15–22.

We habituated subjects to sequences of consonant–vowel syllables that followed one of two patterns, AAB (e.g. wi wi di) or ABB (le we we). Following habituation, we presented subjects with two novel 28 test items, one with the same pattern as that presented during habituation and one with a different pattern. Like human infants, tamarins were more likely to dishabituate to the test item with a different pattern. We conclude that the capacity to generalize rule-like patterns, at least at the level demonstrated, did not evolve specifically for language acquisition, though it remains possible that infants might use such rules during language acquisition<sup>65</sup>.

Tali dati confermano dunque l'abilità dei tamarini di astrarre e generalizzare delle regolarità *sintattiche* la cui applicazione, tuttavia, evidentemente non abbraccia ancora la effettiva complessità di acquisizione e produzione vocale del linguaggio umano. Il fatto che gli esseri umani condividano con talune specie di scimmie (o di uccelli) certe abilità cognitive, nello specifico, l'abilità di acquisire, riconoscere e generalizzare una regola sintattica, ci segnala la presenza di possibili precursori (ovvero delle condizioni necessarie alla evoluzione del linguaggio proposizionale, ma di certo non sufficienti) di quella che si sarebbe evoluta nella capacità di percepire una forma assai più complessa ed articolata di organizzazione sintattica: il linguaggio verbale. Rimane allora da capire esattamente quale pattern sintattico rappresenta la configurazione la cui padronanza è in grado di rivelarci l'area di confine fra le abilità cognitive delle due specie.

---

<sup>65</sup> *Ivi*, p. 1178.

In tale direzione, vale la pena di prendere in esame lo studio di ricerca condotto da Newport *et al.*<sup>66</sup> sempre sui tamarini, al fine di esplorarne l'abilità di tracciare - all'interno di un flusso di sequenze foniche - dipendenze strutturali tra elementi non immediatamente adiacenti. L'idea degli studiosi era di applicare su una specie non umana lo stesso studio sperimentale che abbiamo descritto nel paragrafo 3.1 sulle abilità di tracciare la distribuzione statisticamente regolata di elementi non adiacenti. Come abbiamo visto, gli umani sono in grado di discriminare dipendenze tra elementi non adiacenti laddove questi sono segmenti fonici (vocali o consonanti), ma non sillabe. Interessantemente, dalla ricerca in questione emerge che i tamarini siano effettivamente in grado di computare le transizioni probabilistiche anche fra elementi non adiacenti, ma soltanto se si tratta di sillabe o di vocali. Dunque, sebbene le due diverse specie mostrano una comune abilità di tracciare dipendenze strutturali tra elementi spazialmente distanti, emergono delle differenze di fondo a proposito della natura degli stimoli che sono in grado di processare computazionalmente. La ragione di tale diversità può essere rintracciata in diversi ordini di spiegazione. Anzitutto, bisogna tenere presente che per i tamarini si trattava

---

<sup>66</sup> Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N., *Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate*, "Cognitive psychology", 49, 2004, pp. 85-117.

di stimoli non specifici della propria specie<sup>67</sup>. Ne consegue che non possiamo esser certi sulla effettiva natura di ciò che era effettivamente percepito dalle scimmie. Così, la maggiore prominenza percettuale delle vocali e delle sillabe rispetto alle consonanti potrebbe aver reso il compito di computazione più agevole per le scimmie. Inoltre, gli studiosi osservano quanto segue:

[...] the concepts of ‘distance’ and ‘non-adjacency’ differ greatly, depending on the representations a learner has of the materials to be learned. This possibility encompasses some of the issues already raised above, but with a somewhat different emphasis. If tamarins hear human speech as relatively large, holistic chunks of sound, occurring in a linear sequence, then the acquisition of non-adjacent syllables reduces to keeping track of those syllables and how often they occur together; this problem is only slightly more difficult than acquiring adjacent syllables, in that all the syllables one away would have to be computed. In contrast, the much more complex and well-articulated representation that adult humans have of human speech might make non-adjacent syllables much more distant from one another than they are for tamarins.<sup>68</sup>

In altri termini, dunque, è possibile che in effetti il tipo di computazioni che gli umani effettuano all’esposizione dei medesimi stimoli si esplichino a più livelli di complessità, ovvero nella discriminazione

---

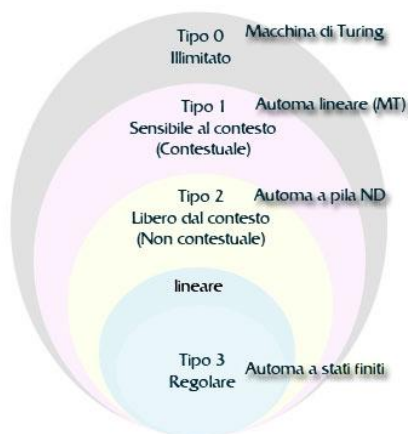
<sup>67</sup> *Ivi*, p. 111: “Our findings suggest that non-adjacency per se is not a problem for tamarins: they are able to compute both non-adjacent syllable regularities and non-adjacent segment (vowel) regularities; only consonants give them difficulty, likely arising from the acoustic complexities of consonants themselves. What is surprising, then, is that certain types of non-adjacency are extremely difficult for human listeners. Why should non-adjacency—particularly syllable non-adjacency—be difficult for human listeners and relatively easy for tamarin monkeys? As noted above, this is not likely to be because tamarins are in general more cognitively capable than adult humans. It must therefore be because human speech is processed in a different way by humans than by tamarins, and particularly in such a way that the computation of non-adjacent syllable regularities becomes more complex for human adults.”

<sup>68</sup> *Ivi*, p. 112.

implicita di pattern complessi fra segmenti non adiacenti, e al contempo, nella segmentazione di unità ben formate; inoltre, è possibile che, mentre i tamarini abbiano una percezione olistica e generale degli stimoli linguistici che ne sfuma la complessità acustica, nel caso degli umani, la percezione fine degli stimoli in tutte le loro caratteristiche segmentali ne può aver reso rendere più complicato ed impegnativo il processamento. Tuttavia, al di là di tali differenze di natura percettuale, l'abilità dei tamarini di distinguere relazioni strutturali tra elementi spazialmente non sequenziali, e regolati da pattern distribuzionali specifici, rappresenta un indizio significativo relativamente al percorso evolutivo del linguaggio proposizionale. In quest'ultimo infatti, come abbiamo osservato nel paragrafo precedente, le forme relazionali che interessano le parole si dispiegano su una sequenza logico-sintattica che può interessare elementi spazialmente o temporalmente non contigui. Il modo in cui si sia affermata la corrispondenza tra il pattern logico-sintattico che coinvolge le parole, e i loro pattern distribuzionali sul piano percettuale rimane un tema suscettibile di ulteriori approfondimenti sul piano sia teorico che sperimentale.

## 4. Percezione di pattern fono-sintattici: il paradigma chomskiano

All'interno del paradigma dell'*artificial grammar learning*, oltre all'approccio dell'apprendimento statistico, diversi studi hanno adottato un metodo differente, al fine di esplorare le abilità sintattiche comuni agli umani e ad altre specie animali. Si tratta del metodo di progettazione sperimentale che fa capo a quella distinzione dei sistemi computazionali effettuata da Chomsky, che va sotto il nome di “gerarchia di Chomsky”.



**Fig. 1.** Immagine modificata da [http://it.wikipedia.org/wiki/Grammatica\\_formale](http://it.wikipedia.org/wiki/Grammatica_formale)

In “Three Models for the Description of Language”<sup>69</sup>, un articolo centrale nel dibattito intorno all’intelligenza artificiale, e in generale sulla natura computazionale del linguaggio – Chomsky afferma che la capacità di distinguere una frase corretta da una che non lo è si basa sull’abilità di riconoscere le proprietà strutturali e formali da cui sono generate. In particolare, tre nozioni sono fondamentali nel suo approccio formale allo studio del linguaggio:

The first step in the linguistic analysis of a language is to provide a finite system of representation for its sentences. We shall assume that this step has been carried out, and we shall deal with language only in phonemic or alphabetic transcription. By a language, then, we shall mean a set (finite or infinite) of *sentences*, each of finite length, all constructed from a finite *alphabet* of symbols. If *A* is an alphabet, we shall say that anything formed by concatenating the symbols of *A* is a string in *A*. By a grammar of the *language* *L* we mean a device of some sort that produces all of the strings that are sentences of *L* and only these.<sup>70</sup>

L’idea centrale è che ogni insieme di stringhe o linguaggi possano essere ordinati secondo un livello crescente di complessità computazionale, in maniera tale che il livello inferiore sia sempre un sottoinsieme di quello superiore (figura 1). Tale classificazione adotta la nozione di complessità nei termini della quantità di informazione (e di memoria) di cui un sistema artificiale necessita relativamente al contesto locale e al suo stato interno, per riconoscere un particolare pattern sintattico in quanto corretto.

---

<sup>69</sup> Chomsky, N., Three models for the description of language, cit.

<sup>70</sup> *Ivi*, p. 114 (corsivi miei).

L'obiettivo di fondo di tale operazione formale è di determinare la natura delle grammatiche in grado di generare le frasi di una lingua storico-naturale. Nello specifico, Chomsky distingue tre classi di grammatiche, in grado di generare altrettanti tipi di linguaggi, attraverso la manipolazione matematica di un insieme di sequenze finite di simboli (stringhe) appartenenti ad un insieme finito di simboli (alfabeto). Ai fini della nostra ricerca, le due classi di grammatiche formali che vale la pena di mettere in evidenza sono le grammatiche regolari (o "a stati finiti") e quelle libere dal contesto (strutture frasali). Le prime possono essere elaborate da automi a stati finiti (FSA), un modello di calcolo semplice, rappresentabile come un piccolo dispositivo, che mediante una testina legge una stringa di input su un nastro e la elabora usando un apposito algoritmo, e di una memoria limitata, in quanto le transizioni da uno stato all'altro dipendono *esclusivamente* dallo stato attuale dell'automa. Consideriamo, ad esempio, la grammatica  $(AB)^n$ , la cui unica regola è quella di far seguire a ciascuna occorrenza di A, una occorrenza di B, per un numero  $n$  di volte. Per creare un simile linguaggio, o essere in grado di riconoscerne la correttezza formale, ad una macchina a stati finiti basterebbero l'informazione relativa all'input iniziale (l'inizio della stringa) e alla transizione allo stato successivo, che riconoscerebbe come lo stato finale. Ciò vuol dire che tali transizioni sono computabili attraverso una catena di Markov, ovvero



attraverso un processo nel quale la probabilità di transizione che determina il passaggio ad uno stato di sistema dipende unicamente dallo stato di sistema immediatamente precedente (proprietà di Markov) e non dal *come* si è giunti a tale stato. Quest'ultimo punto è molto importante, in quanto seguendo le regole formali, una macchina a stati finiti è in grado di generare un insieme potenzialmente infinito di stringhe ben formate, ma non necessita di alcune capacità di memoria che vadano oltre la possibilità di ricordare lo stato corrente della macchina.

È opportuno tenere presente, ai fini della nostra ricerca, che le grammatiche regolari, presentano al loro interno diversi livelli di complessità. Si danno infatti, entro tale classe i linguaggi detti “*strictly locally 2-testable languages*”, in cui rientra l'esempio di cui sopra  $(AB)^n$ , il cui riconoscimento formale richiede semplicemente la padronanza del bigramma  $ab$ , processabile bilateralmente:

The set of strings of the form  $(AB)^n$  is an example of an SL2 stringset, being definable by the (minimal) set of 2-factors:

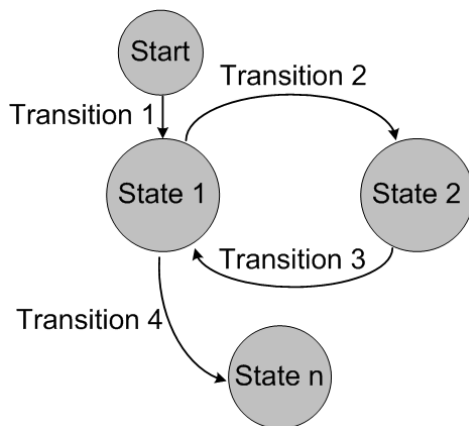
$$(2) \quad D(AB)^n = \{\times A, AB, BA, B \times\},$$

which asserts that the string must begin with an A, end with a B, that every A is followed by a B and that every B other than the last is followed by an A. The set of strings licensed by this definition, denoted  $L(D(AB)^n)$ , is  $\{(AB)^i \mid i > 0\}$ .<sup>71</sup>

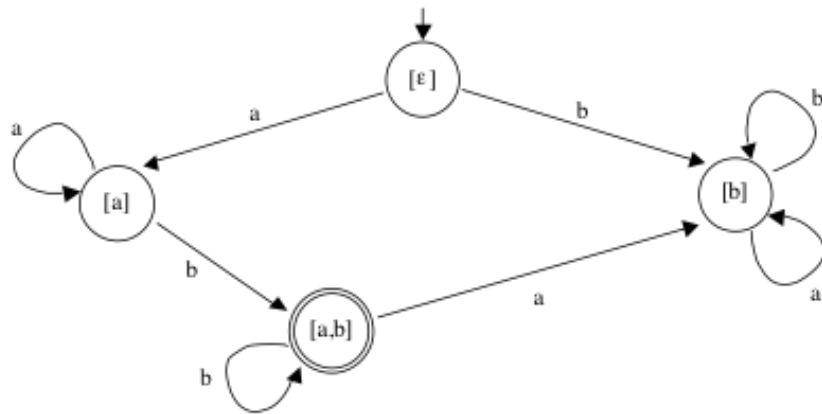
---

<sup>71</sup> Rogers, J., & Hauser, M. D., *The use of formal language theory in studies of artificial language learning: A proposal for distinguishing the differences between human and nonhuman animal learners*, in H. van der Hulst (ed.), “Recursion and Human Language”, Gruyter, Walter De GmbH, Berlin/New York, 2010, pp. 213-232.

Un secondo tipo di grammatica regolare, più complesso rispetto a ad un ordine di descrizione bigrammatica, è quello in cui, le transizioni tra gli stati sono stocastici, ovvero, non deterministici. Ciò vuol dire che, ogni stato presenta più di una possibile computazione per determinati caratteri in ingresso. Entrambi i tipi di grammatiche a stati finiti possono essere rappresentati mediante un grafo, in cui i nodi rappresentano gli stati e gli archi le transizioni, o in una versione tabellata, sulle righe lo stato corrente e sulle colonne l'ingresso, e l'elemento della tabella rappresenta la transizione verso il prossimo stato e l'uscita dell'automa:



**Fig. 2.** Automa a stati finiti deterministico



**Fig. 3.** Automa a stati finiti non deterministico

Tradotto nei termini di un linguaggio naturale un automa di tale complessità potrebbe paragonarsi ad una persona che produce frasi semplicemente legando una parola all’ultima proferita, a prescindere da tutto ciò che occorre nel resto della frase. Chiaramente la complessità delle lingue storico-naturali va ben oltre tale meccanismo. Si pensi ad esempio alla frase “Giovanni, che è bravo a suonare il piano, stasera terrà un concerto”; in quest’esempio le parole sono legate da relazioni logiche più complesse, che richiedono una capacità di memoria più ampia, ovvero in grado di abbracciare i termini (i simboli, nel caso degli automi) la cui consequenzialità logica non trova una corrispettiva contiguità spaziale all’interno della frase/stringa. Il secondo tipo di grammatiche che vale la pena di prendere in esame, ai fini del nostro studio, sono le grammatiche delle strutture frasali. Così ne parla lo stesso Chomsky:

Costumarily, syntactic description is given in terms of what is called “immediate constituent analysis.” In description of this sort the words of a sentence are grouped into phrases, these are grouped into smaller constituents phrases and so on, until the ultimate constituents (generally morphemes) are reached. These phrases are then classified as noun phrases (NP), verb phrases (VP), etc. Evidently, description of sentences in such terms permits considerable simplification over the word-by-word model, since the composition of a complex class can be used as a building block at various points in the construction of sentences.<sup>72</sup>

the	man	took	the	book
NP		Verb	NP	
		VP		
Sentence				

**Fig. 4.** Immagine modificata da Chomsky (1956)

L'idea dello studioso è dunque che questo genere di grammatica permette di sostituire delle variabili applicando delle regole di sostituzione su categorie più ampie che sono i sintagmi nominali e verbali. Vediamo adesso in che modo il linguaggio verbale possa trovare una traduzione in termini computazionali. Il processamento di una frase ordinaria del linguaggio come, ad esempio “se non piove e Gianni mi accompagna, allora verrò a prenderti in aeroporto” va oltre le competenze di una macchina a stati finiti, in quanto richiede che *tutti* gli stati relativi agli elementi interni alle frasi connesse vengono tenuti in memoria fino allo

<sup>72</sup> Chomsky, N., Three models for the description of language, cit, p. 116.

stato finale. Fitch ne spiega molto bene il processo in un passo che vale la pena di riportare nella sua lunghezza:

Consider the notion of a conversational aside: John got fired might be rephrased as John, who works downstairs, got fired or John, who works (or I should say used to work) downstairs, got fired. Each of these sentences conveys the same basic message, but the later ones include an “embedded” specification or clarification. Formally, forgetting about their meaning, sentence structures of this form can be captured by strings of embedded parentheses: (), (()), and ((())). Intuitively, we know that a string of parentheses with a form like ( () or )() is incomplete or ill-formed, and that with time and patience, and paper and pencil, we could extend such “grammaticality” judgments to any number of parentheses. Our ability to do this with very long strings is limited only by extraneous factors like how long we can stay awake. Perhaps surprisingly, no FSA can be programmed to duplicate this feat. In formal terms, we say that the language of “balanced” parentheses (a “Dyck language”) cannot be recognized by any possible FSA. The reason is that the only way an FSA can keep track of the past parentheses is via the current state it is in, and since (by definition) its states are limited, we can easily “break” any FSA by feeding it any finite, well-formed parenthetical expression that has more opening parentheses than our FSA has states.<sup>73</sup>

Allo stesso modo, un linguaggio che va oltre le abilità di un automa a stati finiti è quello che consiste di frasi in cui il numero delle occorrenze degli elementi della prima stringa di elementi di un insieme A è seguito (esclusivamente) da una stringa con lo stesso numero di occorrenze di elementi dell'insieme B. Tale grammatica - rappresentabile dalla regola  $A^n B^n$  - rientra fra le grammatiche dette “libere dal contesto”, o delle strutture frasali. Come vedremo nei prossimi paragrafi, le regole sintattico-

---

<sup>73</sup> Fitch, W. T., *The evolution of language*, Cambridge University Press, 2010, pp. 112-113.

formali implementate nella derivazione di grammatiche a stati finiti e libere dal contesto sono utilizzate nella creazione degli stimoli su cui testare le abilità sintattiche di specie animali differenti, per fini comparativi. Nello specifico, in tale paradigma di studi, una delle idee fondanti è l'ipotesi che vi sia un tratto cognitivo, in particolare, che fonda la specie-specificità della cognizione umana: la ricorsività.

#### **4.1 Facoltà comuni e innovazioni evolutive**

In *The faculty of language: what is it, who has it, and how did it evolve?*<sup>74</sup>, articolo centrale nell'odierno dibattito intorno alla evoluzione del linguaggio, Hauser, Chomsky, e Fitch mettono a tema un importante questione: la individuazione del tratto cognitivo (nello specifico, sintattico-computazionale) che nel percorso evolutivo della facoltà di linguaggio ha segnato la speciazione dell'uomo. Procedendo con quel rigore metodologico necessario per l'analisi di una facoltà tanto complessa come quella del "linguaggio", essi distinguono due accezioni del termine. La prima è la FLN, ovvero la facoltà di linguaggio, intesa in senso stretto (*faculty of language – narrow sense*). Essa comprende un nucleo di

---

<sup>74</sup> Hauser M. D., Chomsky N., Fitch W. T., *The faculty of language: what is it, who has it, and how did it evolve?*, "Science", 298, 2002, pp 1569-79.

<sup>48</sup> *Ivi*, pp. 1569, 1571, 1573 (corsivo mio).

meccanismi computazionali in grado di costruire una serie potenzialmente infinita di espressioni connesse all'interfaccia concettuale-intenzionale da trasmettere all'esterno per mezzo del canale acustico-visivo, ovvero all'interfaccia senso-motoria. In altri termini, attraverso una serie discreta di regole implementate sulle espressioni finite dell'interfaccia concettuale-intenzionale, la FLN connetterebbe i due sistemi fornendo dei significati concettuali esternabili ed interpretabili a livello senso-motorio, realizzando in tale modo la possibilità specie-specifica umana di articolare discorsi intersoggettivamente comprensibili. Gli autori affermano:

We hypothesize that FLN only includes recursion and is the only uniquely human component of the faculty of language. [...] A key component of FLN is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system. [...] All approaches agree that *a core property of FLN is recursion*, attributed to narrow syntax in the conception just outlined. [...] We propose in this hypothesis that *FLN comprises only the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces*.<sup>75</sup>

Fondando il proprio carattere di unicità su tale abilità specificamente umana, gli autori assumono che il linguaggio umano differisca radicalmente da quello delle altre specie in termini qualitativi. Tuttavia, nel tessuto teorico delle idee di Chomsky, Fitch e Hauser, è possibile

---

<sup>49</sup> Vedi Ivi, p. 1576.

individuare un'importante idea continuista<sup>76</sup>, concernente la seconda accezione di linguaggio, ovvero il dominio generale del linguaggio, inteso in senso ampio, comprendente, cioè, al suo interno tanto le facoltà cognitive condivise con le altre specie animali, quanto la FLN, appannaggio unicamente umano. Si tratta della FLB (*faculty of language - broad sense*), e di essa così ne parlano gli autori:

FLB includes an internal computational system (FLN) combined with at least two other organism-internal systems, which we call “sensory-motor” and “conceptual- intentional”. Despite debate on the precise nature of these systems, and about whether they are substantially shared with other vertebrates or uniquely adapted to the exigencies of language, we take as uncontroversial the existence of some biological capacity of humans that allows us (and not, for example, chimpanzees) to readily master any human language without explicit instruction. FLB includes this capacity [FLN], but excludes other organism-internal systems that are necessary but not sufficient for language (e.g., memory, respiration, digestion, circulation, etc.). [...]  
By this hypothesis, FLB contains a wide variety of cognitive and perceptual mechanism shared with other species, but only those mechanisms underlying FLN – particularly its capacity for discrete infinity – are uniquely human.<sup>77</sup>

Le interfacce senso-motoria e concettuo-intenzionale, che nell'uomo sono connesse al sistema ricorsivo di operazioni computazionali, in sostanza comprendono la capacità di percepire e produrre suoni articolati, di orientarsi nello spazio, di enumerazione, di imitazione vocale, di ritmo, l'intelligenza sociale: tutti meccanismi impiegati nei domini cognitivi di diverse specie animali oltre a quella umana, ma che in

---

<sup>50</sup>Ivi, pp. 1570-71, 1573.



quest'ultima trovano una configurazione del tutto eccezionale. Chomsky, Hauser e Fitch, a tal proposito affermano: «Something about the faculty of language must be unique in order to explain the differences between humans and other animals – if only the particular combination of mechanisms in FLB»<sup>78</sup>.

Ora, questo è proprio il *punctum saliens* attorno al quale ruota la loro trattazione. Ciò che distingue il linguaggio umano da quello degli altri animali è, al minimo, che in esso le componenti della FLB (comuni alle diverse specie) sono organizzate secondo un'architettura complessa, un'architettura che trova la sua matrice nella ricorsività.

We hypothesize that most, if not all, of FLB is based on mechanisms shared with nonhuman animals [...]. In contrast, we suggest that FLN - the computational mechanism of recursion – is recently evolved and unique to our species. [...]

The innovation that yielded the faculty of language was the evolution of the computational system that links [the conceptual-intentional and the sensory-motor interface systems].<sup>79</sup>

Tale passo presenta due punti concettuali che vale la pena di sottolineare. Il primo è che la maggior parte delle componenti della FLB sono qualitativamente simili negli umani e nelle altre specie animali;

---

<sup>78</sup> Fitch W. T., Hauser M. D., Chomsky N., *The evolution of the language faculty: clarifications and implications*, "Cognition", 97, 2005, p. 182.

<sup>52</sup> Hauser, M. D., Chomsky, N., Fitch, W. T., *The faculty of language: what is it, who has it, and how did it evolve?*, in cit., pp. 1573, 1578. Vedi Fitch, W. T., Hauser, M. D., Chomsky, N., Op. cit., p. 182: "But our hypothesis focuses on a known property of human language that provides its most powerful and unusual signature: discrete infinity".

dunque la storia evolutiva di tali tratti potrebbe affondare le sue origini in un periodo molto più antico rispetto alla nascita del linguaggio umano. Il secondo consiste nell'assunto secondo il quale la ricorsività è emersa in una fase recente della storia evolutiva dell'uomo permeando di sé tali tratti condivisi. In tal modo il linguaggio avrebbe acquisito un tratto unico: la infinità discreta, ossia la capacità di combinare secondo una organizzazione gerarchica unità significanti in una varietà potenzialmente infinita di strutture più ampie, il cui significato differisce sistematicamente. Tale caratteristica contraddistingue il linguaggio umano, facendone un potente congegno creativo, e regolato allo stesso tempo.

## **4.2 Percezione di grammatiche artificiali nelle scimmie**

La distinzione delle grammatiche formali, effettuata da Chomsky, sta alla base di uno studio di ricerca condotto da Fitch e Hauser sulle abilità sintattiche dei tamarini dal ciuffo bianco<sup>80</sup>; gli autori, infatti, hanno

---

<sup>27</sup> Fitch, W. T., Hauser, M. D., *Computational constraints on syntactic processing in a nonhuman primate*, "Science", 303, 2004, pp. 377-380. Vale la pena di riportare per intero il seguente passo, p. 377: "Syntax is one key component of human language, with no equivalent in animal communication systems. The limitless expressive power of human language requires structures, termed phrases or sentences, above the word level (or, by analogy, above the single call level in animals). Linguistic syntax involves the rearrangement and permutation of such abstract hierarchical structures, often with concomitant changes in meaning. The production and perception of these hierarchical syntactic structures is a core capability underlying human linguistic competence. This level of organisation goes far beyond the simple concatenation procedures sometimes called "syntax" in animal communication. However, the evolution of language faculty presumably involved the incorporation of some ancestral primate cognitive capabilities. Thus, a critical question is whether hierarchical processing was one of these preexisting

individuato proprio nei test sulla capacità di riconoscere strutture gerarchiche un buon luogo in grado di indicare i “limiti” sintattici di tale specie di scimmie, ed eventualmente, di confermare l'appannaggio esclusivamente umano della operazione della ricorsività. Per testare l'ipotesi che i tamarini siano in grado di riconoscere sequenze lineari, ma non di identificare le relazioni logico-sintattiche intercorrenti tra le sottounità che costituiscono le diverse parti delle sequenze, i due ricercatori hanno selezionato due tipi di grammatiche, ovvero  $A^nB^n$  e  $(AB)^n$ , appartenenti rispettivamente all'insieme delle grammatiche a stati finiti, e all'insieme di grammatiche libere dal contesto. Vediamo più da vicino la logica sottesa. Nella struttura  $A^nB^n$ , ad ogni  $a$  segue linearmente una  $b$ . Ciò vuol dire che gli elementi che vi occorrono non sono legati da alcuna relazione di tipo gerarchico, bensì dalla regola deterministica “se  $a$ , allora  $b$ ”. Al contrario, le grammatiche in cui rientra la struttura  $(AB)^n$  - ad esempio  $(AB)^3$ , a cui corrisponde la stringa  $aaa-bbb$  - stando a questo paradigma teorico, sarebbero governate da una regola di tipo gerarchico-ricorsivo, in quanto dato il numero delle occorrenze di  $b$  è vincolato al numero delle occorrenze di  $a$ , secondo uno stesso principio di annidamento entro la struttura globale. Ciò vorrebbe dire che esiste un principio logico che, nel regolare la struttura delle due sottostringhe, determina la

---

abilities, perhaps evolved to serve non communicative functions (e.g., motor control, number, or social cognition).”

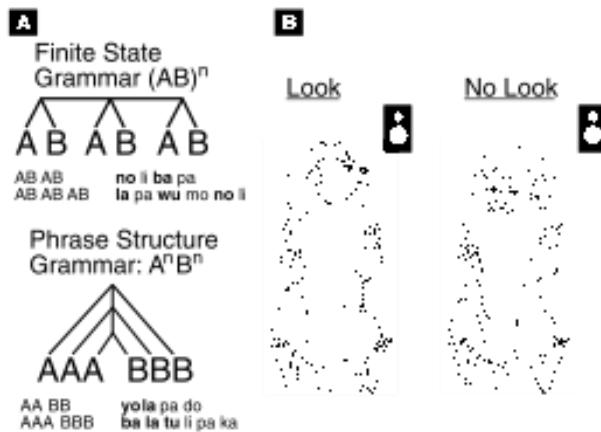
correttezza (e la completezza) della data sequenza nella sua globalità, come avviene nella costruzione del senso globale di una frase attraverso l'elaborazione delle parole che vi occorrono. Così ne parlano gli autori:

The weakest class in this hierarchy are finite state grammars (FSGs), which can be fully specified by transitions probabilities between a finite number of "states" (e. g., corresponding to words or calls). Recent evidence suggests that parsing procedures at this level of complexity are spontaneously available to both human and nonhuman primates. However, FSGs are inadequate to generate all the structures of any human language, because all languages minimally require procedures at next level of complexity, termed phrase structure grammars (PSGs)<sup>81</sup>.

L'idea alla base di tale studio è che le grammatiche a stati finiti, il cui meccanismo di connessione consiste in una mera concatenazione di simboli e stringhe, rappresentino una sorta di "ponte evolutivo", che ha condotto le capacità cognitive dell'uomo allo stadio attuale, ovvero alle *phrase-structure grammars*, le grammatiche delle strutture frasali dalla complessa struttura gerarchica, in cui, cioè, un elemento si annida in se stesso, ed in cui è possibile connettere unità anche distanti fra loro, in funzione del principio della "dipendenza dalla struttura". L'ipotesi avanzata da Fitch e Hauser, infatti, è che i tamarini dai ciuffi bianchi fossero in grado di distinguere una stringa a stati finiti, ma non una grammatica gerarchica (figura 5).

---

<sup>28</sup> *Ivi*, pp. 377-378.



**Fig. 5.** Immagine modificata da Fitch, Hauser (2004). Adottando il paradigma metodologico della violazione delle aspettative, gli autori hanno mostrato che i tamarini sono in grado di discernere le grammatiche a stati finiti ma non quelle frasali.

Nello specifico, per mezzo dell'associazione di sillabe appartenenti a due classi distinte A e B, hanno creato un insieme di frasi formali (ovvero stringhe di suoni prive di senso) regolati da una grammatica a stati finiti ed un insieme regolato da una grammatica "context free". Siano *a* e *b*, sillabe appartenenti rispettivamente ad A (sillabe proferite da una voce umana femminile), e a B (sillabe proferite da una voce maschile), la grammatica a stati finiti consisteva in una stringa di sillabe, tale che ad una *a* seguiva sempre una *b* (*ab*, *ab-ab*, *ab-ab-ab*, ...), mentre la grammatica gerarchica corrispondente alla struttura frasale consisteva in una stringa tale che ad un gruppo di *a*, seguiva un gruppo con tante *b*, quante erano le *a* del primo

gruppo; ad esempio, alla stringa *aaa* seguiva *bbb*, ad *aaaa* seguiva *bbbb*, e così via<sup>82</sup>. Così gli autori descrivono le conclusioni tratte:

[...] despite a clear ability to process sequential regularities in acoustic strings, tamarins are unable to process a simple phrase structure, where components at one portion of string are related to other components some distance away. [...] Tamarins suffer from a specific and fundamental computational limitation on their ability to spontaneously recognize or remember hierarchically organized acoustic structure.<sup>83</sup>

I tamarini, dunque, non sono in grado di riconoscere (contando) la “dipendenza strutturale” intercorrente tra il gruppo di *a* e quello di *b*, che in questo caso consiste, appunto, nel numero di elementi, e che sarebbe l’analogo, in linguistica, della dipendenza morfo-sintattica fra gli elementi di un enunciato, regolata dai principi ricorsivi. Tale esperimento spiegherebbe perché tale specie manca della capacità di acquisire una lingua naturale, la cui struttura prevede appunto una complessità sintattica gerarchica che pone gli elementi che occorrono al suo interno in una relazione logica, tale che la determinazione della correttezza formale di una frase non può prescindere dal valore delle unità che vi occorrono. In conclusione, affermano gli autori, l’abilità di riconoscere tale tipo di

---

<sup>29</sup> Vedi *Ivi*, p. 378: “the  $A^nB^n$  grammar produces center-embedded constructions that, although less common in human language than other (e.g. right-branching) structures, are ubiquitous in mathematics (e.g., nested parentheses in formulas) or computer programming languages (e.g., BEGIN-END statements).”

<sup>31</sup> *Ivi*, pp. 379-80. Vedi. Hauser M. D., Chomsky N., Fitch W. T., *Op. cit.*, p. 1576: “It seems relatively clear, after nearly a century of intensive research on animal communication, that no species other than humans has a comparable capacity to recombine meaningful units into an unlimited variety of larger structures, each differing systematically in meaning.”

strutture logico-sintattiche sarebbe il tratto che segna la divergenza cognitiva fra scimmie - nel caso specifico, fra tamarini - ed esseri umani<sup>84</sup>.

### 4.3 Percezione di grammatiche artificiali negli uccelli canori

Sorprendentemente, sebbene i tamarini abbiano fallito nella individuazione di strutture gerarchiche ricorsive, da uno studio condotto nel 2006 dal neuroetologo Timothy Gentner<sup>85</sup>, emerge che una specie di passerai (*Sturnus vulgaris*), esposti allo stesso tipo di stimoli usati da Fitch e Hauser, sia in grado di individuare la regola logica sottesa nelle grammatiche fraseologiche, e persino di generalizzarne l'applicazione su stimoli diversi. Gentner e i suoi collaboratori hanno creato due insiemi di sequenze sottese rispettivamente dalla grammatica a stati finiti e da quella "libera dal contesto", generando le *a* e le *b* a partire da due tipologie di unità che formano i loro canti, i cosiddetti "rattle" e i "warble".

---

<sup>84</sup> Conway, C., *Sequential learning in non-human primates*, "Trends in Cognitive Sciences", 5, 2001, pp. 539-546.

<sup>85</sup> Gentner, T. Q., Fenn, K. M., Margoliash, D., Nusbaum, H. C., *Recursive syntactic pattern learning by songbirds*, "Nature", 440, 2006, pp. 1204-1207. Sebbene gli stimoli a cui le due specie sono state esposte erano i medesimi, la procedura sperimentale consisteva nel metodo della "violazione delle aspettative" nei tamarini e nel metodo "go/no go" nei passerai. Ciò vuol dire che la capacità di distinguere le due grammatiche nei tamarini è stata studiata abituando gli individui ad un certo tipo di stimoli emessi da un altoparlante, per poi notare la loro reazione all'esposizione di stimoli diversi. L'idea è che se sono in grado di discernere la differenza tra le due classi di stimoli a cui sono sottoposti, una volta esposti ad un tipo di stringhe a cui non sono stati abituati vi presteranno più attenzione. Nello specifico, ciò viene misurato quantificando il tempo in cui i tamarini guardano nella direzione dell'altoparlante. Viceversa, se non sono in grado di individuare la differenza tra due classi di stimoli, allora guarderanno per quantità di tempo non significativamente diverse ad entrambi i tipi di stringhe. Nel caso dei passerai, Gentner *et al.* hanno adottato invece la procedura "go/no go", in cui gli uccelli hanno imparato a reagire agli stimoli corretti (rinforzati) e a non reagire a quelli non corretti.

Tuttavia, come gli stessi studiosi osservano, il “successo” dei passeri nella distinzione dei due insiemi di stimoli può essere imputabile a tre ulteriori fattori che poco hanno a che fare con l’individuazione dello schema sintattico; è possibile, infatti, che gli uccelli stessero: a) semplicemente memorizzando i suoni complessivi delle stringhe, senza badare al pattern sintattico, b) facendo caso solo alla presenza della adiacenza di un qualsiasi numero di  $a$  consecutive seguite da un qualsiasi numero di  $b$  - c) notando unicamente la presenza di coppie iniziali di “AA” (o finali di “BB”) per distinguere la classe di  $A^nB^n$  e la presenza di le coppie di “AB” per riconoscere i pattern a stati finiti, oppure le transizioni B/A (assenti nei pattern  $A^nB^n$ , in cui all’ultima  $a$  segue la sequenza delle  $b$ ), o A/B (i pattern  $A^nB^n$  ne presentano soltanto una, nel punto in cui si passa dalla sequenza delle  $a$ , alla sequenza delle  $b$ ). Attraverso accurati test di controllo<sup>86</sup>, i ricercatori hanno escluso ogni ipotesi alternativa alla loro tesi centrale, dimostrando così che i passeri hanno appreso la regola di organizzazione gerarchico-ricorsiva  $A^nB^n$ , o in generale, una regola ad essa funzionalmente equivalente, traendone le seguenti conclusioni teoretiche:

---

<sup>86</sup> Nello specifico gli studiosi hanno creato delle sequenze di diversi “rattle” e “warble” per escludere l’ipotesi che i passeri stessero utilizzando dei meccanismi di mera memorizzazione fonica; per controllare ce non stessero meramente riconoscendo l’adiacenza di qualsiasi numero di  $a$  e  $b$ , hanno creato degli stimoli in cui il numero delle  $a$  non combaciava con quello delle  $b$ ; hanno poi escluso la terza ipotesi analizzando i dati statistici relativi ai pattern di risposta ai diversi test. A tal proposito, vedi Ivi, pp. 1205-1206.



Although uniquely human syntactic processing capabilities, if any, may reflect more complex context-free grammars or higher levels in the Chomsky grammatical hierarchy, it may prove more useful to consider species differences as *quantitative rather than qualitative* distinctions in cognitive mechanisms. Such mechanisms (for example, memory capacity) need not map precisely onto strict formal grammars and automata theories. There might be no single property or processing capacity that marks the many ways in which the complexity and detail of human language differs from non-human communication systems<sup>87</sup>.

Stando a tali conclusioni, dunque, la divergenza tra il linguaggio umano e quello delle altre specie animali potrebbe fondarsi su diverse abilità cognitive oltre a quelle meramente sintattico-computazionali. È inoltre possibile, secondo l'autore, che tali proprietà differiscano meramente in termini quantitativi (ad esempio, in termini di memoria necessaria a processare le informazioni), ma non qualitativi. A questo punto, è bene sottolineare come la ricerca di Gentner, seppure molto suggestiva, presenti delle falle metodologiche che, se da una parte potrebbero minarne la validità scientifica, dall'altra offrono degli spunti per la progettazione di nuovi esperimenti in grado di sospingere la conoscenza delle abilità sintattiche delle specie non umane (e allo stesso tempo di quelle degli umani stessi) in avanti. Vale la pena, dunque, di prendere in esame i dettagli dello studio problematici da un punto di vista concettuale. A tale proposito, troviamo molto interessanti le osservazioni di Micheal

---

<sup>87</sup> *Ivi*, p. 1206 (corsivo mio).

Corballis<sup>88</sup>, il quale vede gli stimoli utilizzati da Gentner (quindi anche da Fitch e Hauser nello studio sui tamarini) non strutturati in funzione del pattern della ricorsività nidificata al centro, bensì come un mero esempio di doppia iterazione (Fig. 6):

To decide whether a given sequence obeys the rule, the birds need simply count the number of successive As, and then the number of successive Bs, and accept the sequence as belonging to  $A^nB^n$  if the numbers are equal. This structure does not conform to a FSG, but does not involve center-embedded recursion, and does not require place-holders. However it does require a counting mechanism, and memory for the number of elements.

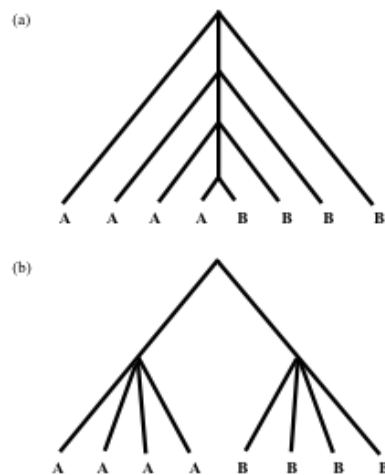


Fig. 1. Tree structure for (a) center-embedded recursion, and (b) double iteration.

**Fig. 6.** Immagine modificata da Corballis (2007)

Inoltre, Corballis mostra una serie di possibili meccanismi cognitivi alternativi in grado di spiegare “il successo” dei passeri di Gentner come la capacità di “subitazione”, ovvero un’estimazione immediata della quantità

<sup>88</sup> Corballis, M. C., *Recursion, language, and starlings*, “Cognitive Science”, 31, 2007, pp. 697-704.

di stimoli di un certo tipo – che non richiede alcun processo di calcolo esplicito. Oppure, è possibile che abbiano una strategia di riconoscimento di un pattern strutturale generale, che nulla a che fare con l'abilità di riconoscere le relazioni di interdipendenza fra gli stimoli, necessarie a fare di uno schema un pattern ricorsivo nidificato al centro. Per dimostrare ciò, l'autore suggerisce un approccio sperimentale diverso:

An alternative nonlinguistic approach, perhaps appropriate for further testing with the starlings, might be to establish connections between AB pairs, and then test whether the birds can still recognize pairings when other pairs are nested within them. For example, if  $A_iB_i$  are pairs for all  $i$ , then sequences like  $A_iA_jB_jB_i$ , or  $A_iA_jA_kB_kB_jB_i$  obey the rule, whereas any other combinations do not. [...] To demonstrate parsing of  $A_nB_n$  sequences based on center-embedded recursion, it would be necessary to show that elements are “bound” from the outside in, such that the outermost elements are correctly associated, the next-to-outermost elements correctly associated, and so on. The article by Gentner et al. does demonstrate the ability of starlings to discriminate sequences based on a CFG [context free grammar], but does not necessarily demonstrate any understanding of the recursive nature of center-embedded sequences.<sup>89</sup>

Proprio sulla base di tali riflessioni critiche, Abe e Watanabe hanno realizzato uno studio di ricerca<sup>90</sup> sui passeri del Giappone, modificando artificialmente il pattern di sillabe tipiche dei loro canti. In particolare, obiettivo degli studiosi era di studiarne la capacità di processare: a) specifici pattern *sintattici*, ovvero sequenze lineari di sillabe, b) dipendenze

---

<sup>89</sup> *Ivi*, pp. 702-703.

<sup>90</sup> Abe, K., Watanabe, D., *Songbirds possess the spontaneous ability to discriminate syntactic rules*, “Nature Neuroscience”, 14, 2011, pp. 1067–1074.

fra elementi non adiacenti, e, interessante, c) strutture ricorsivamente nidificate.

Relativamente al primo punto, gli studiosi hanno sostituito alcune sillabe nell'ordine originale delle sillabe dei canti (figura 7), osservandone la reazione sulla base del paradigma delle violazioni delle aspettative. In altri termini, per ciascuna delle tre ipotesi, i ricercatori hanno sottoposto i passeri ad una fase di abituazione, lasciandoli ascoltare per due ore consecutive i loro stessi canti attraverso un amplificatore, per poi presentare una versione modificata dei canti, in cui talune sillabe adiacenti venivano sostituite con delle nuove unità in base a specifiche regole di sostituzione. Il parametro utilizzato per attestare la loro capacità di discriminare le sequenze della fase di abituazione da quelle della fase del test era il numero di vocalizzazioni prodotte in risposta allo stimolo, comportamento tipicamente elicitato nei primi minuti in cui sentono un canto. Dal test sull'ordine sequenziale delle sillabe è risultato che gli uccelli reagiscono significativamente alla sostituzione di specifiche sillabe poste in un ordine adiacente, ma non ad altre:

Artificial modification of syllable sequences might create novel syllable transitions that did not appear in the original songs. It is therefore possible that the subjects responded to those uncommon syllable transitions. To analyze such possibility, we tested the subjects with further SM songs that contained the novel syllable transitions found in SEQ2: J→B and L→G transitions [...]. The J→B transition also appeared in SEQ2D songs, but did not cause significant responses [...]. The L→G transition in SEQ1 also did not cause significant responses [...].

Further analysis of systematic modification of syllable sequence also failed to find a simple linear sequence that the birds might use to discriminate the sequence-modified songs [...]. These results indicate that the differential behavioral response to SM songs is not a consequence of a simple response to local linear sequence of syllables.<sup>91</sup>

Original JOLENEKE JOGLENEKE JOOL  
 SEQ1 JOLENEKE JOGLENEKE JOOL  
 SEQ2 JOLENEKE JOGLENEKE JOOL  
 SEQ3 JOLENEKE JOGLENEKE JOOL  
 SEQ2B JOLENEKE JOGLENEKE JOOL  
 SEQ2C JOLENEKE JOGLENEKE JOOL  
 SEQ2D JOLENEKE JOGLENEKE JOOL

**Fig. 7.** Immagine modificata da Abe (2011)

Così, sulla base di tali osservazioni, Abe e Watanabe hanno condotto un ulteriore test sperimentale, in cui l'ordine delle sillabe era governato da una grammatica artificiale del tutto arbitraria, in cui le sillabe *sintatticamente* connesse erano poste in posizioni non adiacenti, al fine di esplorarne l'abilità di estrarre le regole di dipendenza strutturale. Da tale esperimento è emerso che effettivamente i soggetti erano in grado di riconoscere tale tipo di regolarità sintattica, discriminando quei canti che non ne presentavano il relativo pattern.

Infine, gli autori si sono spinti ancora oltre, realizzando un esperimento al fine di testare la presenza in tale specie di passeri, di

---

<sup>91</sup> Ivi, p. 1068-1069.

un'abilità ritenuta appannaggio cognitivo specifico della specie umana<sup>92</sup>: la capacità di annidare ricorsivamente talune strutture entro se stesse, ovvero di discriminare le grammatiche “libere dal contesto”:

To analyze whether Bengalese finches could learn and distinguish syllable strings obeying such complex grammar, we synthesized artificial syllable strings obeying center-embedding context-free grammar (center-embedding language) and test strings violating its rules [...]. Center-embedding language resembles the structure of the Japanese language, in which a subject-object-verb (S1-O1-V1) phrase can be embedded into another S2-O2-V2 sentence, creating a S2-S1-O1-V1-V2 sentence [...]. To ensure that the birds actually distinguished the embedded structures, we designed the strings to have a pair-wise relationship between the elements (for example, AX-FX pair). Each AX and FX pair always co-occurred in a string; the appearance of F1 [...] is expected by the appearance of A1. [...] Again, test strings obeying the correct grammatical rule did not cause a significant response [...], but those violating the center-embedding syntactic rule caused a significant decrease in the call responses [...]. Together, these data indicate that Bengalese finches were able to learn syntactic rules of syllables by listening to auditory information and discriminate novel auditory information according to these rules. *They even correctly distinguished syntactically complex information such as syllable strings created according to the rule of context-free grammar.*<sup>93</sup>

Stando a quanto osservato dai ricercatori, dunque, i passeri del Giappone sarebbero sensibili - alla stregua degli esseri umani - a pattern che presentano una struttura ricorsiva. Tale affermazione, seppur molto forte e sorprendente per quanto concerne la prospettiva comparativa sulla cognizione animale merita un'opportuna analisi critica. Crediamo, infatti, che lo studio di Abe e Watanabe, sebbene mostri *importanti* competenze sintattiche dei passeri del Giappone, sia lungi dal mostrarne l'abilità

---

<sup>92</sup> Hauser, M. D., Chomsky, N., Fitch, T., Op. cit.

<sup>93</sup> Abe, K., Op. cit., p. 1070 (corsivo mio).

precipuamente umana di annidare ricorsivamente strutture entro se stesse. Invero, è importante osservare che, in una stringa del tipo “ $A_X-A_Y-C_Z-F_Y-F_X$ ” la connessione tra gli elementi “appaiati”, come “ $A_X-F_X$ ”, o “ $A_Y-F_Y$ ” è del tutto arbitraria, e rinforzata sul piano meramente percettuale. Inoltre, a nostro avviso, l’ordine delle sillabe rispecchia una struttura complessa, in cui gli elementi presentano sì delle connessioni, ma queste, pur non dispiegandosi secondo una mera sequenzialità di adiacenze lineari, sono nondimeno collocate su un livello privo di quella profondità distintiva delle strutture gerarchiche del linguaggio umano. Una profondità che, dispiegandosi in funzione della natura semantica delle categorie, determina la specie-specificità del linguaggio umano. Ciò vuol dire che la ricorsività delle strutture proposizionali non è districabile dalla precipua natura categoriale delle unità articolabili all’interno di una proposizione dotata di senso, e per questo motivo, le pur complesse strutture che i passeri sono in grado di discriminare non sono assimilabili alle medesime strutture ritenute appannaggio cognitivo esclusivo della specie umana.

In tale direzione, vanno altresì le significative osservazioni di van Heijningen<sup>94</sup>, secondo la quale, per appurare che le specie non umane abbiano la abilità di discriminare strutture ricorsive nidificate al centro,

---

<sup>94</sup> Heijningen, van, C., Visser, de, J., Zuidema, W., Cate, de, C., *Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species*, “Proceedings of the National Academy of Sciences”, 106., 2009, pp. 20538-20543.

bisogna specificare il tipo di relazione che intercorre fra gli stimoli che sebbene temporalmente (o spazialmente) distanti, sono “appaiati”. Infatti, le connessioni che hanno luogo nella realizzazione di una proposizione hanno carattere anzitutto logico-concettuale, legato al significato dei termini fra loro connessi. I suoni-stimolo utilizzati (tanto da Fitch e Hauser, quanto da Gentner) mancano di una simile caratterizzazione logico-concettuale, in quanto basati su similarità di tipo meramente percettuale, come il tipo di voce (femminile o maschile) o i tipi di note (“rattle” vs. “warbles”). Così ne parla la studiosa:

But while the discussion on the experiment in terms of whether it is an appropriate test of the presence of recursion lingers on, a more basic question about what the starlings have actually learned has received little attention [...]: did they learn an abstract rule allowing them to generalize from familiar to novel stimuli? Gentner et al. [...] showed that starlings generalized from familiar rattle and warble structures to unfamiliar ones, but not that they generalized to other phrase types as well. Therefore, the birds could also have based their discrimination on a phonetic categorization, restricted to rattle and warble phrase types. That is, they might classify all “rattle-warble-rattle-warble” sequences as one type, different from “rattle-rattle-warble-warble” sequences, based on the overall acoustic similarity alone [and with the appropriate similarity metric, this would generalize to larger n without a need for a syntactic analysis [...]. So the question whether the birds have the abstract computational abilities that enable them to discriminate between the two stimuli sets strictly based on order information is still open.<sup>95</sup>

Alla luce di tali importanti critiche alla metodologia adottata negli studi comparativi qui presi in esame (e di conseguenza, alle presupposizioni teoriche che vi stanno alla base), vale la pena di chiedersi:

---

<sup>95</sup> *Ivi*, p. 20539.



a) in quale misura l'approccio "formale" sia in grado di rivelarci la reale entità delle proprietà cognitive che fondano i sistemi di comunicazione animale, e b) quale valore epistemologico sia effettivamente attribuibile alla riduzione dello studio sulla facoltà di linguaggio ad un mero agglomerato più o meno complesso di operazioni computazionali che prescindono dal valore semantico delle unità *sintatticamente* connesse. Nel prossimo capitolo metteremo a fuoco l'analisi su quest'ultimo punto, essenziale, a nostro avviso, per aumentare il livello di lucidità teorica degli studi comparativi qui presi in esame.

## **5. Comprensione dei pattern sintattici, ovvero la co-determinazione tra sintassi e uso**

### **5.1 Vedere una regola: comprensione o categorizzazione percettuale?**

Come abbiamo visto nel capitolo precedente, un meccanismo fondamentale alla base dell'acquisizione di una lingua storico naturale da parte degli infanti (a partire dalla esposizione sensoriale ad essa) - e delle grammatiche artificiali da parte delle specie non umane - è l'abilità di riconoscere dei pattern strutturali e di generalizzarne l'applicazione a stimoli *percettualmente* diversi. Ad esempio, come emerge dallo studio di Marcus *et al.*<sup>96</sup>, gli infanti pre-verbali sono perfettamente in grado di generalizzare il pattern ABA sotteso alla stringa “*fe wo fe*”, ad una stringa percettualmente diversa come “*gi ni gi*”. Da tale esempio emerge chiaramente come la capacità di generalizzazione di una regola possa agire

---

<sup>96</sup> Marcus, G., Vijayan, S., Rao, S. B., Vishton, P. M., *Rule Learning by Seven-Month-Old Infants*, cit.

a partire da un livello base che è quello meramente percettuale. Gli infanti sono cioè in grado di riconoscere operazioni logiche del tipo “la prima e l’ultima sillaba della sequenza sono *fisicamente identiche*”, ed applicare tale schema a qualsiasi sequenza finita di sillabe. Tuttavia, la relazione logica di identità, così come, ad esempio, le relazioni “maggiore, o minore di” può tradursi in un dominio non immediatamente percepibile al livello sensoriale. Ad esempio, l’identità tra “Alessandro” e “Daniele” si pone *non* ad un livello fisico, in quanto nè la realizzazione fonica delle parole, nè le persone designate dai nomi si somigliano fra loro; lo statuto di identità tra le due unità si pone invece ad un livello più astratto, consistente nella sussunzione dei due elementi entro un’unica categoria, che abbraccia in sé diversi riferimenti: la categoria dei nomi<sup>97</sup>. Lo stesso vale per “correre” e “giocare”, appartenenti alla medesima categoria “azione”.

Dunque, alla luce di tale distinzione, possiamo affermare che ciò che permette ai bambini di orientarsi nel linguaggio, non è *sic et simpliciter* l’abilità di riconoscere due istanze in quanto identiche, come afferma Marcus; l’acquisizione del linguaggio nei bambini, presuppone, infatti, in primo luogo la capacità di riconoscere due differenti stringhe di suoni in quanto rappresentazioni foniche di oggetti appartenenti alla medesima

---

<sup>97</sup> Vedi Gómez, R., *Infant artificial language learning and language acquisition*, “Trends in Cognitive Sciences”, 4, 2000, pp. 178-186.

categoria e, in secondo luogo, di riconoscere le frasi “Marco legge un libro” e “Alice vede un cane” come istanze dello stesso pattern strutturale “nome-verbo-nome”, nonostante le evidenti differenze esperibili sul piano sensoriale tra “Marco” e “Alice”, o tra unità foniche “cane” e “libro”. Così, negli insiemi delle variabili A ed B usati da Marcus, gli elementi non sono rispettivamente sussumibili entro una categoria comune, in quanto non presentano alcuna similarità categoriale; inoltre, la relazione di dipendenza tra le due classi non fa capo ad alcuna connessione di ordine logico-sintattico o algebrico, ma opera su un livello meramente percettuale, ovvero di identità o differenza sul piano fonico. Al contrario, nel linguaggio naturale, è l’effettiva natura della classe categoriale a determinare la correttezza del pattern, e ciò rappresenta un aspetto essenziale ed in quanto tale non trascurabile negli studi empirici volti ad esplorare le dinamiche dell’acquisizione del linguaggio. Alla luce di tali considerazioni, è opportuno, a questo punto, definire con chiarezza i confini di due abilità cognitive, spesso sovrapposte nella letteratura scientifica intorno allo studio dello sviluppo onto- e filo-genetico del linguaggio: si tratta dell’abilità di *percepire* le unità costitutive di una lingua storico-naturale (o di un linguaggio artificiale), che va metodologicamente distinta dalla abilità di *comprenderne* il significato all’interno del contesto fraseologico, e pragmatico.

A tale proposito, consideriamo, ad esempio, una famosa frase inventata da Noam Chomsky<sup>98</sup> al fine di enfatizzare la natura innata delle abilità sintattiche: “Colorless green ideas sleep furiously”. Tale frase *suona* come una frase ben formata, nonostante sia completamente priva di senso (non comprensibile). In altre parole, siamo anzitutto perfettamente in grado di distinguerne le unità costitutive - e ciò, come abbiamo visto, fa capo a delle abilità di computazioni statistiche distribuzionali - e di associare ogni termine ad una specifica categoria - aggettivi, nomi, verbi, etc; tuttavia, le nostre abilità, così come quelle di un infante non si limitano a questo. Crucialmente, infatti, siamo altresì in grado di percepire il pattern strutturale che connette i diversi termini in quanto corretto, per quanto non ne comprendiamo il significato. Vale la pena di chiedersi, dunque, quale meccanismo sottenda questa ultima abilità.

Ancora una volta, crediamo sia necessario chiamare in causa le abilità di computazioni statistiche, le quali operano su diversi livelli del linguaggio. Fin qui abbiamo preso in esame quello fonotattico e della sequenza di sillabe, ma a ben guardare, le computazioni statistiche operano oltre che sul piano delle unità lessemiche ben formate, anche sul piano dei suffissi e delle radici che ne determinano la morfologia sintattica e, di conseguenza, sul piano delle categorie grammaticali astratte. È in virtù di

---

<sup>98</sup> Chomsky, N., Three models for the description of language, cit., p. 116.

tale meccanismo, infatti, che percepiamo come ben formata (secondo le regole grammaticali della lingua italiana) una frase del tipo “Giulia nuota in piscina” piuttosto che “piscina Giulia in nuotamente”. In tale direzione vanno le osservazioni di Mark Seidenberg, il quale, prendendo spunto proprio dell’esempio classico delle “colorless ideas”, afferma:

But language exhibits structure at multiple levels, each of which has its own statistical character. The “colorless” sentence is less puzzling when one looks beyond transition probabilities to other information that is used in comprehension. For example, words fall into general types—green is a property or adjective, and sleep is an action or verb—that exhibit characteristic distributions. The “colorless” sentence conforms to these distributions in English, whereas “Ideas colorless sleep furiously green” does not.<sup>99</sup>

In altri termini, sono esattamente le proprietà statisticamente distribuzionali dei suoni della lingua (o delle lingue) a cui veniamo esposti a determinare la capacità di *percepire* una frase come ben formata, pur non comprendendola<sup>100</sup>. Chiarito il meccanismo responsabile della percezione della correttezza formale della frase in questione, rimane, a questo punto, da esplorare la ragione per cui la comprensione sul piano semantico ne è compromessa.

---

<sup>99</sup> Seidenberg, M. S., MacDonald, M., *Does grammar start where statistics stop?*, “Science”, 298, 2002, p. 554.

<sup>100</sup> Tuttavia, è opportuno sottolineare che le computazioni statistiche che operano ai diversi livelli della frase presentano differenti ordini di complessità. A tal proposito vedi Heinz, J., Idsardi, W., *Sentence and Word Complexity*, “Science”, 333, 2011, pp. 295–297.

## 5.2 Scheletri computazionali: similarità e differenze logico-sintattiche tra regole percettuali, algebriche e fraseologiche

Prendiamo in considerazione, ancora una volta, la ricerca condotta da Marcus.<sup>101</sup> Come abbiamo visto nel paragrafo 4.2, lo studioso ed i suoi collaboratori enfatizzano l'idea secondo la quale l'abilità propria degli infanti pre-verbali di applicare uno stesso pattern strutturale su elementi di natura diversa - come ad esempio discriminare l'identità fra la prima e l'ultima unità di una stringa di suoni – trovi un corrispettivo nel dominio algebrico:

For instance, we can substitute any value of  $x$  into the equation  $y = x + 2$ . Similarly, if we know that in English a sentence can be formed by concatenating any plural noun phrase with any verb phrase with plural agreement, then as soon as we discover that “the three blickets” is a well-formed plural noun phrase and that “reminded Sam of Tibetan art” is a well-formed verb phrase with plural agreement, we can infer that “The three blickets reminded Sam of Tibetan art” is a well-formed sentence.

Dunque, l'idea che possiamo applicare un pattern grammaticale che determini la concordanza tra nome e verbo (qualsiasi essi siano) troverebbe, secondo gli autori, un corrispettivo nei principi algebrici che regolano le espressioni del tipo  $x = y + 2$ , in cui uno stesso algoritmo può

---

<sup>101</sup> Marcus, G., *et al.*, *Rule Learning by Seven-Month-Old Infants*, cit.

essere implementato su numeri diversi. Tale osservazione cattura un aspetto molto importante, ovvero l'analogia funzionale tra i domini linguistico ed algebrico. Tuttavia, ancora una volta gli studiosi mancano di discriminare delle caratteristiche non trascurabili. È importante sottolineare che a differenza della applicazione di una regola sul piano meramente percettuale – quale è il caso degli stimoli adottati da Marcus *et al.* (1999) - l'implementazione di un algoritmo, ovvero di una predicazione su un nome sortisce una *Bedeutung* significato (il valore dell'espressione) *che può avere una corrispondenza sul piano della realtà*, e aggiungerei, sul piano pragmatico. E, in effetti, è proprio quest'ultima proprietà ad assumere il ruolo chiave nel rendere comprensibile una mera stringa di suoni.

Tale affermazione può essere chiarita da quanto scrive Ludwig Wittgenstein nel *Tractatus Logico-Philosophicus*: “Der Konfiguration der einfachen Zeichen im Satzzeichen entspricht die Konfiguration der Gegenstände in der Sachlage”<sup>102</sup>. La combinazione *sintattica* delle parole all'interno della proposizione deve essere in accordo, in qualche modo, con una relazione logicamente possibile all'interno degli stati cosali del mondo. Non è logicamente possibile, infatti, che le idee siano allo stesso tempo

---

<sup>102</sup> Wittgenstein L., *Tractatus logico-philosophicus*, cit., § 3.21. Trad. It.: “Alla configurazione dei segni semplici nel segno proposizionale corrisponde la configurazione degli oggetti nella situazione.” Cfr. Strawson, P., *On Referring*, “Mind”, 1950, in P. Strawson, *Logico-Linguistic Papers*, Methuen, London, 1971; trad. it.: *Sul riferimento*, in A. Bonomi, *La Struttura logica del linguaggio*, Valentino Bompiani, 2001.



colorate e prive di colore, o che dormano e siano in uno stato attivo di furia. Ciò implica che possiamo parlare anche di oggetti che non esistono nel mondo reale, o attribuirvi azioni proprie ad altre categorie di oggetti - ma soltanto nella misura in cui ne collochiamo la descrizione all'interno di proposizioni le cui connessioni siano logicamente non contraddittorie. Per tornare, dunque, all'esempio della frase di Chomsky, possiamo affermare che essa appare priva di senso ed incomprensibile (almeno su un piano non metaforico) in quanto non trova alcun valore significazionale sul piano delle relazioni logicamente possibili fra gli stati di cose del mondo reale, e ciò mina il valore della correttezza formale delle combinazioni sintattiche che occorrono in essa. Al fine di chiarire ulteriormente tale importante assunto, vorrei riportare un significativo passo di Bertrand Russell:

An infant very soon learns to be differently affected by the statement "cats eat mice" from the way he would be affected by the statement "mice eat cats"; and not much later he learns to make one of these statements rather than the other. [...] It may be that one part of the environment is sufficient to cause one word, while the other is sufficient to cause another, but *it is only the two parts in their relation that can cause the whole sentence*. [...] Thus wherever sentences come in we have a causal relation between two complex facts, namely the facts asserted and the sentences asserting it; the facts as wholes enter into the cause-and-effect relation, which cannot be explained wholly as compounded of relations between their parts. Moreover, as soon as the child has learned to use correctly relational words, such as 'eat', he has become capable of being causally affected by a relational feature of the environment, which involves a new degree of complexity not required for the use of ordinary names. Thus the correct use of relational words, i.e. of sentences, involves what may be correctly

termed ‘perception of form’, i.e. it involves a definite reaction to a stimulus which is a form.<sup>103</sup>

Possiamo dunque affermare, con Russell, che è la precipua relazione tra le parti a stabilire il significato di una frase, ovvero il suo corrispettivo sul piano oggettuale delle connessioni di fatti. L’uso delle parole in un certo ordine, ovvero l’uso sintattico delle parole – determina la “percezione della forma” dei fatti del mondo, o, in altri termini, la comprensione del modo in cui sono fra loro inter-relati. In linea con tali idee, è bene riportare alcune osservazioni di Terrence Deacon, decisive per una comprensione più profonda della differenza tra i sistemi di comunicazione non verbale, e degli umani. Riprendendo la distinzione peirciana di indice, icona e simbolo, lo studioso sostiene che la mente degli animali non umani non vada oltre la “soglia simbolica”. Ciò vuol dire che gli animali non umani associano secondo un rapporto *uno a uno* un segnale ad un oggetto, in funzione di una connessione reale di natura fisica (rapporto indicale), o di una somiglianza strutturale (rapporto iconico), ma non sono in grado di associare i segni fra di loro, né di fare di un segnale l’oggetto di un altro segnale. Tal genere di associazioni, nello specifico è

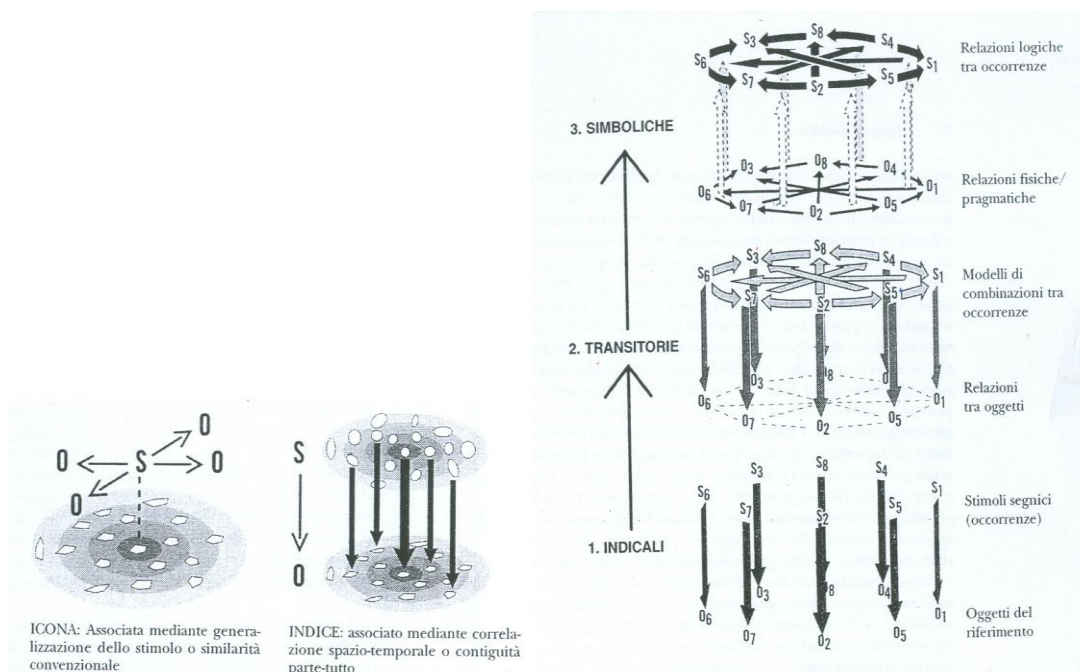
---

<sup>103</sup> Russell, B., *An outline of Philosophy*, Allen & Unwin, London 1927, in Egner R. E., Dennon L. E., (ed.), *The Basic Writings of Bertrand Russell*, Routledge Classics, London and New York, 2010, pp. 87-88 (corsivo mio).

ciò che fa - secondo il quadro teorico di Deacon - di un segnale una parola, ovvero un simbolo. Vediamo in che senso.

La corrispondenza tra parole e oggetti è una relazione secondaria, subordinata a una trama di relazioni associative di genere differente, che permette addirittura il riferimento a cose impossibili. [...]

Questa relazione referenziale tra le parole – parole che sistematicamente indicano altre parole – forma un sistema di relazioni di ordine superiore che consente loro di *implicare* relazioni indicali, e non solamente gli indici in sé. [...] Il loro potere indicale è, per così dire, *distribuito* nelle relazioni tra parole. *Il riferimento simbolico deriva da possibilità e impossibilità combinatorie, e perciò dipendiamo da combinazioni sia per scoprirlo (nell'apprendimento) sia per farne uso (nella comunicazione)*<sup>104</sup>.



**Fig. 1.** Immagini modificate da Deacon (1997)

<sup>104</sup> Deacon, T. W., *The Symbolic Specie. The coevolution of Language and the Brain*, Norton & Company, New York, 1997; trad. it. a cura di S. Ferraresi, *La specie simbolica. Coevoluzione di linguaggio e cervello*, Giovanni Fioriti, Roma, 2001, pp. 52, 64 (corsivo mio).

In tal prospettiva, possiamo osservare che il sistema comunicativo dei primati, e, in generale, degli animali non umani, consente di effettuare delle associazioni specifiche tra un lessigramma e un oggetto o evento specifico, o fra due eventi, e dunque, di prevedere secondo un rapporto di causalità diretta, quale evento seguirà ad un dato comportamento manifesto<sup>105</sup>. Diversa è invece la natura del sistema proposizionale simbolico degli uomini, la cui essenza consiste nella proprietà tale per cui «la relazione tra un lessigramma e un oggetto è *funzione* della relazione che intercorre con altri lessigrammi, e non solo della comparsa correlata di lessigramma e oggetto» (figura 2)<sup>106</sup>. Da un punto di vista evolutivo, gli effetti positivi di tale sistema di comunicazione derivano dalla possibilità in esso garantita, di fornire una notevole quantità di conoscenza implicita:

E poiché tale ricodificazione è fondata su relazioni di ordine superiore, e non sui dettagli, spesso semplifica enormemente il problema mnemonico e potenzia altrettanto le possibilità rappresentazionali. [...] Poiché le regole combinatorie non codificano oggetti, ma modi di correlarli, nuovi simboli possono venire immediatamente incorporati e combinati con altri fondati su una conoscenza indipendente da ciò che essi simbolizzano.<sup>107</sup>

Infine, una descrizione che mira a chiarire la natura dei meccanismi cognitivi dedicati alla percezione e alla comprensione di una frase non può

---

<sup>106</sup> *Ivi*, p. 68.

<sup>107</sup> *Ivi*, p. 71.

prescindere dalla considerazione di una sfera cruciale nella determinazione del senso: l'uso intersoggettivamente condiviso del linguaggio. Riporto a tal proposito un passo fondamentale delle *Ricerche Filosofiche*:

Per una *grande* classe di casi – anche se non per *tutti* i casi- in cui ce ne serviamo, la parola «significato» si può definire così: «Il significato di una parola è il suo uso nel linguaggio. E talvolta il *significato* di un nome si definisce indicando il suo *portatore*»

Ciò vuol dire che parlare di qualcosa non è semplicemente un'azione che riguarda la frase in se stessa, e le connessioni logico-sintattiche più o meno lecite che vi occorrono. Il parlare è un'attività propria degli uomini, che in quanto tale si intesse dei contesti socio-pragmatici. Oltre che da regole grammaticali, la comprensibilità di una frase è resa possibile dalle abitudini sociali e dalle convenzioni che ne determinano i significati:

Una volta, quando ero bambino, chiesi il significato della parola “qualcosa” [o era “forse”?]. mi risposero: “Non puoi ancora capirlo .” Ma come avrebbero potuto spiegarlo! Con una definizione? Oppure avrebbero dovuto dire che la parola è indefinibile? Non so come in seguito l'abbia imparato, ma certamente ho imparato ad applicare le frasi in cui ricorre la parola. E questo imparare somigliava quasi sempre a un addestramento (a un diventare esperto).<sup>108</sup>

---

<sup>108</sup> Wittgenstein, L., *The Big Typescript*, Springer-Verlag, Vienna, 2000; trad. it. di A. De Palma, *The Big Typescript*, Einaudi, Torino, 2002, II, 8. nota 4.

È possibile, come vedremo nei prossimi paragrafi, che l'effettiva connessione tra un termine e il suo uso avvenga attraverso un implicito calcolo distribuzionale tra le frasi e le diverse situazioni pragmatiche in cui occorre. Ciò che è importante sottolineare qui è che l'uso e la fruizione del linguaggio nei contesti situazionali sono in grado di stabilire quelle associazioni tra proposizioni e stati di cose tali da costituire un significato condiviso - quindi in linea di principio, comprensibile - al di là della effettiva correttezza formale dell'espressione stessa.

### **5.3 Strutture fraseologiche. Critiche al paradigma chomskiano**

Obiettivo del presente paragrafo è quello di tradurre nel paradigma chomskiano le idee esposte nei due precedenti paragrafi. Come abbiamo visto, infatti, è possibile individuare delle regole logico-sintattiche, che a diversi livelli di astrazione operano *nel* linguaggio, determinandone la specie-specificità. In termini chomskiani, è possibile descrivere la complessità strutturale tipica del linguaggio nel seguente modo

Natural languages go beyond purely local structure by including a capacity for recursive embedding of phrases within phrases, which can lead to statistical regularities that are separated by an arbitrary number of words or phrases. Such long-distance, hierarchical relationships are found in all natural languages for which, at a minimum, a “phrase-structure grammar” is necessary.<sup>109</sup>

---

<sup>109</sup> Hauser M. D., Chomsky N., Fitch W. T., Op. cit., p. 1577.

Dato tale assunto di fondo, crediamo sia opportuno muovere due critiche di base al paradigma chomskiano. In primo luogo, è necessario tematizzare in che misura sia corretto intravedere nella ricorsività nidificata - o, in generale, nelle strutture gerarchiche - una proprietà cruciale del linguaggio, se con tale termine ci riferiamo al linguaggio che si incarna nell'uso che ne fanno gli esseri umani, e non a quello utilizzato dai sistemi artificiali per l'elaborazione di informazioni. In effetti, crediamo non vada trascurato l'osservazione del fatto che gli umani mostrano significative difficoltà a comprendere una struttura annidata in cui il numero degli elementi dipendenti è maggiore a tre, che peraltro, raramente occorre nell'uso ordinario e non del linguaggio.

In secondo luogo, vale la pena di sollevare dei dubbi critici sulla identificazione del tratto distintivo del linguaggio umano (rispetto a quello delle altre specie animali) nella intrinseca natura ricorsiva delle sue strutture sintattiche<sup>110</sup>. In tale direzione vanno le critiche mosse da Pierre Perruchet e Arnand Rey proprio agli esperimenti condotti all'interno di tale paradigma di ricerca. Gli autori affrontano tale questione in un articolo dal titolo emblematico: "Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates?"<sup>111</sup>. Vale la pena

---

<sup>110</sup> Hauser M.D., *et al.*, The faculty of language: what is it, who has it, and how did it evolve?, cit.; Fitch W. T., *et al.*, *The evolution of the language faculty: clarifications and implications*, cit.

<sup>111</sup> Perruchet, P., Rey, A., *Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates*, "Psychonomic Bulletin & Review", 12, 2005, pp. 307-313.

di riportarne un passo significativo:

[...] is it still worthwhile to pursue experimental investigations about tamarins' abilities to master PSG grammars? A positive response would entail that such grammars are (1) commonly mastered by humans and (2) actually crucial for describing human language structures. We believe that neither of these conditions is met. [...] Let us add simply one more embedding to our initial sentence, and we get the oft-cited "the rat the cat the dog chased ate stole the cheese," which is unintelligible to most English speakers, as Miller and Chomsky (1963) themselves noted about a very similar example [...].

[...] Let us assume, for the sake of argument, that F&H [Fitch & Hauser] (2004) were right in claiming that humans easily master a PSG, whereas tamarins can only master an FSG. This would explain, at best, why tamarins fail to process a sentence such as "the rat the cat ate stole the cheese," but this would not account for their failure to process "the cat ate the rat that stole the cheese" or "the dog chased the cat that ate the rat that stole the cheese." Indeed, these right-branching structures are normally generated by an FSG, which is easily mastered by tamarins, according to F&H. Clearly, tamarins do not fail to master only center-embedded sentences. They fail to master even the most elementary linguistic utterance.<sup>112</sup>

Così, se vogliamo aderire al paradigma chomskiano, dovremmo far rientrare le frasi ramificate a destra (ad esempio, "il cane che rincorse il gatto che mangiò il topo si chiama Fido") nella classe delle strutture generabili da una grammatica a stati finiti. E ciò sortirebbe immediatamente un paradosso metodologico: data, infatti, la dimostrazione che i tamarini e i passerini siano in grado di riconoscere tale tipo di strutture, bisognerebbe spiegare perché non hanno la minima comprensione di quelle frasi del linguaggio umano in linea di principio generabili dalla più semplice delle grammatiche a stati finiti.

---

<sup>112</sup> *Ivi*, p. 311.



In conclusione, vale la pena di considerare un'ultimo punto debole del paradigma chomskiano, che consiste nella definizione di complessità, intesa nei termini della quantità di memoria necessaria a processare le informazioni del dato pattern. Il fatto, ad esempio, che i tamarini studiati da Fitch e Hauser (2004)<sup>113</sup> fallissero nell'individuazione di una struttura del tipo (AB)<sup>n</sup> di certo non può essere imputabile alla abilità di trovare connessioni logiche tra le *a* e le corrispettive *b*, che nota Corballis (2007) non presentavano alcuna dipendenza interna, nè è imputabile *tout court* alle loro abilità mnemoniche. Queste ultime, infatti, si manifestano a livelli di complessità talvolta molto più elevati di quelli umani (ma in domini diversi da quello inerente la mera esposizione a suoni non con-specifici) come ad esempio la navigazione spaziale, o l'ordinamento seriale di simboli grafici

Alla luce di tali importanti critiche al paradigma chomskiano, in generale - e alla metodologia adottata negli studi comparativi qui presi in esame, in particolare - vale la pena di chiedersi: a) in quale misura l'approccio "formale" sia in grado di rivelarci la reale entità delle proprietà cognitive che fondano i sistemi di comunicazione animale, e b) quale valore epistemologico sia effettivamente attribuibile alla riduzione dello studio sulla facoltà di linguaggio ad un mero agglomerato più o meno

---

<sup>113</sup> Fitch, W. T., Hauser, M. D., *Computational Constraints on Syntactic Processing in a Nonhuman Primate. Science*, cit.

complesso di operazioni computazionali. In altri termini, sebbene ad oggi esso sia l'unico approccio implementabile in disegni sperimentali da applicare sullo studio delle facoltà *sintattiche* degli animali, crediamo sia opportuno rivedere l'appropriatezza di alcuni assunti relativamente allo studio del linguaggio effettivamente *usato* dagli individui umani, e, alla luce di tali considerazioni, correggere l'approccio sperimentale comparativo. In tale direzione, vale la pena di chiedersi se esistono dei meccanismi cognitivi analoghi, o omologhi che, seppur non riducibili in termini computazionali, rappresentino un nucleo bio-cognitivo comune, condiviso di accesso alla realtà, il cui esame possa affiancare lo studio delle abilità sintattiche delle specie animali. Per rispondere a tale quesito, e allo stesso tempo, guadagnare una effettiva comprensione delle abilità sintattiche in grado di rivelarci la specificità della cognizione umana, è opportuno e consigliabile, a nostro avviso, cercare di ottenere un compromesso metodologico tra la traduzione in termini formali delle sue proprietà logico-sintattico, e la dimensione *naturale* del suo uso ordinario.

In tale direzione, troviamo conveniente mantenere l'approccio dell'apprendimento statistico che - da una parte rende ben conto, da un punto di vista epistemologico, della naturale predisposizione degli esseri umani a trovare regolarità relazionali fra gli stimoli dell'ambiente circostante, incluso il linguaggio vocale - dall'altra, trova un corrispettivo

ben esaminabile fra le abilità cognitive degli animali non umani. A nostro avviso, ciò sarebbe certamente utile ad alimentare la validità epistemologica di uno studio comparativo. Per quanto concerne la *naturalità* del linguaggio, crediamo che un aspetto fondamentale del linguaggio - comune tanto agli umani quanto alle scimmie e agli uccelli - sia la caratterizzazione prosodica delle articolazione di suoni, che prescinde dalla diversità inter-specifica dei tratti segmentali delle vocalizzazioni. Di fatti, la linea intonazionale, sebbene spesso non opportunamente sottolineata nell'analisi comparativa fra i diversi sistemi di comunicazione, è un parametro essenziale, in grado di veicolare significati che portano in sé un colore emozionale (con valore spesso altamente informativo) da cui nessuno scambio comunicativo nel mondo animale prescinde.

Nella seconda sezione del presente lavoro, tematizzeremo come la caratterizzazione prosodica e i *pattern* strutturali delle articolazioni vocali si intrecciano dando luogo ad espressioni la cui complessità non è riducibile alle singole parti costitutive, ma emerge dalla enunciazione del proferimento considerato nella sua totalità. In tale analisi, prenderò in esame il ruolo della prosodia nell'origine evolutiva ed ontogenetica dell'abilità di produrre e comprendere il linguaggio verbale. Tuttavia, prima di analizzare tale importante aspetto del linguaggio, è opportuno guardare da vicino gli studi inerenti i sistemi comunicativi di talune specie

di scimmie e uccelli, al fine di acquisire una visione più consapevole delle abilità “spontanee” delle specie studiate. Gli studi sperimentali fin qui descritti dimostrano, infatti, che alcune specie di scimmie e uccelli canori possono sì acquisire l'abilità di *percepire* grammatiche complesse, ma ciò spesso richiede lunghe fasi di addestramento da parte degli umani. Inoltre, gli stimoli a cui le specie studiate vengono esposte non sempre sono di natura conspecifica, ma spesso consistono in sequenze foniche proferite dagli esseri umani stessi, o al massimo, create artificialmente al computer. Vale la pena, dunque, di esplorare da vicino le spontanee abilità di *produzione* sintatticamente regolata di messaggi comunicativi in talune specie di primati e di uccelli canori.

## **6. Sintassi: studi comparativi sulla produzione di pattern fonici**

### **6.1 Strutture fonologiche: requisiti anatomico-neurali**

A discapito della condivisione di più del 98% del patrimonio genetico (e della appartenenza alla medesima area filogenetica), le scimmie e gli umani presentano delle notevoli differenze sul piano cognitivo. Fra queste, due delle più importanti sono: la quasi totale assenza di controllo volontario sull'articolazione vocale dei suoni, e l'incapacità di organizzare gerarchicamente i segmenti vocali (consonanti e vocali) entro strutture di ordine superiore come parole o frasi. Ciò ci induce a pensare che l'abilità di *sintassi* non fosse ancora presente nell'ultimo antenato comune agli umani e ai primati non umani, vissuto circa 6-7 milioni di anni fa. Le ragioni di tale assenza sono di ordine sia anatomico che neurale.

La collocazione del tratto laringeo in una posizione alta, e la conformazione (in particolare l'estensione e la forma) del tratto vocale delle scimmie, di fatti, non permettono alla loro lingua quella libertà di movimento necessaria a modulare la vasta gamma di suoni e frequenze che invece caratterizza il linguaggio umano. A tal proposito, è curioso notare

che la posizione bassa della laringe, tipico della specie umana, è ravvisabile in diverse specie oltre all'uomo, e l'analisi delle convergenze evolutive che interessano le diverse specie in questione rivela che essa fornisca indizi efficaci (talvolta anche esagerati) sulle dimensioni del corpo del segnalante, con evidenti conseguenze nelle situazioni di conflitto fra gli individui. L'idea dunque, è che tale conformazione sia stata una conquista evolutiva dell'uomo, in seguito esattata per l'evoluzione del linguaggio:

One effect of a lowered larynx is to increase vocal tract length (and, consequently, to decrease formant frequencies). An animal with a lowered larynx can duplicate the vocalizations of a larger animal that lacks this feature, thus exaggerating the impression of size conveyed by its vocalizations. According to this 'size exag-geration' hypothesis, the original selective advantage of laryngeal lowering was to exaggerate size and had nothing to do with speech. [...] Once the larynx was lowered, the increased range of possible formant patterns was co-opted for use in speech.<sup>114</sup>

Oltre che a dei limiti di natura anatomica, l'impossibilità di articolare una vasta gamma di suoni nelle scimmie fa capo a delle caratteristiche di natura neurale su cui molto possono dirci i dati comparativi con le specie canore di uccelli (in particolare, dei passeri). Infatti, seppur filogeneticamente molto distanti dagli uomini, esse sono abilissime nel modulare suoni discreti secondo strutture *sintattiche* definite, quindi di controllarne l'articolazione da un punto di vista motorio.

---

<sup>114</sup> Fitch, W. T., *The evolution of speech: a comparative review*, "Trends in Cognitive Sciences", 4, 2000, p. 263.

## 6.2 Strutture proto-sintattiche fra le vocalizzazioni delle scimmie

Le abilità sintattiche fra i primati non umani, come abbiamo visto, sono estremamente limitate per ragioni legate tanto alla fisionomia del loro tratto laringeo-vocale, quanto alla rudimentale capacità neurale di controllo motorio. Tuttavia, la letteratura scientifica presenta diversi esempi di abilità sintattiche, che vale la pena di riportare. In questa direzione si collocano le ricerche di Peter Marler<sup>115</sup>, il quale riconosce degli esempi di sequenze sonore regolate nei *pant-hoots*, vocalizzi ritmati, emessi ad alto volume dagli scimpanzé o dai gibboni in situazioni emozionalmente intense come l'incontro fra gruppi, in seguito alla cattura di una preda, o per rimanere in contatto attraverso la foresta. È interessante notare che nonostante tali ululati siano diversi da soggetto a soggetto e che, anche nello stesso individuo, cambino di circostanza in circostanza, presentino delle costanti sequenziali. Una sequenza di *pant-hoots*, che in media dura dieci secondi, e che può vantare dai quindici ai trenta suoni distinti (caratterizzati come mugolii, urla, ululati), presenta infatti di base quattro parti strutturali: un'introduzione, una fase in crescendo, un climax, e un abbassamento del tono finale.

---

<sup>115</sup> Marler, P., *Origins of Music and Speech: Insights from Animals*, in N. L. Wallin, B. Merker, S. Brown, (ed.), *The Origins of Music*, Mit Press, Cambridge, 2001, pp. 31-48.

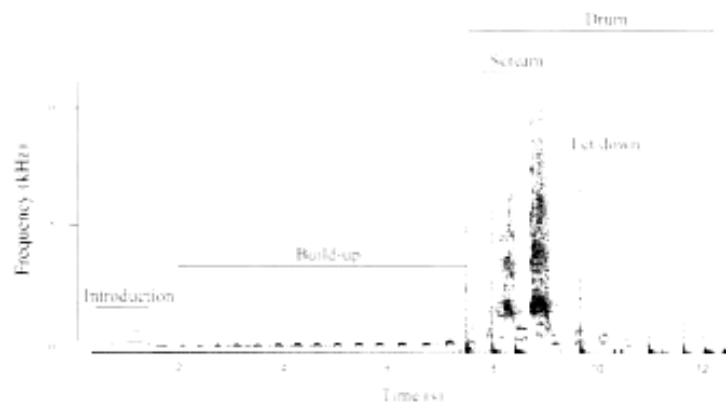


Fig. 1: Spectrogram of a pant hoot showing its four part (hust) phases, produced in association with a tree-drumming

**Fig. 1.** Immagine modificata da Crockford (2004)

Inoltre, in un interessante articolo inerente la capacità degli scimpanzé di combinare i propri richiami in stringhe composte, Crockford e Boesch<sup>116</sup> osservano che fra i vocalizzi che hanno avuto modo di registrare in uno studio direttamente sul campo, la metà occorre in combinazione con altri richiami, o con tamburellamenti. Il dato sorprendente è che il contesto in cui ciascuna unità veniva usata cambiava a seconda che si trovasse connessa ad un altro vocalizzo, o meno. In tale fenomeno *sintattico*, gli autori identificano in una forma di proto-sintassi lessicale, in cui l'annessione di unità diverse entro un unico proferimento ha dei rimarchevoli effetti di senso. Prendiamo in considerazione, ad esempio, lo

<sup>116</sup> Crockford, C., Boesch, C., *Call combinations in wild chimpanzees*, "Behaviour" 142, 2005, pp. 397–421.



stesso *pant hoot*, il richiamo più comune emesso dagli scimpanzé, generalmente impiegato, come abbiamo visto, nei contesti di spostamenti di gruppo, incontri fra gruppi diversi, o in occasione della ricerca di cibo, e il *grunt*, - vocalizzo emesso in presenza di lievi pericoli, o prima di iniziare uno spostamento di gruppo, o in occasione della ricerca di cibo<sup>117</sup>. La giustapposizione di tali vocalizzi in un richiamo composto, restringe il campo semantico ad un contesto specifico comune ad entrambi i richiami, la presenza di cibo, generando maggiore accuratezza nel riferimento al dato contesto:

Only one single call was predominantly produced in a single context. In contrast, more than half of the different call combinations were context specific, each being mainly produced in a single context. One explanation for this is that the act of combining calls refined the likelihood of a particular combination being produced in a particular context<sup>118</sup>.

---

<sup>117</sup> *Ivi*, p. 415: “Pant hoots were produced more than expected in ‘responses during feeding and travelling’, whilst grunts were produced more than expected in food contexts. Pant hoot and grunts were produced more in ‘responses and meeting at food’ than expected and in more ‘responses at food’ than grunts. Thus pant hoot and grunt is produced in the response context associated with pant hoots, but only during feeding, the context associated with grunts.”

<sup>118</sup> *Ivi*, p. 417.

**Table 1.** Descriptions of chimpanzee vocalisation and drum structures and their contexts of occurrence.

Vocalisation name	Acoustic characteristics*	Contexts of production
Bark	<400 ms; 600-2000 Hz; sometimes non-linear phenomena	aggression, travelling, hearing other parties of chimpanzees, encountering chimpanzees from neighbouring communities
Short bark	<170 ms; 400-1500 Hz	hunting colobus monkeys
Tonal bark	tonality: tonal; position of max. F0: early, <0.25, as a proportion of bark length	seeing large snakes
Grunt	40-150 ms; 100-350 Hz; tonality: broad range from quite tonal to highly noisy	feeding, greeting, close contact calls at the start of travel, mild threats
'Aaa' grunt	90-150 ms; 500-700 Hz	feeding
Deep grunt	170-1200 ms; 70-400 Hz	feeding
Hoo	170-1200 ms; 200-700 Hz; highly tonal with clear harmonic structure	feeding, mild fear or surprise such as finding animal carcass or small snake, travelling near nesting time
Hoo grunt	series of hoos, grunts or intermediate forms joined with voiced inhaled elements; 80-200 ms; 200-350 Hz; hoo/grunt elements are generally more tonal than the grunt elements of the pant grunt	nesting, occasionally feeding
Laughter	rapid inhaled and exhaled elements, generally quiet and voiced, grading to non-voiced; 50-100 ms; 80-300 Hz; rate of 4-7 exhaled elements/second	play
Pant	unvoiced, rapid inhaled and exhaled elements; duration and F0 fall within the range for pant grunts	occasionally grooming or feeding excitement
Pant grunt	series of grunts joined together by voiced inhaled elements; 30-200 ms; 100-200 Hz; tonality: highly noisy	greeting up the hierarchy

\* First two measures listed for each call type are the duration of a single element, followed by the fundamental frequency (unless otherwise stated). Measures are intended as a guideline and are approximate rather than definitive.

Pant hoot	comprised of up to four phases: introduction: hoos; build-up: series of hoos interspersed with voiced inhaled elements; climax: screams and sometimes barks; let down: similar to build-up but hoos having decreasing pitch	feeding, travelling, meeting chimpanzees from other parties or other communities
Scream	500-1000 ms; 800-2000 Hz; sometimes non-linear phenomena	receiving aggression, fear, as part of pant hoot and sometimes during copulation
Squeak	shorter and more tonal than screams; 50-250 ms; 800-2000 Hz	when threatened by a dominant
Whimper	highly tonal series of hoo-like calls; 100-700 ms, 250-1300 Hz	begging; by infants and juveniles during weaning or separation from the mother; by young adult males when separated from other males
Drum	pounding of hands and/or feet against large, resonant tree-buttresses; 1-30 beats at 2-13 beats/minute; lowest recordable frequencies <20 Hz (Moebius, unpubl. data)	travelling, display, encounters with chimpanzees from neighbouring communities, occasionally arriving at large food sources

**Table 8.** Significant contextual changes in the production of call combinations compared with the component calls<sup>a</sup>.

Component call	Contexts where component calls occurred more than expected	Call combination	Contexts in which call combinations occurred more than component calls	Classification of call combinations (see table legend)
'Aaa' grunt (GRa) Deep grunt (GRd)	Food Food	GRa + GRd	Food	4
Pant hoot (PH) Drum (DM)	Response in food and travel Travel, aggression	PH + DM	Travel (Aggression)	1
Pant hoot Grunt	Response in food and travel Food	PH + GR	(Response at food)	2
Pant hoot Pant grunt (PG)	Response in food and travel Greet, Meet in travel	PH + PG	Greet (Meet at food & travel)	1, 3
Pant hoot Scream (SM)	Response in food and travel Receive aggression, sex	PH + SM	Response at food (Response in travel)	1
Pant hoot Scream Drum	Response in food and travel Receive aggression, sex Travel, aggression	PH + SM + DM	Travel	1

<sup>a</sup>paired *t* tests; contexts without brackets:  $p < 0.05$  and ( )  $p < 0.1$  after Hochberg-Bonferroni correction; = call combination is not produced significantly more in any context than the component call. Relative to the component calls, call combinations were produced (1) in the same contexts as only one of the component calls (2) in the sum of the contexts of the component calls (3) in a similar, related context to each of the component calls (4) more specifically in the same context than either of the component calls (numbering is in keeping with that in the discussion).

**Fig. 2.** Immagini modificate da Crockford *et al.* (2004). In Table 1 viene fornita una lista dei diversi tipi di vocalizzi, delle rispettive caratteristiche acustiche e dei contesti di produzione. In Table 8, viene fornita una lista di combinazioni “sintattiche” di vocalizzi, e dei loro impieghi d’uso.

Tale “effetto di significato”, osservano gli autori, è soltanto uno dei diversi effetti cui l’operazione di composizione dei richiami può dar luogo. Una ulteriore funzione della composizione di richiami diversi, classicamente riportata nella letteratura sui sistemi di comunicazione delle scimmie è quella di *modificare* il significato dei singoli vocalizzi che compongono il richiamo. Rientrano in questa funzione quelle forme composte che rinviano a contesti “intermediari” rispetto a quelli in cui vengono utilizzate le singole componenti dello stesso richiamo. Ciò vuol

dire che si ricorre ai richiami composti in contesti simili a quelli delle singole unità giustapposte, ma non identici. Un ulteriore caso di modificatori di senso, è dato da quelle giustapposizioni che intensificano l'urgenza della situazione. Queste due ultime funzioni sono largamente riconosciute nella letteratura sul tema. Tuttavia, nell'articolo in esame, gli autori mettono in luce ulteriori aspetti funzionali che caratterizzano l'uso dei richiami composti. Ad esempio:

(...) the pant hoot and pant grunt combination, a longer range and a shorter range call respectively, may convey information simultaneously to different audiences. The combination is produced exclusively by females in Tai, specifically while greeting males during meeting contexts (...). The pant hoot usually starts as the female joins the new party whilst the pant grunt part only begins during the final approach to the male. As such, the pant hoot part may announce her arrival to the party as a whole and the pant grunt act as a greeting, exclusively directed at the male<sup>119</sup>.

Anche in questo caso, l'unione delle due unità sfocia nella in un effetto di senso che in qualche modo può rappresentare l'intersezione semantica delle due unità, rendendo così più specifico il contesto d'uso, come nell'esempio di cui sopra. Tuttavia, oltre ad essere un modo per salutare, la giustapposizione di *pant hoots* e *pant grunts* determina il senso specifico dei due vocalizzi, che all'interno della combinazione sintattica, si restringono reciprocamente il contesto d'uso. Soltanto quando proferiti l'uno di seguito all'altro, infatti, tali unità desumono un significato

---

<sup>119</sup> *Ibidem*, corsivo mio.

specifico, ovvero il *pant hoot* funge da saluto generale al gruppo, e il *pant grunt* da saluto diretto specificatamente al maschio<sup>120</sup>. Infine, una ulteriore funzione che vale la pena di riportare fra quelle enucleate dagli autori riguarda ancora una volta i *pant hoots*, principalmente connessi ad altri richiami usati per lunghe distanze, e ai tamburellamenti. Gli studiosi, nel tentativo di trovare una spiegazione a tale tipo di composizione, ipotizzano quanto segue:

pant hoots act as signature calls for signallers, which can be combined with context information in the calls or drums which follow. If this is the case, this may represent a primitive form of 'agent' combined with 'context'<sup>121</sup>.

Stando a questa ipotesi, dunque, i *pant hoots* indicherebbero l'identità del soggetto che compie l'azione, mentre i tamburellamenti ne definirebbero il contesto di azione. Se così fosse, si potrebbe dunque interpretare tale combinazione come la radice evolutiva delle moderne forme predicative. Troviamo che sebbene molto curiosa, tale ipotesi vada trattata con una buona dose di cautela metodologica. In particolare, è necessario tenere presente che a differenza delle interazioni comunicative umane, tali forme di comunicazione prescindono, per quanto ne sappiamo, dalla condizione di mutuo riconoscimento delle intenzioni, fra gli

---

<sup>120</sup> Troviamo che tale esempio possa essere considerato un interessante caso di intreccio tra uso, proto-sintassi e quello che forse rappresenta il germe evolutivo del principio di contesto.

<sup>121</sup> *Ivi*, p. 418, corsivo mio.

interagenti. La trasmissione delle informazioni può essere considerata un *effetto collaterale* del fatto che i segnalanti stiano reagendo emozionalmente e vocalmente ad eventi evolutivamente significativi, in quanto potenziali minacce per la sopravvivenza. Inoltre, un ulteriore punto che è opportuno sottolineare, è che - come gli stessi ricercatori puntualizzano - i contesti cui le combinazioni di unità si riferiscono non sono mai completamente diversi rispetto alle aree semantiche proprie delle unità singole. L'effetto semantico di tale meccanismo proto-sintattico consiste infatti in una più specifica determinazione del contesto in cui le unità costitutive vengono impiegate o sul segnalante stesso<sup>122</sup>, ma mai in un rinvio ad un contesto totalmente diverso. Operazione semantica, questa, che al contrario, interessa i canti dei gibboni.

Come vedremo più da vicino nella seconda sezione, fra i gibboni le coppie di partner realizzano dei duetti, ovvero dei canti altamente sincronizzati al fine di segnalare ai vicini il loro legame, e allo stesso tempo per la difesa del proprio territorio contro attacchi esterni. In un recente studio, il gruppo di ricerca di Klaus Zuberbühler<sup>123</sup> si è chiesto se tali duetti differissero strutturalmente dai canti realizzati in risposta alla

---

<sup>122</sup> Ivi, p. 419: "(...) it seems likely in chimpanzees that call combinations serve several functions, all of which enable listeners to extract more precise information about the signaller or the current situation of the signaller."

<sup>123</sup> Clarke, E., Reichard, U. H., & Zuberbühler, K., *The syntax and meaning of wild gibbon songs* "PLoS ONE", 1, 2006, pp. 1-10.

presenza di potenziali predatori. Nonostante, ad “orecchio nudo” tali canti vengono percepiti dall’uomo come sostanzialmente uguali, dettagliate analisi effettuate al computer hanno rivelato che, in effetti, i due tipi di contesti funzionali elicitano canti dalla complessità strutturale sottilmente diversa. Ad esempio, i canti emessi in risposta alla presenza di un predatore sono caratterizzati da un numero maggiore di note chiamate “hoo”, presentano una durata maggiore. Inoltre, probabilmente in risposta all’urgenza della situazione, i maschi rispondono alle parti cantate dalle femmine con maggiore prontezza, mentre le femmine, nella parte iniziale, aspettano un po’ più a lungo - rispetto a quanto avviene nei duetti - prima di emettere il loro “*great call*”. Inoltre, una nota, detta “*sharp wow*” sembra avere un valore particolare; come gli stessi autori affermano,

Whatever the perceptually relevant cues, our observations also demonstrated that neighbouring groups were able to differentiate between songs given in the two contexts. In particular, we observed a predator-specific delay in the production of the first great call, as well as the inclusion of ‘sharp wow’ notes in all cases in which neighbours responded to the predator-induced song. We never observed these patterns in the response songs of neighbours to normal duets. In all observed cases, absent males began to sing while returning to the rest of the group. Again, we never observed such behaviour during normal duets. The returning males’ songs always included ‘sharp wow’ notes, a convincing sign that they understood the meaning of the song produced by their group<sup>124</sup>.

A questo punto, vale la pena chiedersi se oltre a costituire una mera risposta fisiologica innata, tali canti, opportunamente adattati al contesto, e alla identità degli interagenti, non rappresentino anche una forma molto

---

<sup>124</sup> Ivi, p. 8.

rudimentale di comunicazione intenzionale, in cui i soggetti, pur non consapevoli dello stato mentale degli altri, ne vogliono preservare la vita, mettendoli all'erta del potenziale pericolo. In effetti, come notano gli autori dello studio, gli individui dello Khao Yai che hanno osservato, cambiano spesso gruppo, e di conseguenza i membri dei vari gruppi che occupano territori vicini sono spesso imparentati. Così, una delle spinte evolutive in grado di selezionare l'emissione dei canti a lunga distanza, può essere proprio la *inclusive fitness*, ovvero la possibilità di favorire la riproduzione dei propri geni aiutando quegli individui che portano geni simili: i parenti, e la prole dei parenti.

Il tema della sintassi nei sistemi di comunicazione delle scimmie è uno dei principali fuochi di studio di alcuni gruppi di ricerca guidati da Klaus Zuberbühler. In particolare, questi ultimi hanno concentrato la propria ricerca sulle alcune specie di scimmie non antropomorfe, che vantano, come vedremo, l'uso di forme sintattiche che, sebbene assai rudimentali da un punto di vista semantico, sono più complesse rispetto a quelle riconosciute fra i richiami degli scimpanzé. Ciò mostra che la capacità di combinare diversi elementi con riferimenti contestuali diversi può essere a buon diritto riconosciuta come una competenza evolutasi sotto la pressione di diverse spinte ambientali e sociali che hanno interessato tanto gli uomini, quanto specie di scimmie filogeneticamente ad essi



distanti. Esempari, a tal proposito, gli studi effettuati sui cercopitechi di Campbell<sup>125</sup>, i quali vantano un sistema comunicativo in cui a due diversi tipi di predatori – leopardi ed aquile coronate rispettivamente corrisponde un particolare richiamo<sup>126</sup>. Tuttavia, tale sistema di segnalazione vanta di un meccanismo sintattico molto interessante, che rappresenta, secondo l'interpretazione di Ouattare *et al.*<sup>127</sup>, una vera e propria forma di affissazione. Infatti, osservano gli autori dell'articolo, tale specie essere in grado di anettere un suffisso ai richiami utilizzati in risposta al potenziale attacco di un predatore, operazione che sfocia nell'alterazione del significato generale del richiamo. Nello specifico, la giustapposizione di una unità acustica, detta “oo” – mai usata singolarmente – genera una nuova espressione indicante la presenza di un non specificato elemento di disturbo. Così ne parlano i ricercatori:

The key finding of this study was that males adhered to a simple affixation rule, which increased their small basic vocal repertoire. “Krak” and “krak-oo” as well as “hok” and “hok-oo” calls were composed of the same call stem elements, while the rapid addition of the “oo” affix generated a significant

---

<sup>125</sup> Cfr. Clay, Z., Zuberbühler, K., *Bonobos Extract Meaning from Call Sequences*, “PLoS ONE”, 6, 2011, p. 6: “In contrast to chimpanzees, who produce an acoustically graded call type that co-varies with food quality, bonobos regularly mix several acoustically distinct food call types into heterogeneous strings of vocalisations. Rather than at the level of individual calls, food quality appears related to the probabilistic composition of heterogeneous call sequences [31]. Results from our playback experiment indicate that rather than attending to individual call types, receivers took into account the relative proportions of different calls within the sequence and extracted meaning by integrating information from across the call units.”

<sup>126</sup> Cfr. Seyfarth, R., Cheney, D., Marler, P., *Vervet monkey alarm calls: Semantic communication in a free-ranging primate*, “Animal Behaviour”, 28, 1980, pp. 1070–1094.

<sup>127</sup> Ouattara, K., Lemasson, A., Zuberbühler, K., *Campbell's Monkeys Use Affixation to Alter Call Meaning*, “PLoS ONE”, 4, 2009, pp. 1-7.

change in the semantic content in terms of the types of external events the calls referred to. While “krak” and “hok” were predator-specific calls, the suffixed versions were produced in less specific contexts.

[...] However, despite these results, it is still largely unclear whether non-human primates intentionally inform their audience about the event they have just experienced, or whether their vocal response is more directly driven by the psychological processes triggered by external events, the currently prevailing hypothesis. What our results show is that callers appear to make some judgements about the nature of the event (tree fall, group gathering to travel, conspecific intruder, eagle, leopard), and that this assessment determines whether or not affixation takes place<sup>128</sup>.

È opportuno soffermarsi su due aspetti caratteristici di tale sistema di comunicazione: come notano gli stessi autori, a prescindere dalla capacità di *volere* informare gli altri, i cercopitechi di Campbell sono in grado di trarre delle informazioni derivanti dalla annessione sintattica di un suffisso al richiamo di allarme. In particolare, l’effetto di senso generato dai richiami così ottenuti consiste nel rinvio ad un’area semantica più generale, ma pur sempre, nella maggior parte dei casi, connessa a quella propria della prima parte del proferimento. Ad esempio, mentre il richiamo detto “krack” viene emesso in risposta alla percezione auditiva o visuale di un leopardo, il “krack-oo” rinvia alla presenza di un pericolo generale, sia esso un’aquila, un leopardo, o in risposta ad un richiamo di pericolo emesso dai gruppi vicini<sup>129</sup>.

---

<sup>128</sup> *Ivi*, p. 5.

<sup>129</sup> Vedi *Ivi*, p. 3: “Some calls were given to a broad, others to a narrow range of events. Crucially, “krak” calls were exclusively given after detecting a leopard, suggesting that it functioned as a leopard alarm call, whereas the “krak-oo” was given to almost any disturbance, suggesting it functioned as a general alert call. Similarly, “hok” calls were almost exclusively associated with the presence of a

Oltre alla capacità di modificare - generalizzandoli - i tratti semantici dei richiami attraverso l'aggiunta di un suffisso, i cercopitechi di Campbell (a differenza di quanto rilevato sui richiami degli scimpanzé) vantano la importante capacità di combinare due richiami, generando un riferimento contestuale nuovo e non direttamente connesso al valore semantico dei singoli elementi accorpati, rinviando, cioè, a contesti diversi da quelli specificati da ciascuna unità di per sé. È ciò che emerge da una ulteriore ricerca effettuata dagli stessi Ouattara *et al.*, che ne parlano nei seguenti termini:

---

crowned eagle (either a real eagle attack or in response to another monkey's eagle alarm calls), while "hok-oo" calls were given to a range of disturbances within the canopy, including the presence of an eagle or a neighbouring group (whose presence could sometimes be inferred by the vocal behaviour of the females)." Su tali studi convergono evidenze sperimentali provenienti da differenti ricerche relative alla medesima ipotesi. Ad esempio, da un'ulteriore ricerca condotta dallo stesso Zuberbühler (*A syntactic rule in forest monkey communication*, "Animal Behaviour", 63, 2002, pp. 293–299) emerge che i cercopitechi Diana comprendono le variazioni semantiche determinate da regole combinatorie presenti nella comunicazione naturale di un'altra specie, il cercopiteco di Campbell, che vive in territori contigui. Curiosamente, questo emette degli allarmi in corrispondenza all'avvistamento di leopardi o aquile, che i Diana sono in grado di tradurre nel proprio sistema di allarmi, sortendo la appropriata reazione comportamentale. In particolare, i cercopitechi di Campbell, in situazioni poco rischiose antepongono ai tipici richiami di pericolo un paio di vocalizzazioni, i "boom", che precedono i susseguenti allarmi di circa 25 secondi. La serie di richiami introdotti dal boom sono emessi, infatti, in corrispondenza di fenomeni generali di disturbo come un ramo che si sta spezzando, o l'avvistamento di un predatore in lontananza. In altri termini, hanno effetti semantici sui successivi richiami di allarme paragonabili a quelli delle dichiarazioni linguistiche vaghe, o ad unità semantiche come "forse" o "una sorta di", hanno sulle parti della frase che seguono. Entrambi, infatti, indicano una situazione generica che non richiede una risposta diretta. Il dato interessante è che i cercopitechi Diana comprendono tale combinazione, rispondendo con un comportamento adeguato, ovvero non emettendo il richiamo d'allarme corrispondente al richiamo di pericolo imminente connesso alla presenza dei predatori. Da tale dato empirico, l'autore deduce che i "boom" condizionino la specificità semantica del susseguente richiamo d'allarme dei cercopitechi Diana.

The different call sequences were not randomly assembled but ordered in specific ways, with entire sequences serving as units to build more complicated sequences. As mentioned, pairs of booms alone instigate group movements toward the calling male, while K+ [krack-oo] series functioned as general alarm calls. If combined, the resulting sequence carried an entirely different meaning, by referring to falling wood. In all cases, the booms preceded the K+series. We also found that another sequence, the H+ [hok-oo]series, could be added to boom–K+sequences, something that callers did when detecting a neighboring group. H+series were never given by themselves<sup>130</sup>.

In particolare, gli autori distinguono alcune specifiche regole *sintattiche* - oltre alla suddetta regola di affissazione - in uso nel sistema di comunicazione di tale specie di cercopitechi. Vale la pena di riportarne le più salienti. In primo luogo, i segnalanti sono in grado di combinare due diverse sequenze di per sé significative in una sequenza ancora più complessa, dal significato nuovo. Ad esempio, dati una serie di “boom” - generalmente adottati dai maschi del gruppo che non hanno alcun contatto visuale con il resto del gruppo – e una sequenza di “krack-oo”, che come abbiamo visto indica la presenza di un pericolo in generale, il significato generale generato dall'accorpamento delle due sequenze in un'unica espressione consiste nel riferimento ad un ramo (o un albero) che sta per cadere. In un contesto ancora diverso, ovvero in risposta ad un richiamo dei gruppi vicini, i cercopitechi di Campbell, usano aggiungere a tale sequenza un'ulteriore allarme composto, lo “hok'oo”. Inoltre, come nel sistema di

---

<sup>130</sup> Ouattara, K, Lemasson, A, Zuberbuhler, K., *Campbell's monkeys concatenate vocalizations into context-specific call sequences*, “Proceedings of the national academy of science”, 106, 2009, p. 4.

allarmi degli scimpanzé, anche questi cercopitechi usano definire un riferimento contestuale combinando sequenze le cui aree semantiche sono diverse, ma intersecanti. In questi casi, il riferimento consiste nel significato comune alle due sequenze che si intersecano. Gli autori dello studio sottolineano l'importanza dell'ordine delle sequenze utilizzate in grado di veicolare specifici riferimenti in particolari contesti. In effetti, il principio sintattico secondo cui l'ordine delle unità discrete *determina* il significato dell'espressione - e, allo stesso tempo, il valore delle stesse unità entro tale combinazione - vige anche nel sistema di comunicazione dei cercopitechi nasobianco maggiore.

Recenti ricerche effettuate da Kate Arnold e Klaus Zuberbühler<sup>131</sup> rivelano infatti dei dati molto interessanti relativamente alla presenza di strutture rudimentali che interessano l'ordine di unità discrete, in grado di *significare* la presenza di determinati tipi di predatori:

Series consisting of 'pyows' are a common response to leopards, while 'hacks' or 'hacks' followed by 'pyows' are regularly given to crowned eagles. Sometimes, males produce a further sequence, consisting of 1-4 'pyows' followed by 1-4 'hacks'. These 'pyows-hack' (P-H) sequences can occur alone, or they are inserted at or near the beginning of another call series. Regardless of the context, P-H sequences reliably predict forthcoming group progression<sup>132</sup>.

---

<sup>131</sup> Arnold, K., Zuberbühler, K., *Meaningful call combinations in a non-human primate*, "Current Biology", 18, 2008, pp. R202-R203. Cfr. Arnold, K., Zuberbühler, K., *Semantic combinations in primate calls*, "Nature", 441, 2006, p. 303.

<sup>132</sup> *Ivi*, pp. 202-203.

L'inversione della sequenza di suoni "pyows-hacks" genera differenti effetti di significato. Infatti, la sequenza costituita da una serie di vocalizzi "hack" seguita da diversi "pyows" rinvia alla presenza dell'aquila coronata. Al contrario, una sequenza di "pyows" (da uno a quattro) seguita da "hacks" (da uno a quattro) rappresenta l'invito, da parte del maschio del gruppo, di spostarsi verso una zona più sicura in seguito all'avvistamento di un oggetto specifico, o per trovare del cibo. Tale combinazione sintattica, come confermato dai lavori sperimentali effettuati sul campo, dunque non è casuale, ma, al contrario, si fa latrice di specifici valori significazionali. Gli studi fin qui esposti sulle connessioni sintattiche che interessano le espressioni funzionalmente referenziali delle scimmie, sono assai significativi, in quanto in grado di rivelare le radici evolutive delle connessioni sintattico-proposizionali del linguaggio umano. Come abbiamo visto, infatti, i sistemi di comunicazione delle scimmie sono regolati da diversi principi combinatori statisticamente computabili, che modificano i tratti semantici delle unità referenziali. Tuttavia, nella tematizzazione di tale fenomeno, è bene tenere in considerazione alcuni punti fondamentali.

In primo luogo, bisogna osservare che il principio che fonda il riferimento contestuale esterno è il particolare ordine in cui le diverse unità (o sequenze di unità) occorrono. Infatti, il riferimento stesso non è funzionale al valore discreto degli elementi costitutivi – la cui area

semantica rimane spesso volubile ed olistica – bensì alla particolare sequenza fonica in cui occorrono. Potremmo a buon ragione affermare, dunque, che il *prius* significazionale è reso dalla connessione sintattica delle unità foniche, a prescindere dal loro valore semantico. Per quanto ne sappiamo ad oggi, nessun sistema linguistico al di fuori di quello umano presenta una natura combinatoria e contestuale allo stesso tempo, tale cioè che il valore semantico di ciascuna unità discreta confluisca nella determinazione del senso globale del contesto fraseologico in cui occorre, desumendone, allo stesso tempo una precipua denotazione. In secondo luogo, vale la pena di osservare che tali segnali vengono emessi in risposta ad una gamma di esperienze reali ristretta agli eventi significativi da un punto di vista evolutivo, che potrebbero, cioè, compromettere la sopravvivenza. Per tale motivo, non possiamo attribuire a questo tipo di segnalazione uno statuto intenzionale *tout court*. Non possiamo cioè appurare con certezza definitiva il fatto che i segnalanti emettano i vocalizzi *al fine* di informare gli altri della presenza di un pericolo, o di cibo - piuttosto che come mera, spontanea espressione di uno stato emozionale interno. Infine, bisogna tenere presente che le configurazioni sintattiche in cui le unità occorrono, sebbene molto semplici, sono in grado di ovviare ai vincoli articolatori imposti dalla natura anatomico-neurale delle scimmie, permettendo di generare significati nuovi a partire dalla

semplici regole di sequenziamento delle unità già presenti nel loro stesso repertorio.

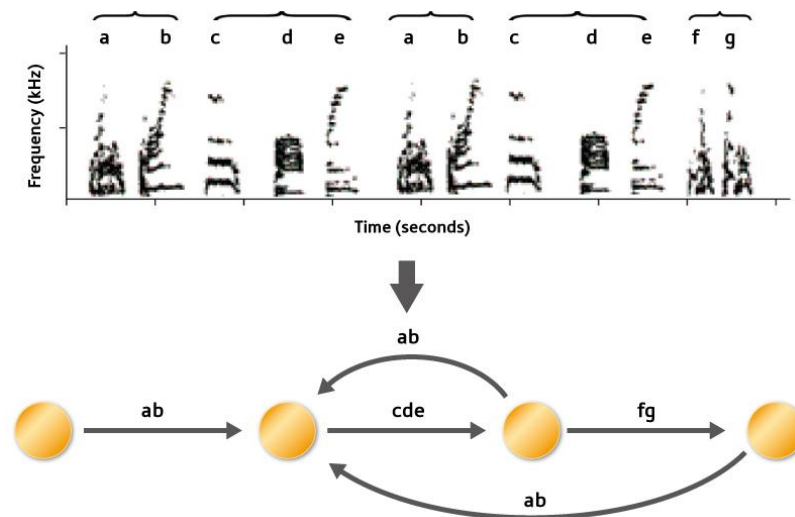
### **6.3 Strutture fonologiche tra i canti degli uccelli**

Gli esseri umani e molte specie di uccelli canori condividono molte similarità cognitive che si estendono dal dominio della acquisizione ontogenetica dei rispettivi linguaggi, a quello dei dialetti “culturalmente” trasmessi, o alla centralità del ruolo delle interazioni sociali nello sviluppo del linguaggio. Rimandando l’approfondimento di tali aspetti alla seconda sezione del presente lavoro, è importante enfatizzare qui un ulteriore tratto cognitivo condiviso, inerente il dominio fonologico e i meccanismi sintattici che ne regolano l’articolazione. Gli studi inerenti tale significativo parallelismo cognitivo fanno capo al paradigma chomskiano, alla luce del quale, la definizione di “canto” acquista una fisionomia completamente nuova (figura 3). Così ne parla Okanoya, uno dei maggiori studiosi del tema, in un significativo articolo sulla sintassi dei canti del passero del Giappone (*Lonchura striata domestica*):

Traditionally, bird song has been analyzed using terms, such as notes, syllables, and phrases [...]. A song note is the smallest unit of utterance recognizable as a continuous trace on the sonogram. A syllable is a minimum unit of repetition within the song, usually a combination of a few notes, or sometimes equal to a particular note itself when such a note is itself a unit of repetition. A song phrase



is built of a number of syllables. A motif refers to a stereotyped sequence within a song [...] and a strophe refers to a bout of motifs [...]. [...] We found it difficult to analyze the song of Bengalese finches using these traditional terms. A stereotyped song motif could not be firmly identified in the song episodes of a Bengalese finch. This is because songs of Bengalese finches have complex note-to-note transition patterns. One note could be followed by several possible notes in a song. Multiple notes are organized into chunks, but these chunks do not have a stereotyped structure; a note could be used in several different chunks and some notes could be repeated several times. These chunks are in turn sung with probabilistic rules; one chunk could be followed by several different chunks. The most suitable notation to describe this complex organization turned out to be from linguistics. A finite state automaton, or syntax, describes several states connected by arrows [Fig. 1]. One state can transit to several possible states, each associated with some probability.<sup>133</sup>

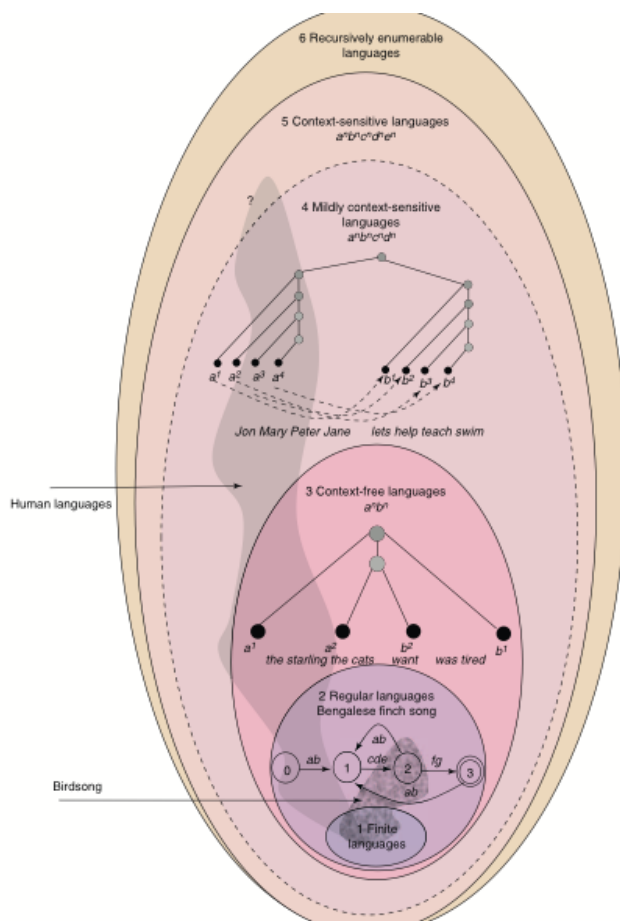


**Fig. 3.** Immagine modificata da Okanoya *et al.* (2004a). Le note dei canti dei passeri del Giappone sono raggruppabili in insiemi di due o tre elementi: i cosiddetti “*chunks*”. La distribuzione dei *chunks* è statisticamente regolata, ed analizzabile calcolando le probabilità di transizione da un *chunk* all’altro. In maniera interessante, il modo più efficace di descrivere la struttura dei canti degli uccelli consiste nella notazione adottata in linguistica computazionale, ovvero come la rappresentazione di automi a stati finiti, in cui ogni stato (cerchio arancione in figura) rappresenta il gruppo di note trascritto sulla freccia che lo segue.

Descrivere i canti dei dei passeri del Giappone come un mero

<sup>133</sup> Okanoya, K., *Song Syntax in Bengalese Finches: Proximate and Ultimate Analyses*, “Advances in the Study of Behavior”, 14, 2004, pp. 304-305.

agglomerato gerarchicamente organizzato di suoni, dunque, è, agli occhi dello studioso, un approccio riduzionistico, che non ne rivelerebbe in pieno la complessità strutturale. Complessità che di fatto, è rappresentabile attraverso un'architettura sintattica condivisa con il linguaggio verbale (figura 4) e, nello specifico, assimilabile alle grammatiche a stati finiti della gerarchia di Chomsky. Vediamo in che senso.

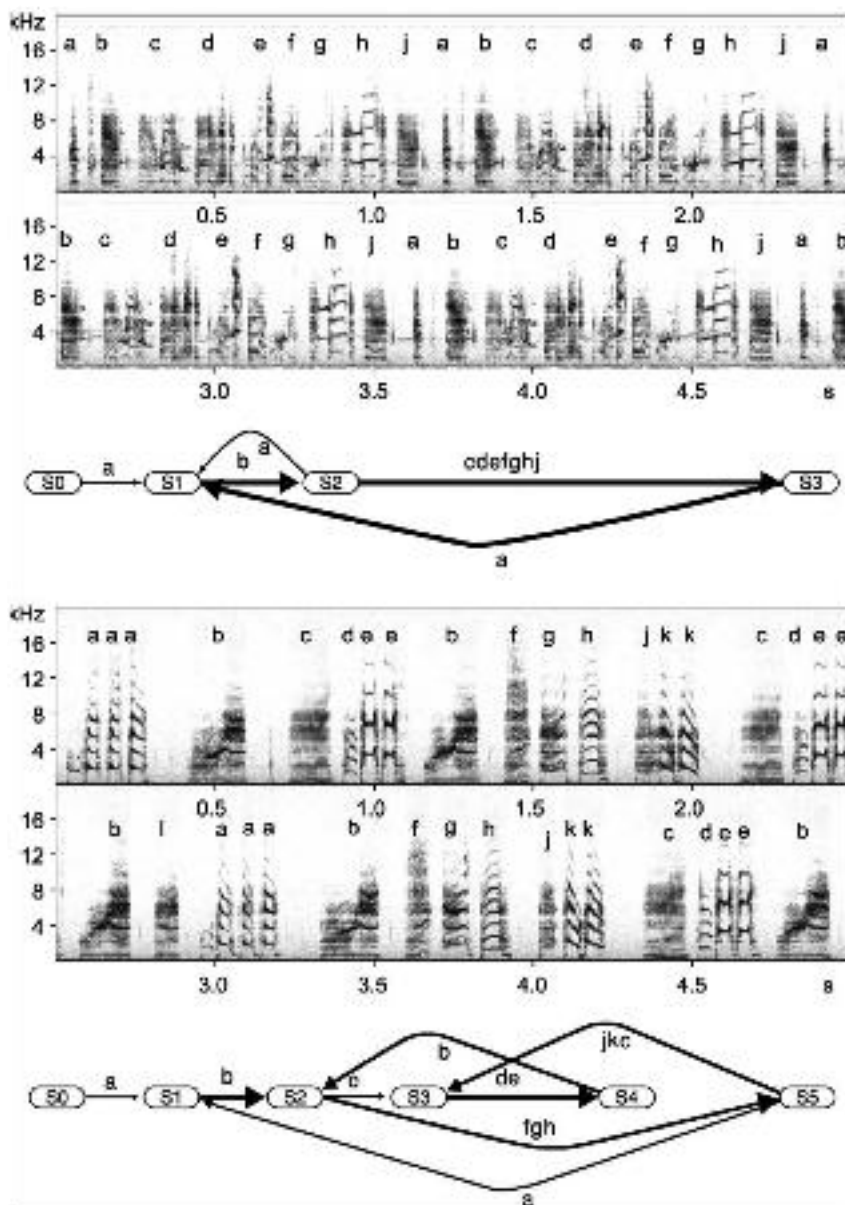


**Fig. 4.** Immagine modificata da Berwick *et al.* (2011)<sup>134</sup>. Rappresentazione della gerarchia di Chomsky, e della ipotetica localizzazione tanto del linguaggio verbale umano, quanto dei canti degli uccelli al suo interno.

<sup>134</sup> Berwick, R. C., Okanoya, K., Beckers, G. J., L., Bolhuis, J., *Songs to syntax: the linguistics of birdsong*, "Trends in Cognitive Science", 15, 2011, pp. 113-121.



rispetto a tali esempi. Ciò salta immediatamente agli occhi nella seguente immagine, in cui un canto di tale specie viene messo a confronto con un canto del passero striato (*Lochura domestica*) da cui si è filogeneticamente evoluto:



**Fig. 4.** Immagine modificata da Okanoya (2004a).

Come si può notare, la sintassi dei passeri striati presenta una dinamica essenzialmente lineare, in cui le transizioni da uno stato all'altro sono governate da un numero esiguo di regole. Al contrario, la sintassi dei canti dei passeri del Giappone presenta più transizioni possibili ad ogni stato. Tale caratteristica ne eleva la sintassi al rango dei processi stocastici, che rientrano, nell'ambito delle grammatiche a stati finiti, in un insieme dalla complessità strutturale più sofisticata, in quanto non deterministico. Infatti, la transizione tra stati diversi è regolata da probabilità di distribuzione statisticamente quantificabili e descrivibili nei termini di una catena di Markov<sup>136</sup>. Così ne parla lo stesso Okanoya:

The most widely used oscine song system models (the zebra finch, white-crowned sparrow, song sparrow, and swamp sparrow) could all be identified as having linear song syntax. When there are some variations introduced in the ordering of song notes, such a song should be called as a non-deterministic song. Species with non-deterministic song repertoires include the nightingale, starling, willow warbler, and Bengalese finch. Among these species, Bengalese finches are unique in that their songs are characterized by finite-state syntax. Finite-state syntax refers to a simple form of syntax in which finite numbers of state are interconnected by arrows and a string of letters is produced when state transition occurs. [...] Finite-state syntax can be expressed as a Markov model of note-to-note transitions, in which transition probabilities of certain combinations of notes are high, while that between some note and others are low, reflecting chunking, recursive loop, and complex state transitions of the song production.<sup>137</sup>

<sup>136</sup> È opportuno precisare che vi sono specie i cui canti presentano una complessità ancora più elevata, computabile per mezzo di processi di Markov di ordini superiori al primo. Tuttavia, ci limitiamo ad analizzare, con Okanoya, i canti del passero del Giappone per ragioni di chiarezza espositiva. Vedi Gentner, T., *Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris**, "Animal Behaviour", 56, 1998, pp. 579-594.

<sup>137</sup> Okanoya, K., *The Bengalese Finch: A Window on the Behavioral Neurobiology of Birdsong Syntax*, "Annals of the New York Academy of Sciences", 1016, 2004, p. 725. Cfr. Berwick, R. C., *et al.*, *Songs to syntax: the linguistics of birdsong*, cit., p. 117: "The Bengalese finch song automaton [...], which encompasses the full song sequence repertoire extracted from a single bird [...], indicates that birdsong structure can be more complicated than a simple bigram description. [...] In fact, the transitions between states are stochastic; that is, the finch can vary its song by choosing to go from state 2 back to state 1 with some likelihood that is measurably different from the likelihood of continuing on to state 3. In any case, formally this means that the notes *cde fg* can appear in the 'middle' of a song, arbitrarily far from either end, bracketed on both sides by some arbitrarily long number of *cde ab* repetitions. Such a note pattern is no longer strictly locally testable because now there can be no fixed-length 'window' that can check whether a note sequence 'passes'. Instead of checking the note sequences directly, one must

In ultima analisi, dunque, possiamo affermare che esiste un nucleo di operazioni sintattiche - definibile nelle regole di una grammatica a stati finiti - che la specie umana condivide con talune specie di uccelli canori. Si tratta di un parallelismo assai significativo da un punto di vista evolutivo, in quanto dimostra che è possibile ravvisare in tale nucleo cognitivo un tratto analogo (evoluto, cioè, in entrambe le specie a fronte delle medesime pressioni selettive) costituito dall'abilità di manipolare sintatticamente delle unità. Competenza che si è verosimilmente evoluta in forme sempre più complesse, fino a sfociare nell'attuale configurazione del linguaggio verbale. Tali dati, dunque, sembrano confermare l'ipotesi avanzata da Darwin, secondo il quale il preludio evolutivo della sintassi proposizionale è da identificare in articolazioni analoghe ai canti realizzati da altre specie animali in contesti che hanno a che fare con la conquista/difesa di un partner sessuale, o con la difesa del territorio. Date tali premesse, ed illustrati i tratti di sintassi condivisi fra le due specie, è opportuno osservarne le differenze costitutive. Nello specifico, ci porremo una domanda fondamentale per il nostro lavoro di ricerca, inerente la natura precipua delle unità interessate dalle operazioni sintattiche nei

---

use the memory of the FSA indirectly to 'wait' after encountering the first of a possibly arbitrarily long sequence of *cde abs*. The automaton must then stay in this state until the required *cde fg* sequence appears. Such a language pattern remains recognizable by a restricted FSA, but one more powerful than a simple bigram checking machine."

linguaggi degli esseri umani e degli uccelli canori. Tali questioni saranno l'oggetto delle argomentazioni nel prossimo paragrafo.

#### **6.4 Dai canti alle parole-note**

L'idea che l'evoluzione della complessità strutturale di articolazioni sonore trovi le sue radici nella selezione sessuale è, come abbiamo visto, ben giustificata empiricamente dagli studi condotti sugli uccelli canori. Fin qui abbiamo descritto il tipo di complessità strutturale che interessa i canti degli uccelli, e le dinamiche evolutive che ne avrebbero guidato la selezione; rimane tuttavia ancora da spiegare in che modo la teoria della selezione sessuale dei canti degli uccelli si incontri con l'evoluzione del linguaggio verbale, la cui natura va evidentemente oltre la mera struttura combinatoria degli elementi costitutivi. Sebbene la tesi secondo la quale il linguaggio proposizionale trovi le sue radici nell'evoluzione delle complessità strutturali di note, o meglio di *chunks*) possa trovare una buona giustificazione empirica nello studio dei numerosi tratti analoghi evolutisi negli uccelli canori, emergono, in effetti, delle differenze sostanziali tra i sistemi di comunicazione degli umani e degli uccelli che ne segnano importanti divergenze. I domini in cui il linguaggio umano diverge dai sistemi di comunicazione animale sono essenzialmente legati alle sfere

pragmatico, semantico ed intenzionale. È importante sottolineare che tali ambiti si intrecciano a formare un unico fenomeno *bio-logico*, la cui configurazione non è discretamente riducibile alle componenti. Infatti, le unità del linguaggio verbale - siano esse parole, proposizioni, o gesti - si permeano di significati il cui intreccio oltre che formalizzabile in regole grammaticali, è funzionale ad un dominio specifico dell'uomo: la sfera simbolica. Ciò vuol dire che le parole sono connesse fra loro in virtù di una relazione semantica volta a rappresentare stati di cose; essa è tale da render possibile l'individuazione di dipendenze tra parole anche molto distanti, all'interno di una medesima frase, e legate da regolarità distribuzionali più complicate di quelle formalizzabili in catene di Markov del primo ordine.

Un'ulteriore caratteristica distintiva del linguaggio umano è la creatività regolare. Gli animali umani, infatti, vantano la capacità di produrre e comprendere, attraverso un uso infinito di mezzi finiti, una quantità di enunciati diversi l'uno dall'altro, sempre nuovi e ciononostante perfettamente comprensibili. Di fatti, una caratteristica costitutiva delle lingue umane è l'intrinseca illimitatezza, ovvero una estensibilità conforme a regole di combinazione-formazione precipue, in linea di principio sempre aperta. Tale proprietà è nota come "creatività governata da regole", e si situa, da una parte, sul piano sintattico: dall'incassamento ricorsivo di lessemi, sintagmi, o frasi entro frasi emergono infatti enunciati che, a parte



le limitazioni connaturate di memoria, energia, e tempo, non prevedono in linea di principio alcun limite superiore. Dall'altra, tale proprietà si situa sul piano morfologico, dove ad operare è il meccanismo della “doppia articolazione”<sup>138</sup>, indicata da André Martinet come tratto specifico delle lingue. Essa consiste nell'articolazione della lingua in due livelli. Il primo è dato dai costituenti delle parole stesse, ovvero dai monemi (prefissi, suffissi, desinenze, basi lessicali minime); il secondo, invece, dai fonemi, ovvero unità sonore minime, in sé prive di significato, e non ulteriormente scomponibili. Su tali piani, dunque, le parole si combinano (ricorsivamente) nei modi più disparati cosicché ogni unità lessicale o fonetica minima, se integrata ad altre unità del medesimo tipo, diviene *input* di un nuovo lessema, che a sua volta, può farsi *input* di un'ulteriore unità. Si consideri, ad esempio, il caso del termine “bisnonna”, da cui deriva “bis-bisnonna”, da cui, a sua volta deriva “bis-bis-nonna”. Oppure, “giustificabile”, da cui nasce “in-giustificabile”, e da questo, “in-giustificabil-ità”. Da questo, se vogliamo, potremmo derivare “anti-ingiustificabilità”, e così via<sup>139</sup>.

Tale flessibilità semiotica della lingua, se da una parte consente di agganciare sempre nuove sfumature pertinenti della nostra *Umwelt*, dall'altra ne rivela una caratteristica indeterminatezza, che richiede un

---

<sup>138</sup> Vedi De Mauro, T., *Prima lezione sul linguaggio*, Laterza, Roma-Bari, 2002, p. 55.

<sup>139</sup> Vedi Pinker S., *The Language Instinct*, cit., pp. 116-49.

processo continuo di chiarimenti e individuazione del senso esatto. Tale ambiguità semantica, evidente, ad esempio nei termini polisemici, e nelle interiezioni, viene tuttavia usualmente contenuta dal riferimento al contesto esterno, dagli elementi non verbali che co-occorrono nella comunicazione, dalla possibilità data dalla familiarità condivisa verso pratiche sociali e culturali, di inferire le intenzioni comunicative del locutore. Per dirla con De Mauro:

Una lingua deve servire agli esseri umani per trasmettere e comprendere sensi, per comunicare e verbalizzare nelle situazioni più differenti, per discutere come adattarsi a tali situazioni sopravvivendo e sopravvivendo come esseri umani, cioè come esseri capaci, poi, di nuovi, altrettanto imprevedibili riadattamenti. Per rispondere a queste esigenze, le lingue non possono essere rigidi calcoli, con i quali molto si può fare, ma non che con essi si rimetta in discussione essi stessi, la loro semantica, la loro sintassi. Invece le lingue consentono modi di produzione e comprensione dei segni oscillanti da un massimo di informalità, cioè di appello alle circostanze esterne al discorso e al testo, [...] ad un massimo di formalità, cioè di esplicitazione dei rapporti tra parti del segno, senso, emittente e destinatari.<sup>140</sup>

Alla luce di tali precisazioni in merito agli aspetti che differenziano la sintassi dei canti degli uccelli da quella del linguaggio umano, è opportuno chiedersi in che modo l'evoluzione abbia potuto guidare il passaggio dall'abilità di *sintassi* intesa in senso generale, come l'abilità di manipolare e percepire articolazioni di unità connesse in funzione di specifiche regole, all'abilità di sintassi tipicamente umana - che intesse (permeandosene) la

---

<sup>140</sup> De Mauro T., *Minisemantica*, Laterza, Roma-Bari, 2004, p. 138.

sfera semantica - e del suo potenziale di *significare* stati cosali del mondo. In tale direzione, crediamo sia opportuno mettere a fuoco l'analisi sulla natura delle unità relazionali all'interno dei pattern sintattici tanto dei canti degli uccelli, quanto delle proposizioni. Nei primi, come abbiamo visto, le unità interessate dal concatenamento regolato dai processi di Markov sono insiemi di due o tre note (*chunks*), mentre nel caso delle proposizioni, strutture dello stesso ordine di complessità intessono tanto la dimensione fonotattica delle parole (ovvero le regole di ammissibilità di sequenze fonetiche in una data lingua), quanto la possibilità di segmentare le parole ben formate all'interno di un flusso continuo di discorso. Tuttavia, le relazioni logico-sintattiche in funzione del quale si articolano sia i morfemi che le parole, all'interno delle proposizioni, vantano di una complessità che va oltre tale ordine di complessità, in quanto si estendono al dominio simbolico-concettuale.

Tali considerazioni ineriscono il piano strettamente *sintattico*, ovvero il piano dei pattern strutturali tra le unità costitutive, argomento pervasivo nella letteratura intorno alla evoluzione del linguaggio. Tuttavia, il fuoco di tale analisi merita di essere esteso su un aspetto spesso non opportunamente enfatizzato, per quanto centrale nelle riflessioni di Darwin. Scrive lo studioso:

With respect to the origin of articulate language [...] I cannot doubt that language owes its origin to the *imitation and modification* of various natural sounds [...]. Primeval man, or rather some early progenitor of man, probably first used his voice in *producing true musical cadences, that is in singing*, as do some of the gibbon-apes at the present day. [...]

It is not difficult to imagine the steps by which the notes of a bird, primarily used as a mere call or for some other purpose, might have been improved into a melodious love-song. This is somewhat more difficult in the case of the modified feathers, by which the drumming, whistling, or roaring noises are produced.<sup>141</sup>

Decentrando l'attenzione sulla linea segmentale delle proposizioni, si è mancato di riconoscere la cardinalità di un aspetto di significazione essenziale, strettamente intrecciato agli schemi strutturali, e talvolta - come vedremo - addirittura dotato di un valore prioritario nella acquisizione e comprensione di un linguaggio. Si tratta della dimensione prosodica, ovvero sovrasegmentale. Nel prossimo capitolo, focalizzeremo l'indagine sulla ipotesi darwiniana intorno alle pressioni selettive responsabili dell'evoluzione della complessità strutturale delle articolazioni. In seguito, a partire da tali riflessioni illustreremo il ruolo chiave della modulazione prosodica di una frase nella determinazione del suo senso. Secondo la nostra prospettiva di ricerca, la caratterizzazione prosodica delle frasi, infatti, trova le sue radici evolutive proprio nel fatto che le unità originariamente articolate consistevano, come afferma Darwin – in unità sonore prive di valore semantico, ma dotate di un potenziale

---

<sup>141</sup> Darwin, C., *The Descent of Man, and Selection in Relation to Sex*, John Murray, London, 1871, p. 56, 67.

intrinsecamente legato alla dimensione socio-emozionale. L'idea è dunque che la dimensione prosodica (musicale) dei proferimenti linguistici, co-originaria alle strutture sintattiche che ne regolano l'articolazione, costituisce un meccanismo di significazione ancora oggi in vigore nel linguaggio verbale. Nello specifico, essa agisce ad un livello che più che *sovra*-segmentale, definirei *inter*-segmentale, in quanto coabita nel tessuto proposizionale delle unità segmentali, ovvero le parole, evolutivamente recenti rispetto alla *sintassi prosodica*. Quest'ultima invero, come vedremo, guida non soltanto la segmentazione delle unità lessemiche in un flusso fonico per lo più continuo - ma ne suggerisce le potenziali connessioni con altre parole, amplificandone così la possibilità di acquisizione e comprensione. In altri termini, dunque, tale dimensione è in grado di modulare i processi di acquisizione di regole categoriali e di pattern statisticamente regolati. In tal modo, ne amplifica il valore di significazione, collocandone lo statuto al di là della mera articolazione a livello percettuale delle unità segmentali.

**SEZIONE II:**

**IL VALORE (SOCIO-)COGNITIVO**

**DELLA SINTASSI PROSODICA**

Il capire una proposizione è più affine di quanto non si creda al capire un brano musicale [...]. Non dimenticare come stanno le cose quando si legge una proposizione con la cadenza sbagliata, e perciò non la si capisce, - e poi ti viene in mente come si debba leggerla.

**L. WITGENSTEIN**

## 7. L'ipotesi di Darwin: alle origini era il canto

Uno degli aspetti per me più affascinanti della natura è la sua musicalità. È sufficiente fare una passeggiata all'aperto con degli amici, per poterne fruire la bellezza nei canti degli uccelli, o degli insetti. E in modo piuttosto interessante, persino nelle conversazioni fra gli amici stessi. Come non riconoscere infatti nei discorsi verbali una intrinseca musicalità, una costellazione di suoni e tonalità di cui le frasi si modulano, e che al pari dei suoni del resto della natura catturano l'attenzione dell'interlocutore, emozionandolo, o evocando particolari sensazioni? Chissà allora che la musica nei discorsi, quella stessa che durante le passeggiate in aperta natura si intreccia e sovrappone ai canti degli animali dell'ambiente esterno non trovi in essi degli interessanti correlati evolutivi. Una risposta autorevole a tale quesito è rintracciabile in *The Descent Of Man, and Selection in Relation to Sex* di Darwin, specificamente in quello che è attualmente uno dei passi più salienti della letteratura sull'evoluzione del linguaggio:

With respect to the origin of articulate language [...] I cannot doubt that language owes its origin to the *imitation and modification* of various natural sounds [...]. Primeval man, or rather some early progenitor of man, probably first used his voice in *producing true musical cadences, that is in singing*, as do some of the gibbon-apes at the present day; and we may conclude, from a

widely-spread analogy, that this power would have been especially exerted during the courtship between sexes, -- would have expressed various emotions, such as love, jealousy, triumph, -- and would have served as a challenge to rivals.<sup>142</sup>

L'idea dello studioso è, dunque, che, i primi ominidi, non ancora in possesso delle piene facoltà di linguaggio proposizionale comunicassero *musicando* le proprie emozioni; tali espressioni musicali avrebbero avuto, secondo tale prospettiva teorica, un ruolo biologicamente adattativo in quanto, al pari di quanto avviene nel mondo animale, latori di effetti determinanti nelle competizioni tra rivali tanto per la difesa del territorio, quanto per la conquista di un partner sessuale: entrambi ambiti decisivi per la trasmissione del patrimonio genetico, quindi per la sopravvivenza della specie. Le osservazioni di Darwin trovano oggi un'ampia conferma empirica negli studi realizzati nel campo dell'etologia cognitiva che, argomentandone la validità scientifica, ci consentono di ottenere una maggiore comprensione intorno alle dinamiche evolutive del linguaggio proposizionale. Vediamo meglio in che senso.

---

<sup>142</sup> Darwin, C., *The Descent of Man, and Selection in Relation to Sex*, cit., p. 56 (corsivi miei).



## 7.1 Duetti e legami di coppia

Il tema centrale del presente paragrafo è la nozione di “canto”, e il valore che Darwin ne ha messo in luce, relativamente alla loro realizzazioni in contesti legati all’attrazione o alla difesa di un partner sessuale. Per rigore metodologico, vorrei spiegare la definizione operativa di cui mi avvarrò per l’uso di tale termine. Con tale nome, ci si riferisce a vocalizzazioni (o produzione di suoni in generale) che a) comprendano più di due unità di tipo differente - ad esempio, almeno due tipi di note nel caso dei canti degli uccelli, b) che siano funzionali alla difesa del territorio, e alla conquista di un partner sessuale<sup>143</sup> c) la cui struttura, pur facendo capo a predisposizioni geneticamente determinate, sia flessibile, dunque in linea di principio acquisibile per apprendimento<sup>144</sup>.

Fra i primati non umani, l’unica famiglia a cui è attribuita la capacità di “cantare” è quella dei gibboni. In particolare, una caratteristica distintiva di quasi tutte le specie di gibboni, che, curiosamente, peraltro, vantano una struttura sociale monogama, rara fra i mammiferi - è la capacità di *creare* dei duetti a partire da vocalizzazioni specie-tipiche e dalla complessità

---

<sup>143</sup> Vedi Hall, M., *A Review of Vocal Duetting in Birds*, “Advances in the Study of Behavior”, 2009, pp. 67-120.

<sup>144</sup> Fitch, W. T., *On the Biology and Evolution of Music*, “Music Perception”, 24, 2006, pp. 85-88.

strutturale stereotipata. I partner di ogni coppia sincronizzano le proprie vocalizzazioni (geneticamente innate) *componendo* veri e propri “canti a due voci”. Thomas Geissmann, uno dei maggiori ricercatori sul tema, ne offre una dettagliata descrizione nel seguente passo:

Males of many gibbon species produce one or several distinct types of short phrases that often become gradually more complex (e.g. in the number of notes, number of distinct note types, degree of frequency modulation) as the song bout proceeds. At more or less regular intervals, females insert long, female-specific phrases that are commonly referred to as great calls. In most species, great calls consist of a particularly rhythmic series of long notes uttered with increasing tempo and/or increasing peak frequency. Males usually stop vocalizing at the beginning of each great call and provide a special reply phrase (coda) to the great call before resuming their more common short phrases. In addition, one or both partners often exhibit an acrobatic display at the climax of the great call, which may be accompanied by piloerection and branch shaking [...]. The combination of the female great call and the male coda is termed a great call sequence, and it may be repeated many times during a single song bout.<sup>145</sup>

In particolare, da una ulteriore ricerca<sup>146</sup> recentemente condotta dallo studioso sui siamanghi (*Hylobates syndactylus*) - primati della famiglia dei gibboni - emerge una correlazione positiva tra l'intensità quantitativa dei duetti cantati, e la qualità della relazione di coppia. Non appena instaurato il legame di coppia, infatti, i partner investono una cospicua quantità di tempo ed energia nell'esercizio necessario all'acquisizione di una buona

---

<sup>145</sup> Geissmann, T., *Gibbon songs and human music from an evolutionary perspective*, in N. L. Wallin, B. Merker, S. Brown, (ed.), “The Origins of Music”, cit., p. 107.

<sup>146</sup> Geissmann, T., *The relationship between duet songs and pair bonds in siamangs, Hylobates syndactylus*, “Animal Behaviour”, 2000, pp. 805-809.

coordinazione vocale. Tale investimento fungerebbe da deterrente per ogni tentativo di diserzione, in quanto li porterebbe a dover imparare dei nuovi schemi vocali con ogni nuovo compagno/a. Inoltre, l'intensità dell'attività di coordinazione vocale ha dei significativi effetti sul rafforzamento del legame. Geissmann, infatti, ha dimostrato che le coppie che dedicavano più tempo all'attività del duettare, tendevano a coprire spazi più ravvicinati e a praticare più *grooming* reciproco - con rimarchevoli effetti sulla sincronizzazione delle attività legate, ad esempio, alle cure parentali, alla ricerca di cibo, o alla difesa del territorio. Dato, inoltre, il caratteristico volume alto, fra le funzioni attribuite ai duetti vengono elencate la segnalazione ai vicini del nuovo legame di coppia, e, allo stesso tempo, la difesa del territorio occupato dalla stessa. Come nota Darwin, i duetti realizzati dai gibboni rappresentano un indizio molto importante per risalire alle dinamiche dell'evoluzione del linguaggio. Tuttavia, per guadagnare un'analisi più dettagliata del fenomeno, è opportuno spostare il fuoco dell'attenzione dai primati non umani ad alcune specie di uccelli canori che, sepur filogeneticamente più distanti dall'uomo, presentano caratteristiche analoghe assai significative. È ciò di cui ci occuperemo nel prossimo paragrafo.

## 7.2 La selezione dei canti

Una volta illustrati gli studi sui tipi di complessità strutturale che diverse specie di scimmie ed uccelli canori sono in grado tanto di percepire, quanto di produrre, ci apprestiamo a mostrare quali vantaggi evolutivi possa fornire non soltanto l'abilità di duettare, ma quella di “cantare” in generale. In particolare, nel presente paragrafo focalizzeremo l'attenzione sull'evoluzione dei canti degli uccelli, la cui tematizzazione può rivelare la presenza di rimarchevoli tratti evolutivi analoghi, in grado di aggiungere dei tasselli importanti alla nostra ricerca.

È ormai assodato, nella letteratura scientifica, che la complessità strutturale e la larghezza del repertorio dei canti degli uccelli abbiano importanti effetti su almeno tre fronti: la conquista di una femmina, la difesa del territorio, la coesione intragruppale. Per quanto concerne il primo punto, sono interessanti gli studi effettuati da Kazuo Okanoya sui canti di molte specie di uccelli, e in particolare del passero del Giappone (*Lonchura striata domestica*). Nello specifico, l'autore sostiene che i canti di tale specie di uccelli - che presentano la struttura sintattica interna ben definita (che abbiamo analizzato nel paragrafo 6.3), oltre che un vasto repertorio di suoni da combinare – siano stati interessati da una dinamica di selezione evolutiva, proprio in quanto elementi sessualmente attrattivi. Per

dimostrare tale ipotesi, il gruppo di ricerca di Okanoya ha controllato sperimentalmente la presenza di quattro fattori indicativi nelle femmine: il ritmo del battito cardiaco, i livelli di ormoni riproduttivi, i comportamenti recettivi, e l'attività di nidificazione. La presentazione in *playback* dei canti, in effetti, induceva l'aumento del battito cardiaco nelle femmine, ma non nei maschi dei passerini del Giappone (reazione che non occorre alla presentazione del canto di una specie diversa di uccelli canori). Come osserva l'autore:

Presentation of heterospecific (zebra finch) songs did not increase the heart rate. When a novel conspecific song was presented, the heart rate increased only in female birds with each presentation of the stimulus, but not in males (...). These results correspond to the sex differences in response to song in this species; males ignore other birds' songs while females are attentive. (...) Results on females suggest that syntactically complex songs may be more potent than simple songs in keeping female finches in an aroused state.<sup>147</sup>

L'idea è dunque che lo stato di forte arousal induca una maggiore recettività in termini sessuali. Ciò converge con la registrazione della presenza di alti livelli di ormoni riproduttivi, la cui produzione veniva stimolata proprio dall'esposizione ai canti<sup>148</sup>. Inoltre lo stato di forte arousal rinvierebbe ad una maggiore attentività verso le qualità canore del maschio. Come hanno dimostrato diversi test di controllo, le femmine sono

---

<sup>147</sup> Okanoya, K., *Song Syntax in Bengalese Finches: Proximate and Ultimate Analyses*, cit., p. 330.

<sup>148</sup> *Ivi*, p. 334.

tendenzialmente predisposte a preferire i canti dalla complessità maggiore<sup>149</sup>, i quali rappresenterebbero l'indice manifesto di qualità cognitive e fisiche<sup>150</sup>. Di fatti, l'abilità di realizzare canti lunghi e strutturati implica la presenza di una buona memoria, e della capacità di coordinazione neuro-motoria, necessarie all'articolazione delle note entro schemi sintattici discreti.

Un ulteriore dato emergente dalle ricerche sperimentali condotte su tale idea riguarda l'attività di nidificazione. L'ipotesi della selezione della complessità dei canti per scopi riproduttivi è confermata dalla osservazione della quantità di materiale per la costruzione del nido trasportato dalle femmine:

---

<sup>149</sup> In tale direzione vanno anche le ricerche effettuate da Nuebauer, R., il quale in uno studio descritto in *Super-normal length song preferences of female zebra finches (Taeniopygia guttata) and a theory of the evolution of bird song*, "Evolutionary Ecology", 13, 1999, pp. 365-380 – dimostra che le femmine dei diamanti mandarini preferiscono non soltanto i canti con una maggiore eterogeneità strutturale interna, ma anche dalla lunghezza fuori dal comune. Inoltre, da uno studio condotto dal gruppo di ricerca di Okanoya, emerge che i giovani passeri del Giappone, se esposti a canti dalla complessità strutturale diversa, ciascuno realizzato da un diverso adulto, tendenzialmente scelgono di "imitare" il canto internamente più eterogeneo. A tal proposito, cfr. Soma, M., Hiraiwa-Hasegawa, M., Okanoya, K., *Song-learning strategies in the Bengalese finch: do chicks choose tutors based on song complexity*, "Animal Behaviour", 78, 2009, p. 1113: "If we define song evolution in a broad sense as the accumulation of particular song phenotypes that have been transmitted through generations, the following process is responsible for its dynamics. When females choose mates based on male song trait elaboration, males with attractive songs produce more offspring (Catchpole & Slater 2008). As a result, attractive songs are more likely to be transmitted culturally to subsequent generations, and the song-learning capacity that enables individual birds to acquire elaborate song traits are more likely to be inherited by the offspring. Although the criteria used by young birds to choose a model for the development of their own songs are not fully understood, this study suggests that a cognitive basis for social song learning might be sexually selected to favour the acquisition of more complex songs which, in turn, accelerates song evolution."

<sup>150</sup> Tuttavia, è bene tenere presente che in molte specie di uccelli canori le femmine realizzano dei canti sia da sole che in duetti al fine della difesa territoriale, in competizioni fra femmine per la conquista/difesa di un pattern sessuale, o per alimentare la coesione gruppale. A tal proposito, vedi Langmore, N., *Functions of duet and solo songs of female birds*, "Trends in Ecology & Evolution", 13, 1998, pp. 136-140.

Three groups of female Bengalese finches were studied. Each group consisted of four finches, separately caged, and kept together in a sound isolation box. The first group was stimulated with group was not stimulated by any song. The number of items of nesting material carried each day when stimulated with complex or simple syntax songs was counted and compared. Female finches carried more nesting material (...). We examined further whether randomly generated note sequences were more effective than syntactically synthesized ones. Females stimulated with random note sequences were less responsive and were comparable to the females stimulated with the simple sequence. Although random sequences resulted in complex ordering of the song notes, *randomness was not the same as the complexity produced by syntax for the females.*<sup>151</sup>

In ultima analisi, dunque, l'idea che l'evoluzione della complessità strutturale di articolazioni sonore trovi le sue radici nella selezione sessuale è, come abbiamo visto, ben giustificata empiricamente dagli studi condotti sugli uccelli canori. In un interessante articolo sull'argomento, William Searcy e Malte Anderson<sup>152</sup>, chiedendosi quali vantaggi evolutivi abbiano guidato il processo in grado di selezionare tali tratti, hanno notato come la complessità dei canti possa rappresentare un importante indice dei "benefici materiali" apportabili dal maschio:

The size of the song repertoire in red-winged black birds increases with years of breeding experience [...], and experienced males are more likely to feed nestlings [...] and on average have larger, better territories. [...]  
Captive male yellow-bellied sunbirds (*Nectarinia venusta*) vary their rate of singing [...] and their use of certain song phrases [...] according to the concentration of sugar solution provided to them; this suggests that singing behavior might signal abundance of resources on natural territories. (p. 518)

---

<sup>151</sup> *Ivi*, p. 333 (corsivo mio).

<sup>152</sup> Searcy, W. A., Andersson, M., *Sexual selection and the evolution of song*, "Annual Review of Ecology and Systematics", 17, 1986, pp. 507-533.

Inoltre, i canti degli uccelli veicolano importanti informazioni inerenti le dimensioni del corpo, l'età, e lo stato di dominanza del maschio. Di conseguenza, le femmine in grado di decifrare tali indicatori, scegliendone gli esecutori come partner sessuali - avrebbero accresciuto la capacità di adattamento della prole. La complessità dei canti ha assunto un ruolo decisivo nel contesto delle sfide fra i maschi, non soltanto in relazione alla contesa delle femmine, ma anche per la difesa o la conquista dei territori. Ciò è dovuto al fatto che i canti spesso codificano l'identità di specie, cosicché i maschi reagiscono meno aggressivamente a quegli individui appartenenti ad una specie diversa (in quanto non rivali per la contesa della femmina), secondo quanto può essere rivelato dal “dialetto”. Inoltre la capacità di articolare un repertorio vasto e internamente articolato può indurre nel maschio nemico la erronea rappresentazione della presenza di più individui nel territorio minacciato. In effetti, proprio la segnalazione della presenza di un gruppo – oltre che la segnalazione dell'appartenenza ad esso – è una componente centrale nella cognizione degli uccelli canori.



### **7.3 Vincoli ed analogie evolutive: il ruolo delle interazioni sociali**

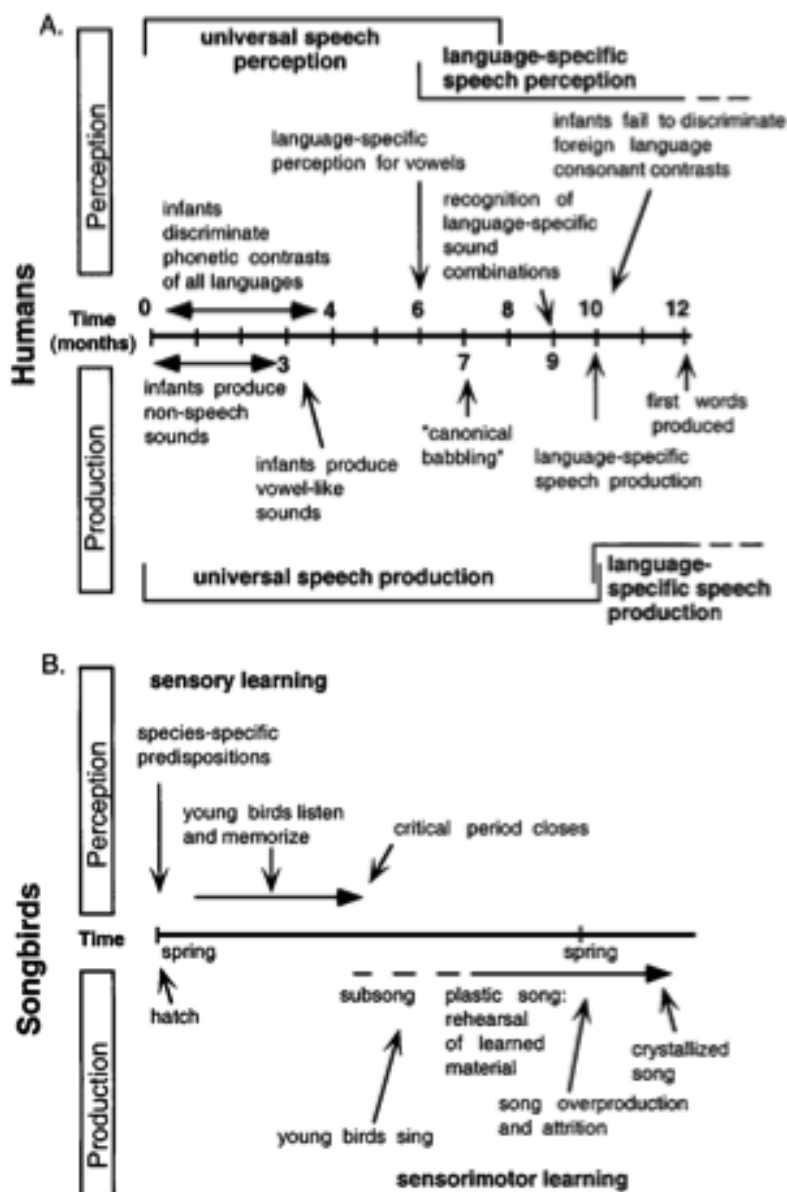
Un aspetto assai significativo che accomuna le espressioni comunicative di uccelli, scimmie ed esseri umani inerisce le loro venature socio-emozionali. A tale proposito, vale la pena di illustrare le riflessioni di Marler<sup>153</sup> il quale, mettendo a confronto i *pant-hoots* e i canti degli uccelli, ha individuato gli aspetti condivisi che passano tra forme di linguaggio evolutivamente imparentate con il linguaggio umano. L'autore nota che entrambi sono di natura affettiva, emessi in uno stato di forte *arousal*, usati per modulare spazi e contatti sociali; tali aspetti messi in luce dall'autore ci forniscono, crediamo, degli indizi significativi circa la centralità del ruolo della dimensione sociale ed emozionale nella dinamica dell'evoluzione del linguaggio. Tuttavia, nonostante la maggiore distanza filogenetica, le convergenze evolutive più marcate interessano la specie umana e gli uccelli canori.

In primo luogo, ad esempio, in entrambi i generi le interazioni socio-emozionali giocano un ruolo chiave nell'acquisizione ontogenetica del linguaggio. In effetti, tanto gli infanti preverbali quanto i piccoli degli

---

<sup>153</sup> Marler, P., *Origins of Music and Speech: Insights from Animals*, in N. L. Wallin, B. Merker, S. Brown, (ed.), "The Origins of Music", cit., pp. 31-48.

uccelli canori, attraverso un continuo feedback con i tutor, perfezioneranno le rispettive espressioni vocali, fino a giungere rispettivamente alla piena padronanza del linguaggio proposizionale e alla produzione del “canto cristallizzato”. In modo interessante, in entrambi i generi è possibile notare la presenza di specifiche fasi di produzione vocale (figura 1); la prima di queste fasi è la lallazione nei bambini, a cui corrisponde la fase di precanto (“*subsong*”) negli uccelli; in questa prima fase di acquisizione del linguaggio, i piccoli producono e ripetono stringhe minime di suoni. Ad essa segue una tappa di perfezionamento in cui gli infanti producono enunciati minimi e, analogamente, gli uccelli producono le cosiddette “*plastic songs*”, in cui il canto viene mischiato a suoni e frasi apprese dagli adulti. Inoltre, lo sviluppo delle competenze linguistico-vocali ha luogo con maggiore naturalezza - tanto gli uccelli quanto gli umani - nella finestra temporale che va dalla nascita al periodo pre-adolescenziale; tale periodo viene detto “critico”, in quanto la mancata esposizione ad un “linguaggio” - sia esso una lingua, nel caso degli umani, o dei canti, nel caso degli uccelli - in tale fascia temporale porta ad una drastica compromissione della produzione/percezione di suoni e accenti tipici del gruppo di appartenenza, fenomeno evidente in quegli individui che si apprestano ad imparare una lingua nuova nella fase adulta.



**Fig. 1.** Immagine modificata da Doupe *et al.* (1999). Gli autori mettono in evidenza le principali tappe di apprendimento di un linguaggio sia sul piano della percezione che della produzione – negli umani (A) e negli uccelli canori (B)

Tale considerazione mette in risalto un ulteriore parallelismo cognitivo tra umani ed uccelli canori: in entrambi i generi, l'identità di gruppo può essere rivelata dal "dialetto", particolari costellazioni di caratteristiche acustiche di cui le vocalizzazioni canore si compongono, rivelando l'area geografica di appartenenza del gruppo<sup>154</sup>. È opportuno inoltre sottolineare che nel graduale processo di acquisizione del linguaggio, in entrambi i generi, la presenza fisica di un tutor che offra un estemporaneo feedback di natura uditiva o visuo-gestuale è di cardinale importanza. In modi diversi, a seconda della specie, negli uccelli canori la presenza fisica di un tutor, infatti, influenza in maniera decisiva l'acquisizione dei canti. Nel seguente passo, Kuhl fornisce una buona panoramica su come ciò avvenga in diverse specie di uccelli:

[...] social contexts can advance song production in birds; male cowbirds respond to the social gestures and displays of females, which affect the rate, quality, and retention of song elements in their repertoires, and white-crowned sparrow tutors provide acoustic feedback that affects the repertoires of young birds. Additional evidence of the impact of social interaction can be adduced from the sensory learning phase in birds. In young zebra finches, visual interaction with a tutor bird

---

<sup>154</sup> In tale direzione, è opportuno citare Crockford, C., Herbinger, I., Vigilant, L. (2004). *Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning*, "Ethology", 110, 2004, pp. 221-243 - il quale rivela che gli scimpanzé siano in grado di modificarne attivamente l'ampiezza e la durata, e, di conseguenza, di farne un indice della comunità di appartenenza – pp. 237-238: "Our results, then, support the learning hypothesis, as the existence of community specific pant hoots between neighbours, but not strangers, suggests that chimpanzees are actively modifying their pant hoots to diverge among neighbours. [...] Chimpanzees must be modifying their pant hoot structure so that it converges with other community members' pant hoots, thus social influence is inherent. [...] Therefore, in using a community specific pant hoot, chimpanzees may be selecting a narrow range from a wide range of possible graded pant hoot types in their repertoire." Così, come nel caso negli uccelli canori, anche negli scimpanzé la capacità di controllo vocale si sarebbe evoluta per ragioni legate alla difesa del territorio, e per marcare l'identità di gruppo.

is typically required to learn [...]; in fact, the impact of social interaction is so potent that young zebra finches will learn an alien song from a Bengalese finch foster father who feeds them [...]. White-crown sparrows will also learn an alien song from a live tutor, even though they reject those songs when presented on audio tape. In barn owls (10) and in white-crowned sparrows [...], the duration of the sensitive period for learning is altered by the richness of their social environments.<sup>155</sup>

In una prospettiva teorica comparativa, il gruppo di ricerca di Micheal Goldstein<sup>156</sup> ha realizzato uno studio finalizzato alla comprensione del modo in cui la mera interazione sociale - realizzata attraverso risposte non vocali come sorrisi, l'agitazione di un giocattolo, la vicinanza e il contatto del corpo, da parte del tutor (un genitore) influenzi l'acquisizione del linguaggio nell'infante preverbale. Nello specifico, gli studiosi hanno realizzato due condizioni sperimentali: una in cui il genitore interagiva con il neonato in maniera spontanea, ovvero rispondendo contingentemente alle sue vocalizzazioni, attraverso le espressioni corporali. Nella seconda condizione, i tutor erano invitati ad attuare comportamenti fisicamente interattivi non sincronizzati con le vocalizzazioni degli infanti, ovvero non come risposta spontanea alle specifiche espressioni dei neonati, ma in modo del tutto random. Ne è emerso che nonostante la breve durata dell'esperimento (appena 30 minuti), i proferimenti vocali degli infanti

---

<sup>155</sup> Kuhl, P. K., *Human speech and birdsong: Communication and the social brain*, "Proceedings of the National Academy of Science", 100, 2003, pp. 9645-9646.

<sup>156</sup> Goldstein, M., King, A., *Social interaction shapes babbling: Testing parallels between birdsong and speech*, "Proceedings of the National Academy of Science", 100, 2003, pp. 8030-8035.

della prima condizione erano nettamente migliori in termini di strutturazione interna, volume, accuratezza nella transizione tra consonanti e vocali, velocità di produzione. Così gli autori commentano tale risultato:

The idea that contingency perception can both reinforce and induce new forms of behavior stands in contrast to the current paradigm of human social learning. The current view relies on imitation, the copying of others' behavior, as a mechanism of socially mediated developmental change [...]. Although studies show that behaviors taken to be evidence of imitative learning are really the products of more general processes like arousal [...], little current research on social development seems to emphasize nonimitative learning [...]. As this study demonstrated, infants' babbling changed in a lawful way without mothers providing behaviors that could serve as the basis for vocal imitation; [...] These data contain striking parallels to social mechanisms of vocal learning in several species of songbirds. [...] Like human infants, sounds receiving social attention are more likely to recur. Contingency of social reinforcement is key: [...] In both taxa, vocal precursors have functional significance as part of a system of reciprocal influences between a young learner with a variable repertoire and a social environment that affords structured feedback.<sup>157</sup>

La descrizione delle convergenze evolutive nella cognizione degli uccelli canori e degli umani, se da una parte rappresenta un indizio assai rilevante sulle pressioni selettive che hanno determinato l'evoluzione del linguaggio verbale, dall'altro mantiene aperte alcune domande decisive. Sebbene, infatti, da una parte la tesi secondo la quale la capacità di linguaggio negli umani trovi le sue radici nell'evoluzione delle complessità strutturali di note trovi una buona giustificazione empirica nello studio dei numerosi tratti analoghi evolutisi negli uccelli canori – dall'altra emergono delle differenze sostanziali tra i rispettivi sistemi di comunicazione, che ne

---

<sup>157</sup> *Ivi*, pp. 8033-8034.

segnano anche importanti divergenze. In tale ambito di indagine troviamo assai proficue, ancora una volta, le osservazioni di Okanoya, il quale sulla scia della teoria darwiniana, afferma che la selezione sessuale avrebbe operato - negli uomini quanto negli uccelli canori anzitutto su un piano meramente sintattico. Ciò vuol dire che nel suo primissimo stadio evolutivo, il linguaggio umano era verosimilmente scevro da qualsiasi potenziale semantico, ma sintatticamente strutturato. Così, l'autore ritiene plausibile l'idea che l'origine del linguaggio verbale sia da identificare in determinate complessità strutturali foniche, ovvero di articolazioni sintattiche che, inizialmente selezionate per fini riproduttivi, *in un secondo momento* avrebbero acquisito un valore semantico discreto<sup>158</sup>.

La nostra ipotesi è che la caratterizzazione prosodica delle sequenze sintattiche (ovvero la modulazione in tempo, frequenza, ampiezza e ritmo dei proferimenti) abbia giocato un ruolo essenziale nell'acquisizione del valore semantico, ovvero nella evoluzione del linguaggio nella sua attuale configurazione. L'esame del concetto di sintassi - inteso, come abbiamo visto, in senso ampio come organizzazione di sequenze di stimoli in base a determinate regole - unitamente a quello di modulazione prosodica

---

<sup>158</sup> Okanoya, K., *Sexual Display as a Syntactical Vehicle: The Evolution of Syntax in Birdsong and Human Language through Sexual Selection*, in A.Wray, "The Transition to Language", Oxford University Press, 2002, pp. 46-47: "I argue that sexual display could be a pre-adaptation for syntax, that sexual display first evolved a finite-state syntax, and that semantic tokens were later inserted as elements of this finite-state syntax."

costituiscono dunque la chiave di violino della presente ricerca, in quanto al contempo essenziali sia nello studio sull'origine del linguaggio, che nella individuazione di quei tratti cognitivi che, differenziandoci dalle altre specie animali, ci contraddistinguono in quanto umani.



## **8. Strutture prosodiche alle origini del linguaggio**

### **8.1 Animali non umani: riconoscimento e uso di indizi prosodici**

Al fine di guadagnare un quadro teorico quanto più completo e nitido dei processi cognitivi alla base della facoltà di linguaggio, e delle sue dinamiche evolutive, è opportuno integrare gli studi comparativi fin qui presi in esame sulle abilità computazionali con un aspetto essenziale della comunicazione animale: la dimensione prosodica. Gli studi inerenti l'abilità di riconoscere indizi o pattern prosodici negli animali non umani sono molto esigui, seppur assai significativi. Vale la pena di riportare, ad esempio, lo studio comparativo condotto da Ramus<sup>159</sup> sull'abilità di discriminare lingue diverse in base alle caratteristiche prosodiche. Nello

---

<sup>159</sup> Ramus, F., Hauser, M. D., Miller, C., Morris, D., Mehler, J., *Language discrimination by human newborns and by cotton-top tamarin monkeys*, "Science", 288, 2000, pp. 349-351.

specifico, gli studiosi hanno esposto un gruppo di neonati umani, e uno di tamarini dal ciuffo bianco ad un insieme di frasi proferite in due lingue diverse – ovvero, giapponese e olandese – in tre diverse condizioni sperimentali: a) utilizzando le medesime frasi, ma proferite da persone di genere diverso, b) utilizzando la sola linea prosodica di tali frasi, estraendola artificialmente da informazioni di carattere lessicale o fonetico, c) riproducendo le frasi al contrario. Da tale studio emerge che tanto i tamarini quanto i neonati vantano la capacità di discriminare le frasi proferite nelle due diverse lingue (sebbene gli umani mostrino maggiori difficoltà rispetto ai tamarini laddove esposti alle frasi proferite da più parlanti di genere diverso). In particolare, entrambe le specie falliscono nel discriminare la differenza di lingua nella terza condizione, ovvero laddove le frasi venivano riprodotte al contrario. Ciò rappresenta un chiaro indizio del fatto che tanto gli umani quanto i tamarini utilizzano pattern di informazioni prosodicamente regolate nel discriminare lingue storico-naturali distinte. Così gli studiosi commentano i propri risultati:

The pattern of our results suggests striking similarities as well as differences between the monkey and the human auditory systems. First, we have shown that tamarins, like human newborns, are able to process not just isolated syllables but also whole strings of continuous speech and to extract enough information to discriminate between Dutch and Japanese. Second, their ability to do so above and beyond speaker variability suggests that they are able to extract auditory equivalence classes - that is, to extract abstract linguistic invariants despite highly variable acoustic shapes [...]. Third, the fact that, like newborns, tamarins fail to discriminate when speech is played backward suggests that their language

discrimination capacity relies not on trivial low-level cues but rather on quite specific properties of speech. Because tamarins have not evolved to process speech, we infer in turn that at least some aspects of human speech perception may have built upon preexisting sensitivities of the primate auditory system.<sup>160</sup>

Dunque, la similarità dei risultati relativi alla percezione delle informazioni prosodiche nelle due specie dimostra che taluni tratti cognitivi sensibili alla percezione auditiva delle informazioni prosodiche del segnale vocale sono comuni alle due specie di primati. Inoltre, la presenza di tale tratto in una specie di scimmia del nuovo mondo rivela una traccia cronologica dell'evoluzione di tale abilità, verosimilmente presente nella linea filogenetica dell'uomo già 48 milioni di anni fa. Le capacità prosodiche sembrano così trovare le proprie radici evolutive in un periodo assai remoto, a conferma dell'originarietà di tale tratto cognitivo.

Al fine di ottenere una più estesa conoscenza sulle origini di tale essenziale proprietà costitutiva del linguaggio, Toro *et al.* (2003), hanno replicato il lavoro di ricerca condotto da Ramus *et al.*<sup>161</sup> sui ratti - così da appurare se oltre che appannaggio dei primati esso costituisca anche una competenza presente nei mammiferi in generale. È interessante notare che da tale studio emerge che anche i ratti siano in grado di discriminare frasi proferite in lingue diverse, alla stregua degli infanti umani, e dei tamarini.

---

<sup>160</sup> *Ivi*, p. 351.

<sup>161</sup> Toro, J, Trobalon, J., Sebastian-Gallés, N., *The use of prosodic cues in language discrimination tasks by rats*, "Animal Cognition", 6, 2003, pp. 131-136.

Tuttavia, ulteriori studi sono necessari al fine di render chiaro quali particolari aspetti prosodici vengono processati dalle diverse specie in questione – ovvero quali tratti rendono le frasi appartenenti a due diverse lingue discriminabili soltanto se proferite secondo la regolare successione di suoni, ma non se riprodotte al contrario. In particolare, sono necessarie ulteriori indagini volte a capire il ruolo dei diversi valori fonetici, oltre che alle proprietà specificamente prosodiche come il ritmo proposizionale, la durata dei suoni, o la curva del profilo intonazionale.

Converge in tale direzione uno studio effettuato da Lisa Heimabauer su uno scimpanzè allevato dagli umani fin dai suoi primissimi giorni di vita, Panzi, dal titolo “A Chimpanzee Recognizes Synthetic Speech with Significantly Reduced Acoustic Cues to Phonetic Content”<sup>162</sup>. La studiosa ha creato due versioni acustiche, artificialmente modificate, delle parole “apricot” e “sparkler”, ovvero una versione sinusoidale, ed una detta “*noise-vocoded*”; la prima presentava tre suoni puri che in qualche modo tracciavano i *formants*<sup>163</sup> più bassi di ciascuna parola, mentre nella versione “*noise-vocoded*” le parole suonavano come rumore bianco<sup>164</sup>, in cui

---

<sup>162</sup> Heimabauer, L., Beran, M., *A Chimpanzee Recognizes Synthetic Speech with Significantly Reduced Acoustic Cues to Phonetic Content*, “Current Biology”, 21, 2011, pp. 1210-1214.

<sup>163</sup> Concentrazioni di energia acustica intorno a particolari frequenze nel tratto vocalico, cui corrispondono i suoni vocalici. Per una chiara spiegazione del termine, all’interno di una prospettiva comparativa, vedi Fitch, W. T., *The evolution of speech: a comparative review*, “Trends in Cognitive Sciences”, 4, 2000, pp. 258-267.

<sup>164</sup> Particolare tipo di rumore che comprende l’intero spettro delle frequenze, presentando dunque assenza di periodicità nel tempo e di ampiezza costante. Per intenderci, un esempio di rumore bianco è quello generato da una televisione mal sintonizzata.

venivano preservate informazioni legate alla dimensione temporale della successione dei suoni (ritmo), e indizi assai rudimentali inerenti le frequenze proprie delle date parole. Panzi è riuscito ad identificare le parole prodotte in tre diverse condizioni (linguaggio naturale, realizzazione *noise-vocoded*, e sinusoidale), associandole al corretto lessigramma corrispondente. Inoltre, mentre lo scimpanzè mostrava maggiore accuratezza nell'identificare le versioni *noise-vocoded* delle parole rispetto a quelle sinusoidali, le prestazione dei soggetti umani coinvolti nel medesimo test, sono state interessate da un andamento opposto, mostrando più precisione nel riconoscimento della versione sinusoidale. Tali dati vengono così commentati dagli autori:

On the one hand, these findings do not show equivalence of processing in humans and the chimpanzee, nor do they rule out the possibility that humans are specialized for speech perception. On the other hand, Panzee's performance does suggest that even quite sophisticated human speech perception phenomena may be within reach for some nonhumans. Such outcomes might, for example, reflect vocal-perception mechanisms shared with chimpanzees or other great apes [...]. [...] Overall, our findings at least suggest that the perceptual capacities required for speech processing in the reduced acoustic forms tested here may be traceable as far back as the common ancestor of chimpanzees and humans—and perhaps even further.<sup>165</sup>

In altri termini, gli autori sostengono che la realizzazione fonica delle lingue presenta in sè delle informazioni verosimilmente legate *in primis*

---

<sup>165</sup> Heimbauer, L., Beran, M., cit., p. 1213.

alla dimensione soprasegmentale, la cui facoltà di percezione può avere radici evolutive temporalmente lontane, e sulla base delle quali ha preso avvio il lungo processo di specializzazione nella percezione (e comprensione) specie-specifica umana dei suoni vocali.

Tali studi trovano un'importante supporto in una ricerca effettuata dal gruppo di ricerca di Robert Shannon sul tipo di informazioni primariamente implicate nel riconoscimento delle unità di una lingua da parte di soggetti umani. Questi, infatti, esposti a degli stimoli linguistici modificati artificialmente in modo da preservare il pattern temporale dello stimolo naturale, ed in cui le informazioni inerenti lo spettro delle frequenze erano notevolmente ridotte (similmente al caso degli stimoli utilizzati nell'esperimento su Panzi) sono riusciti ad identificare le parole da cui derivava tale modificazione artificiale per il 90% delle volte:

No formant structure was present, and formant frequency transitions either were lost completely (if they took place wholly within one of the present analysis bands) or were conveyed as a temporal change in the relative sound level of two adjacent noise bands. [...] Despite this reduced spectral content, the temporal cues were sufficient to produce 90% correct identification of words. This result indicates that minimal spectral information is required for speech recognition as long as temporal cues are available in a few contiguous regions.<sup>166</sup>

---

<sup>166</sup> Shannon, R., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M., *Speech Recognition with Primarily Temporal Cues*, "Science", 270, 1995, 270, pp. 303-304.

Ancora, relativamente alle competenze di natura prosodica diffuse nel regno animale, ulteriori dati assai significativi per i propositi del presente studio provengono ancora una volta dagli uccelli canori. Studi effettuati sul campo da Takahasi<sup>167</sup> dimostrano infatti che i giovani di uccelli usano comporre i loro canti assemblando parti udite da diversi tutor, con evidenti effetti sul piano della difesa territoriale e dell'attrazione di partner sessuali:

The ability to segment and recombine is adaptive to achieve a hierarchical song structure, rather than an exact copy, of songs. Segmentation and recombination are adaptive in territorial context and in sexual selection. In territorial defense, a song that has been compounded from two birds might suggest that two birds are singing. In sexual selection, mixed song may be able to have several features preferred by females. Birds can increase the apparent song complexity by utilizing chunked segments and recombining these in several different ways.<sup>168</sup>

Il punto interessante è che l'abilità di assimilare le parti dei diversi tutor presuppone la capacità di segmentare i canti di questi ultimi in parti diverse, ed in tale processo, è possibile che gli uccelli si servano, oltre che delle proprietà statisticamente regolate di distribuzione delle sillabe, anche di valori prosodici.

In ultima analisi possiamo così affermare che in diverse specie non umane, l'uso di indizi prosodici rappresenta una guida imprescindibile nella discriminazione di unità e pattern fonici di tipo diverso. È interessante

---

<sup>167</sup> Takahasi, M., Yamada, H., & Okanoya, K. *Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches* (*Lonchura striata* var. *domestica*), "Ethology", 116, 2010, pp. 481–489.

<sup>168</sup> *Ivi*, p. 487.

notare, in tale direzione, come i proferimenti vocali che gli uomini rivolgono agli animali presentino caratteristiche prosodiche esageratamente marcate. In modo interessante, tale caratterizzazione prosodica presenta significative similarità con i discorsi diretti agli infanti pre-verbali. Ciò è in effetti quanto emerge da un recente studio in cui Burnham ed il suo *team* di ricerca<sup>169</sup>, tracciano nei discorsi riferiti ai bambini e agli animali la presenza di quattro diverse caratteristiche condivise: il tono, l'intonazione e il ritmo (isolati artificialmente dalle informazioni lessicali e fonetiche) – e l'iperarticolazione delle vocali; gli autori commentano i risultati di tale ricerca nei seguenti termini (figura 1):

These results show that infant- and pet-directed speech are similar and distinctly different from adult-directed speech in terms of heightened pitch and affect. Interestingly, only infant- directed speech contains hyperarticulated vowels. Thus, vowel hyperarticulation does not accompany special registers simply because they differ from adult speech in pitch and affect. Rather, it seems to be a didactic device: Mothers exaggerate their vowels for their infants but not for their pets. Evidently, speakers are sensitive to their audience, both in regard to acoustic preferences and emotional needs, and in terms of potential linguistic ability. [...] In conclusion, we propose that (i) special speech registers differ in their mix of acoustic, affective, and vowel hyperarticulation components and (ii) speakers intuitively perceive the emotional and linguistic needs of their audience and automatically adjust their mix of speech components accordingly.<sup>170</sup>

---

<sup>169</sup> Burnham, D., Kitamura, C. *What's new, pussycat? On talking to babies and animals*, "Science", 296, 2002, p. 1435.

<sup>170</sup> *Ibidem*.



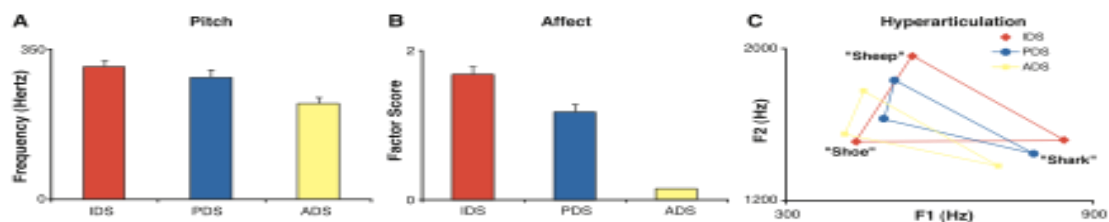


Fig. 1. (A) Pitch (fundamental frequency in hertz), (B) rated affect, and (C) vowel hyperarticulation (in  $F_1$ - $F_2$  vowel space) in infant-, pet-, and adult-directed speech (IDS, PDS, and ADS, respectively).

**Fig. 1.** Immagine modificata da Burnham *et al.* (2002). I grafici mostrano come i discorsi diretti ai bambini (IDS) e agli animali domestici (PDS), rispetto a quelli rivolti agli adulti (ADS):

- (A) presentino una frequenza in Hz più elevata,
- (B) siano percepiti come maggiormente carichi d'affetto, e
- (C) in essi le vocali siano maggiormente articolate.

In altri termini, dunque, gli autori assumono che, dati gli effetti sul piano emozionale dell'esagerazione dell'altezza del tono - essa viene impiegata nelle espressioni vocali rivolte tanto ai bambini quanto agli animali, notoriamente sensibili alla espressività emozionale. D'altro canto, però, l'esagerata articolazione delle vocali sarebbe funzionale ad una migliore percezione dei segmenti linguistici, ed in quanto tale impiegata principalmente nei discorsi rivolti agli infanti. È opportuno, a questo punto, esaminare più da vicino il ruolo della modulazione prosodica dei proferimenti nel facilitare la processazione del discorso (quindi, dell'acquisizione del linguaggio), e quali specifici parametri fonici veicolano tale effetto positivo. Sarà questo l'oggetto di studio dei prossimi paragrafi, contestualmente all'analisi dei discorsi rivolti ai bambini pre-verbali.

## **8.2 La processazione dei segnali linguistici: indizi che fanno la differenza**

Il processo di acquisizione del linguaggio, come abbiamo visto nel precedente capitolo, fa capo ad importanti strategie di segmentazione di unità lessemiche entro un flusso continuo di suoni. La presenza di particolari competenze cognitive analoghe, o omologhe in altre specie animali dimostra che i relativi meccanismi trovano le proprie radici nel dominio della dotazione neurobiologica geneticamente determinata. Tuttavia, posto ciò, è necessario approfondire l'analisi degli aspetti degli stimoli esterni verso (e dai) quali tali meccanismi innati sono orientati. Gli studi di ricerca comparativa effettuati su tale tema si collocano in un paradigma teorico che assume la centralità del ruolo dei processi computazionali (relative all'apprendimento statistico e alle grammatiche chomskiane) che stanno alla base del riconoscimento delle parole, spesso celando il ruolo decisivo di un aspetto fondamentale del linguaggio stesso: le proprietà prosodiche dei proferimenti, ovvero l'intonazione, il ritmo, la durata e l'accento (inteso come rafforzamento dell'intensità della voce) delle loro unità costitutive.

È opportuno osservare che gli studi comparativi volti a indagare le abilità delle specie non umane di riconoscere i tratti prosodici di un

linguaggio artificiale sono in numero nettamente inferiore rispetto agli studi basati sul paradigma chomskiano e dell'apprendimento statistico. Tuttavia, dati assai interessanti rispetto a tale tema provengono dalle ricerche condotte nell'ambito della acquisizione di linguaggi artificiali negli umani. In tale prospettiva si colloca, ad esempio, la ricerca realizzata dal gruppo di ricerca di Saffran - descritta in "Word Segmentation: The Role of Distributional Cues"<sup>171</sup>. Obiettivo di tale studio era di esplorare il valore degli indizi prosodici in quanto marcatori delle unità costitutive di un linguaggio artificiale. Nello specifico gli autori hanno costruito un linguaggio artificiale sulla base delle medesime regole di probabilità di transizione descritte in Saffran *et al.* 1996<sup>172</sup>, aggiungendo alle proprietà transizionali di tale mini linguaggio un ulteriore indizio volto a facilitarne la segmentazione: l'allungamento della durata di una - la prima o l'ultima - delle vocali delle parole (nonsense) trisillabiche. Dallo studio emerge che l'aggiunta di tale parametro prosodico favorisce significativamente la discriminazione delle parole ben formate all'interno del flusso linguistico nonsense. Così gli autori commentano:

Nine-month-old infants appeared initially to form units based on the rhythmic regularities in simple syllabic combinations and subsequently to analyze the distributional regularities. These data indicate that 9-month-old infants attempt to

---

<sup>171</sup> Saffran, J. R., Newport, E., Aslin, R. N., *Word segmentation: The role of distributional cues*, "Journal of Memory and Language", 35, 1996, pp. 606-621.

<sup>172</sup> Saffran, J. R., Aslin, R. N., & Newport, E. L., *Statistical learning by 8-month-old infants*, cit.

integrate prosodic and distributional cues when segmenting speech. More generally, it seems quite likely that language learners take advantage of any and all cues which they can discover in the language input. Correlated sources of information, extracted by appropriately biased learning mechanisms, should eventually converge on solutions to learning problems like the discovery of word boundaries.<sup>173</sup>

Possiamo affermare, dunque, che le proprietà statistiche e gli indizi prosodici (nella fattispecie, l'allungamento delle vocali) co-operano nel processo di segmentazione delle unità costitutive di un discorso privo di senso, quale può essere *prima facie* il flusso di suoni cui tanto gli uomini quanto le specie non umane sono esposti nei primi tempi di acquisizione di una lingua storico-naturale umana. È opportuno notare, a tal proposito, che i soggetti coinvolti negli esperimenti della ricerca in questione erano adulti umani. Per tale motivo, i risultati di tale studio aprono ulteriori domande concernenti il ruolo specifico degli indizi prosodici in relazione alle proprietà distribuzionali nell'effettivo processo di acquisizione ontogenetica di una lingua negli infanti pre-verbali. Cruciale, in tale direzione è lo studio condotto da Johnson e Jusczyk (2001), dal titolo "Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics"<sup>174</sup>, il cui obiettivo è quello di mostrare se gli indizi

---

<sup>173</sup> Saffran, J., *et al.*, *Word segmentation: The role of distributional cues*, cit.

<sup>174</sup> Johnson, E., Jusczyk, P.W., *Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics*, "Journal Of Memory And Language", 44, 2001, pp. 548-567. Su tale tema, cfr.: Thiessen, E. D., Saffran, J.R., *When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants*, "Developmental Psychology", 39, 2003, pp. 706-716; Shukla, M., Nespors, M., *An interaction between prosody and statistics in the segmentation of fluent speech*, "Cognitive psychology",

prosodici vantano un valore superiore alle proprietà statistiche del linguaggio artificiale nel determinarne la segmentazione delle unità costitutive. A tal fine, gli autori dello studio si sono basati sui medesimi materiali fonici utilizzati dal gruppo di Saffran<sup>175</sup>, aggiungendovi dei marcatori prosodici in conflitto con gli indizi legati alle proprietà transizionali. Siano date, ad esempio le unità ben formate *pabiku*, *tibudo*, *golatu*, *daropi*, e le “part words” *budopa*, *bikugo*, *kudaro*, etc. – costruite in modo da contenere le ultime due sillabe di una parola ben formata, e la prima di un’unità diversa. Come nell’esperimento condotto da Saffran, i soggetti vennero coinvolti in una fase di familiarizzazione che prevedeva l’esposizione ad un flusso vocale continuo ed ininterrotto per due minuti, che potrebbe essere rappresentato graficamente nel seguente modo: *tibudopabikudaropipabikugolatutibudopabiku...*

L’idea degli autori è quella di suggerire una strategia di segmentazione in conflitto con quella prevista dalle proprietà statistiche di distribuzione delle sillabe, ovvero accentuando prosodicamente le sillabe iniziali delle *part-words*. Così, nel flusso di familiarizzazione utilizzato in Saffran *et al.* (*tibudopabikudaropipabikugolatutibudopabiku...*) le prime sillabe delle *part-words* vennero sostituite da sillabe accentuate - ovvero

---

54, 2007, pp. 1-32; Johnson, E., Seidl A. H., *At 11 month, prosody still outranks statistics*, “Developmental Science”, 12, 2009, pp. 131-141.

<sup>175</sup> Saffran, J. R., *et al.*, *Statistical learning by 8-month-old infants*, cit.

modificate in frequenza, durata e amplitudine - generando un flusso ortograficamente rappresentabile come segue: *tibuDOpabikudaroPIpabikugolaTUtibudopabiku...*, in cui le sillabe accentuate sono quelle in maiuscolo. Seguendo la procedura della violazione delle aspettative, gli studiosi hanno ottenuto dei risultati estremamente importanti relativamente alle strategie di segmentazione delle parole negli infanti di 8 mesi di età. I dati provenienti da tale ricerca dimostrano infatti che, se esposti ad un flusso di discorso artificiale le cui unità costitutive sono segmentabili allo stesso tempo sulla base di proprietà statistiche di distribuzione delle sillabe, e di indizi prosodici in conflitto fra loro – gli infanti tendono a far capo all'accentuazione prosodica come marcatore primario delle parole ben formate:

In the present experiment, two competing cues to word segmentation were pitted against each other: stress and statistics. In segmenting the familiarization sequence, the infants relied more heavily on the stress cue to indicate word onsets than on the statistical cue relating to the transitional probabilities of successive syllables. The addition of stress as a conflicting segmentation cue in the familiarization sequence resulted in a reversal of looking time preferences from what was observed in the previous experiment. Thus, in the test phase of the present experiment, infants looked significantly longer toward the words than the part-words. [...] Consequently, it appears that although statistical cues are sufficient to segment a simple artificial language, 8-month-olds weigh speech cues such as stress more heavily.<sup>176</sup>

---

<sup>176</sup> Johnson, E., Jusczyk, P. W., *Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics*, cit., p. 558.

In ultima analisi, dunque, sulla base degli studi fin qui descritti possiamo affermare che nelle primissime fasi di acquisizione ontogenetica del linguaggio, i marcatori prosodici rivestono un ruolo dominante rispetto alle proprietà computabili del linguaggio. I neonati, infatti, sembrano seguire maggiormente le caratteristiche prosodiche del flusso discorsivo nel processo di segmentazione delle parole. Vale la pena, a questo punto, di prendere in esame alcune ricerche chiave in merito alla caratterizzazione prosodica dei proferimenti indirizzati proprio agli infanti pre-verbali (*Infant-Directed Speech*), indagandone la capacità di acuire la percezione dei tratti più salienti dello scambio comunicativo, ovvero, non soltanto la segmentazione delle parole, ma anche la dimensione emozionale, e l'indirizzamento dell'attenzione verso oggetti esterni.

### **8.3 La portata emozionale e cognitiva del *Motherese***

Al fine di esplorare la cardinalità del valore dei tratti prosodici delle lingue parlate, è opportuno rivolgere l'analisi al tipo di linguaggio in cui forse più che in ogni altro, le diverse sfumature prosodiche vengono particolarmente enfatizzate. Si tratta del motherese (detto anche *baby talk*, o *infant-directed speech* - da qui in avanti IDS) ovvero di quei proferimenti vocali dai marcati tratti prosodici, che gli adulti *spontaneamente*

indirizzano ai neonati, e che, d'altra parte questi ultimi preferiscono ai discorsi rivolti agli adulti<sup>177</sup>. Nel presente paragrafo analizzeremo gli specifici parametri acustici che interessano l'IDS e il relativo ampliamento degli effetti sia sul piano emozionale che cognitivo - non a caso, come vedremo, funzionali a facilitare l'accesso dei soggetti pre-verbali nell'universo linguistico degli umani.

Anne Fernald e Patricia Kuhl, due delle maggiori studiose in tale ambito di ricerca, hanno condotto uno studio rivolto proprio alla individuazione dei tratti acustici che determinano la preferenza degli infanti per il motherese. Come osservano nel relativo articolo, "Acoustic determinants of infant preference for motherese speech"<sup>178</sup>, in cui espongono i bambini a sequenze linguistiche modulate rispettivamente in durata, intensità e frequenza, i neonati mostrano una spiccata preferenza per quei discorsi la cui frequenza fondamentale ( $F_0$ ) veniva modulata in modo da presentare un'esagerazione intonazionale (figura 2):

The finding that  $F_0$ -contours derived from motherese speech are sufficient to elicit an auditory preference in infants suggests that it is the widely reported "higher pitch" of mothers' speech [...] is especially attractive to infants. However, the "higher pitch" commonly attributed to infant-directed speech usually refers to several different features of motherese intonation, including higher mean-  $F_0$ ,

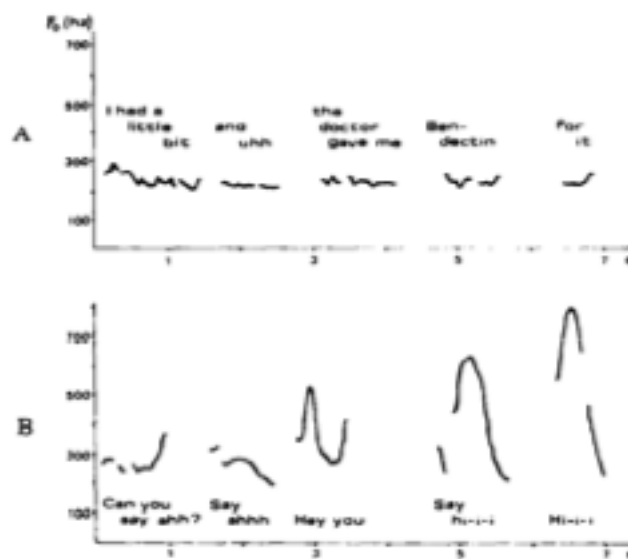
---

<sup>177</sup> Vedi Fernald, A., *Four-month-old infants prefer to listen to motherese*, "Infant Behavior and Development", 8, 1985, pp. 181-195.

<sup>178</sup> Fernald, A., *Acoustic determinants of infant preference for motherese speech*, "Infant Behavior and Development", 10, 1987, pp. 279-293. Cfr. Cooper, R. P., Abraham, J., Berman S., Stask, M., *The development of infants' preference for motherese*, "Infants Behavior and development", 20, 1997, pp. 477-488.



wider  $F_0$ -excursions, and expanded  $F_0$ - range [...]. While these  $F_0$ -characteristics of motherese do tend to co-occur in motherese, they are independent and could have distinct perceptual consequences for the infant. For this reason, it is important to differentiate between the possible contributions of pitch level and pitch movement to the heightened salience of motherese prosody. [...] *It therefore seems more plausible that it is the increase in  $F_0$ -modulation within an expanded range, rather than the higher absolute  $F_0$ -level, that makes motherese speech more interesting to infants than adult-directed speech.*<sup>179</sup>



**Figure 1.** Intonation contours from one of four pairs of natural speech samples used as auditory reinforcers by Fernald (1985): (A) Adult-directed speech; (B) Infant-directed speech, or motherese.

**Fig. 2.** Immagine modificata da Fernald *et al.* (1987). I discorsi rivolti agli infanti (B) mostrano una linea intonazionale dai contorni assai più esagerati rispetto ai discorsi scambiati fra adulti (A).

Sulla base di tale importante dato empirico, il *team* di ricerca guidato dalla stessa Anne Fernald si è in seguito interessato ad esplorare la presenza di simili tratti prosodici nei discorsi rivolti agli infanti in sei diverse lingue (inglese, francese, italiano, tedesco, giapponese, inglese

<sup>179</sup> *Ivi*, p. 290 (corsivo mio).

americano, e inglese britannico) e tanto dalle madri quanto dai padri - per dimostrarne il carattere cross-culturale, ovvero universale. Da tale studio emerge quanto segue:

Across languages, both mothers and fathers used higher mean- $f_0$ ,  $f_0$ -minimum, and  $f_0$ -maximum, greater  $f_0$ -variability, shorter utterances, and longer pauses in infant-directed speech than in adult-directed speech. Mothers, but not fathers, used a wider  $f_0$ -range in speech to infants. American English parents showed the most extreme prosodic modifications, differing from the other language groups in the extent of intonational exaggeration in speech to infants. These results reveal common patterns in caretaker's use of intonation across languages, which may function developmentally to regulate infant arousal and attention, to communicate affect, and to facilitate speech perception and language comprehension.<sup>180</sup>

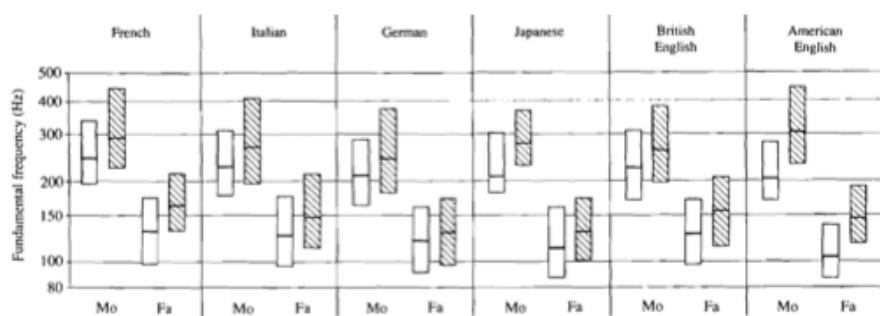


Fig. 3: Cross-language comparison of fundamental frequency ( $f_0$ ) characteristics of mothers' (Mo) and fathers' (Fa) speech to adults and to infants. For each bar, the bottom line represents the mean  $f_0$ -minimum, the top line represents the mean  $f_0$ -maximum, and the intersecting line indicates the mean- $f_0$  per utterance. The extent of the bar corresponds to the  $f_0$ -range. □, Adult-directed speech; ▨, infant-directed speech.

**Fig. 3.** Immagine modificata da Fernald *et al.* (1989). Rispetto ai discorsi diretti agli adulti (barre bianche), gli IDS (barre nere) presentano in sei diverse culture una gamma di frequenze più elevate, tanto da parte delle madri (Mo) che dei padri (Fa).

In altri termini, gli studiosi hanno scoperto che - rispetto alle conversazioni che hanno luogo fra adulti - il *baby talk* è caratterizzato da:

<sup>180</sup> Fernald, A., Taeschner, T., & Dunn, J. *A cross-language study of prosodic modifications in mothers and fathers' speech to preverbal infants*, "Journal of child language", 16, 1989, pp. 477-501.

a) una gamma di frequenze più elevate – e ciò è evidente specialmente nelle espressioni proferite dalle madri (figura 3); b) espressioni più brevi in lunghezza, ma ripetute più volte; c) pause più lunghe. Secondo gli autori, tali precipue caratteristiche sarebbero da associare ad un costitutivo ruolo cognitivo legato, nello specifico, alla modulazione del livello di arousal, assieme alla trasmissione di informazioni di natura primariamente emozionale ed affettiva. Tale effetto, unitamente all’efficacia dell’esagerazione dei contorni prosodici nel richiamare e mantenere l’attenzione degli infanti fa delle modificazioni prosodiche un ottimo candidato per la facilitazione della processazione delle costitutive unità foniche della lingua, dunque della sua acquisizione e comprensione. Nel seguito del paragrafo analizzeremo più da vicino gli studi rivolti ad approfondire tali significativi effetti cognitivi e comunicativi del *baby talk*.

In uno studio condotto nel 1989, “Intonation and communicative intent in mothers' speech to infants: Is the melody the message?”<sup>181</sup>, la stessa Fernald ha creato degli stimoli consistenti in un insieme di frasi tipicamente proferite in cinque diversi contesti d’uso, con i relativi intenti comunicativi: approvazione, proibizione, gioco, richiamo dell’attenzione, conforto. La studiosa ne ha filtrato le informazioni di natura segmentale,

---

<sup>181</sup> Fernald, A. *Intonation and communicative intent in mothers' speech to infants: Is the melody the message?*, “Child development”, 60, 1989, pp. 1497-1510.

lasciando evidente la linea intonazionale, di modo che, per intenderci, tali frasi suonavano come un discorso sentito attraverso un muro, di cui non si distinguono le parole, ma soltanto l'andamento prosodico. Obiettivo guida in tale studio era di confrontare l'accuratezza dei soggetti adulti nell'identificare l'intento comunicativo delle espressioni rivolte ai neonati con quella mostrata nell'identificare il messaggio delle stesse espressioni rivolte agli adulti - sulla base delle sole caratteristiche prosodiche (frequenza media, gamma di frequenza, durata dei proferimenti). Da tale esperimento è emerso che l'identificazione dell'intento comunicativo del parlante avveniva con maggiore precisione quando si trattava di espressioni rivolte agli infanti, piuttosto che a interlocutori adulti:

The melody carries the message in speech addressed to infants to a much greater extent than in speech addressed to adults. The major finding of this founding is that listeners using only prosodic cues were able to identify the communicative intent of the speaker with significantly higher accuracy in infant-directed speech than in adult directed speech. The interpretation of these results proposed here is that the relation of prosodic form to communicative function is made uniquely of salient in the melodies of mothers' speech to infants, and that these characteristic melodies are potentially accessible and informative to the preverbal infant.<sup>182</sup>

È opportuno sottolineare che sebbene ne risulti evidente la portata comunicativa della linea prosodica dei proferimenti, lo studio non rivela quali specifici tratti prosodici siano associati ai diversi intenti comunicativi.

---

<sup>182</sup> *Ivi*, p. 1505.

Una risposta significativa a tale quesito arriva da uno studio pubblicato nel 2001 dal Trainor *et al.*<sup>183</sup>. In esso gli autori, pur riconoscendone la portata informativa, sostengono che il *prius* significazionale del motherese sia di natura squisitamente emozionale, e volto primariamente ad instaurare ed alimentare il legame affettivo tra gli interlocutori. Per dimostrare la validità di tale ipotesi, gli autori hanno utilizzato, come stimolo, una sola frase espressa sia in motherese (“Hey, honey, come over here”) che rivolta ad un adulto – ed, in entrambi i casi, adattata a tre diversi scenari fortemente emozionali: paura, sorpresa, amore/conforto. Dopo aver appurato che la connotazione emozionale di tali proferimenti era ben identificata nei due diversi tipi di linguaggio, gli autori hanno analizzato nello specifico quattro parametri acustici distintivi delle diverse dimensioni emozionali: frequenza fondamentale, linea intonazionale, gamma di frequenze, tempo e contorno ritmico. Da tale analisi emerge che in linea generale, tutti i tratti acustici - tranne la frequenza fondamentale, più alta nel *baby talk* - non cambiano in base al tipo di linguaggio (motherese o indirizzato ad un adulto), bensì in funzione del tipo di emozione espressa. Così gli autori commentano tali dati:

---

<sup>183</sup> Trainor, L. J., Austin, C. M., Desjardins, R. N., *Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion?*, “Psychological science”, 11, 2000, pp. 188–195.

Why might the vocal expression of emotion be so pervasive in ID speech? First, because infants cannot understand the words, the prosody becomes vitally important for communication. Second, *much research indicates that strong emotional ties to caring adults are vital in the infancy period not only for emotional development, but also for physical and intellectual development [...].* [...] We conclude that emotional expression is very similar in ID and AD speech and that previously reported differences between ID and AD speech are in large part the result of comparing emotional ID speech with less emotional AD speech. In other words, ID prosody is not what is special. What is special is the widespread expression of emotion to infants in comparison to the more inhibited expression of emotion in typical adult interactions.<sup>184</sup>

Possiamo dunque affermare, con gli autori, che i tratti acustici distintivi del motherese trovano la loro origine nella vitale necessità di esprimere emozioni nello spazio interpersonale; così, la dimensione emozionale assume una posizione prioritaria sia in termini temporali, che per quanto concerne lo sviluppo delle facoltà cognitive negli infanti ancora privi della possibilità di accesso verbale alla comunicazione con i *caregiver*. Crediamo che tali dati, conciliati con i risultati della ricerca effettuata da Fernald, sottolineano l'importanza della portata informativa della comunicazione emozionale, pervasiva, come abbiamo visto, anche nelle vocalizzazioni dei primati non umani, e per questo, a buon diritto identificabile con uno dei maggiori tratti di continuità evolutiva.

È importante notare che le informazioni espresse dalla modulazione dei tratti prosodici, nel comunicare le stesse emozioni di cui si fanno carico

---

<sup>184</sup> *Ivi*, p. 194 (corsivo mio).

assumono un ruolo performativo. Crediamo, infatti, che Trainor *et al.* manchino di riconoscere un assioma importante della comunicazione, ovvero che ogni espressione emozionale (verbale o meno) vanti una portata significazionale legata al contesto e allo scopo intrinseco alla comunicazione stessa, ovvero alle sue conseguenze pragmatiche. Come nel caso degli stimoli adottati da Fernald, legati a contesti come la proibizione, o il conforto, allo stesso modo, la connotazione prosodica delle diverse emozioni utilizzate dal *team* di Trainor sono inevitabilmente altrettanto legati a contesti d'uso, come possono essere il richiamo dell'attenzione verso un nuovo oggetto o evento nel caso delle espressioni che indicano sorpresa, o la messa in stato di allerta, nel caso della connotazione emozionale legata alla paura. Vale la pena, dunque, di sottolineare l'alto impatto emozionale degli IDS, decisivo tanto nello sviluppo del legame affettivo tra gli interlocutori, essenziale per l'evoluzione delle facoltà cognitive del bambino e del suo equilibrio psichico<sup>185</sup>, quanto nell'acquisizione del linguaggio stesso, ovvero per la comprensione e la produzione dei significati proposizionali.

Un ulteriore vantaggio cognitivo della modulazione prosodica delle espressioni consiste nella sua capacità di richiamare e mantenere

---

<sup>185</sup> Sull'importanza del ruolo della dimensione relazionale ed emozionale nello sviluppo del sè, vedi Siegel, D. J., *The Developing Mind*, The Guilford Press Inc., 1999, trad. it. di L. Madeddu, *La mente relazionale. Neurobiologia dell'esperienza interpersonale*, Raffaello Cortina, Milano, 2001.

l'attenzione nel bambino sul focus del discorso, favorendone l'abilità di segmentare le parole all'interno del flusso linguistico, e infine, di riconoscerne i legami sintattici, e il loro pragmatico contesto d'uso intersoggettivamente condiviso.

A tal proposito, in uno studio dal titolo "Prosody and focus in speech to infants and adults"<sup>186</sup>, Fernald e Mazzei si sono specificamente chieste se l'esagerazione dei tratti prosodici tipici del motherese accresca la salienza degli indizi acustici relativamente alla introduzione di informazioni nuove su cui si vuole focalizzare l'attenzione dell'infante. Così ne descrivono i risultati:

Consistent with previous findings on the global prosodic characteristics of speech to children, the mothers in this sample used higher pitch and wider pitch range when addressing their infants than when addressing an adult [...].

When focusing on target words introduced as "new" information in a story told to preverbal infants, mothers consistently positioned these words at points of perceptual prominence in the speech stream. *Focused words were more likely to occur on F<sub>0</sub>peaks and in utterance-final position in infant-directed speech than in adult-directed speech, and these pitch peaks were significantly higher in fundamental frequency.*<sup>187</sup>

Così, dagli studi fin qui descritti emerge che il motherese sia in grado di canalizzare l'attenzione dell'infante verso determinati oggetti, eventi, o informazioni di natura emozionale. Resta da chiedersi, a questo punto, se

---

<sup>186</sup> Fernald, A., *Prosody and focus in speech to infants and adults*, "Developmental Psychology", 27, 1991, pp. 209-221.

<sup>187</sup> *Ivi*, pp. 212, 214 (corsivo mio).



tale tipo di linguaggio abbia degli effetti positivi sulla processazione del flusso del discorso e la formazione di connessioni tra le frasi e i rispettivi riferimenti oggettuali, o, in altri termini, sul piano della segmentazione delle parole all'interno del flusso linguistico cui gli infanti sono esposti, e sulla abilità di riconoscere e comprendere gli schemi sintattici che ne stabiliscono le interconnessioni. Per quanto concerne il primo punto, Thiessen ed i suoi collaboratori hanno condotto uno studio sperimentale dal significativo titolo "Infant-directed speech facilitates word segmentation"<sup>188</sup> in cui, basandosi sulla ricerca di Saffran *et al.* ("Statistical learning by 8-month old infants") hanno costruito un mini linguaggio artificiale in cui le unità costitutive presentavano specifiche regolarità statistiche il cui calcolo implicito ne permetteva la segmentazione. Obiettivo della ricerca era scoprire se le caratteristiche prosodiche del *baby talk*, nello specifico la frequenza fondamentale e il contorno intonazionale, accrescano la capacità degli infanti di segmentare correttamente il flusso fonico, ovvero di distinguere le parole dalle "non-parole". A tal fine, gli autori hanno esposto gli infanti a due mini linguaggi identici dal punto di vista strutturale e segmentale, ma differenti relativamente alla loro caratterizzazione prosodica: uno dei linguaggi seguiva infatti i parametri prosodici del motherese, mentre l'altro rispettava i parametri prosodici

---

<sup>188</sup> Thiessen, E., Hill, E., Saffran, J. R., *Infant-directed speech facilitates word segmentation*, "Infancy", 7, 2005, pp. 53-71.

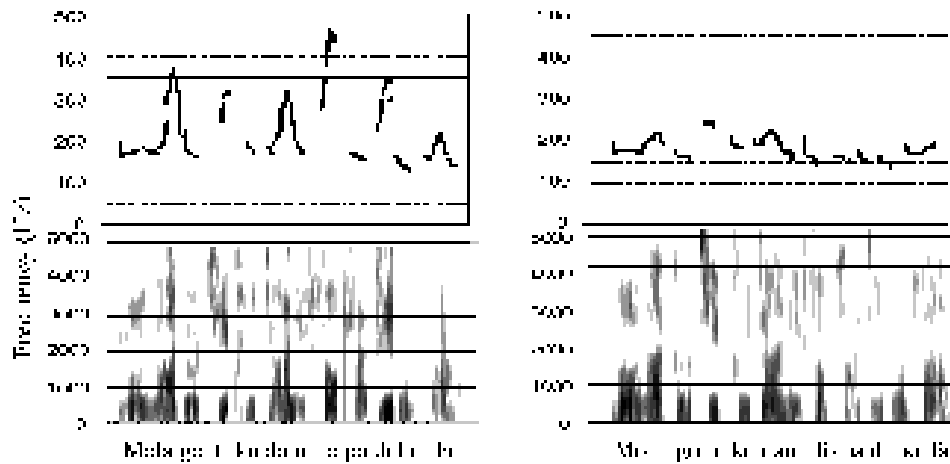
tipici dei discorsi rivolti agli adulti<sup>189</sup> (figura 4). Da tale studio emerge che, in effetti, la modulazione prosodica facilita la corretta segmentazione del flusso fonico da parte degli infanti preverbali, e ciò avviene a discapito della mancanza di specifiche marcazioni in corrispondenza dei confini delle parole, come ad esempio, l'accentuazione della prima sillaba delle parole (forma tipica nella lingua inglese), e sebbene i due linguaggi presentassero i medesimi pattern strutturali. Secondo gli autori, è possibile che tale risultato sia dovuto al fatto che l'innesto di una linea prosodica sulla struttura segmentale amplifichi lo stato attentivo degli infanti, mantenendone più vivo l'interesse verso il mini linguaggio in motherese, quindi verso la sua costitutiva struttura interna - o, in alternativa, è possibile che la modulazione prosodica abbia influito, più che sull'attenzione verso la struttura statisticamente regolata, sulla capacità di *memorizzare* le unità ben formate del flusso fonico. Così gli autori commentano i risultati del loro studio:

Each of these hypotheses is consistent with the current data. Therefore, these experiments do not distinguish between them. However, they do indicate that ID speech enables infants to more easily use statistical information than AD speech, even in situations where AD and ID speech contain the same statistical

---

<sup>189</sup> Vedi Ivi, p. 58: "In the AD speech, the speaker's average F0 was 230 Hz (range = 140–260 Hz). In the ID speech, the speaker's average F0 was 292 Hz (range = 140–480 Hz). The increased range in the ID speech was due to the speaker's exaggerated pitch peaks, which reached an average of 406 Hz. By contrast, pitch peaks in the AD speech reached an average of only 252 Hz. These values are consistent with previous reports on the F0 characteristics of ID speech. For example, Fernald (1989) found that the average F0 of ID speech was 311 Hz, with a range of 338 Hz."

structure.<sup>190</sup>



**Fig. 4.** Immagine modificata da Thiessen *et al.* (2005). Rappresentazione acustica di due frasi non-senso, identiche dal punto di vista fonetico e delle distribuzioni statisticamente regolate delle sillabe, ma modulate rispettivamente secondo i parametri prosodici del *motherese* (colonna sulla sinistra) e dei discorsi rivolti agli adulti (colonna sulla destra).

Una volta illustrato come gli aspetti prosodici possano facilitare il processo di segmentazione delle parole all'interno del flusso linguistico del *motherese*, è opportuno interrogarsi circa i meccanismi che interessano l'abilità di percepire e comprendere lo schema sintattico che ne intesse le connessioni. Tale questione teorica è il focus di ricerca del paradigma di ricerca che va sotto il nome di *bootstrapping* prosodico. Sarà questo l'oggetto del prossimo paragrafo.

<sup>190</sup> *Ivi*, p. 67.

## 8.4 L'ipotesi del *bootstrapping* prosodico

Infants acquiring language have to learn about the lexicon, the phonology, and the syntax of their native language. For each of these domains, being able to rely on knowledge from the other domains would simplify the learner's task. For instance, since syntactic structure spells out the relationships between words in a sentence, it makes sense to assume that infants need to have access to words and their meanings in order to learn about syntax. Conversely, learning about the meaning of words would be greatly facilitated if infants had access to some aspects of syntactic structure [...]. These potential circularities can partially be solved if infants can learn some aspects of the structure of their language through a surface analysis of the speech input they are exposed to.<sup>191</sup>

In tale passo, tratto da “Bootstrapping Lexical and Syntactic Acquisition”, Christophe ed i suoi collaboratori illustrano l'ipotesi che fa capo al problema teorico che va sotto il nome di “*bootstrapping*”<sup>192</sup>. Questo consiste nello spiegare come gli infanti possano acquisire padronanza della dimensione sintattica a partire dalla sfera lessicale - e viceversa, comprendere il significato delle parole a partire dal pattern sintattico che li intesse. Tale circolarità - da cui deriva il nome stesso della teoria - secondo gli autori può essere risolta se si guarda agli aspetti strutturali degli stimoli linguistici cui gli infanti sono ordinariamente

---

<sup>191</sup> Christophe, A., Millotte, S., Bernal, S., Lidz, J., *Bootstrapping Lexical and Syntactic Acquisition*, “Language and Speech”, 51, 2008, p. 62. Sulla tesi del *bootstrapping* fonologico, cfr. Christophe, A., Guasti, T., Nespore, M., Dupoux, E., Van Ooyen, B., *Reflections on phonological bootstrapping: Its role in lexical and syntactic acquisition*, “Language and Cognitive Processes”, 12, 1997, pp. 585-612; Fernald, A., McRoberts, G., *Prosodic bootstrapping: A critical analysis of the argument and the evidence*, in J.L. Morgan, K. Demuth, “Signal to Syntax: from speech to grammar in early Acquisition”, Lawrence Erlbaum Associates, NJ, 1996, pp. 365-387.

<sup>192</sup> “*Bootstrap*”, letteralmente “tirante di stivale”, si riferisce ad una frase idiomatica americana la cui traduzione è “reggersi ai tiranti dei propri stivali”, indica un processo auto-sostenuto, il cui sviluppo cioè, si regge su se stesso, senza l'ausilio di un meccanismo esterno.

esposti<sup>193</sup>. A tal proposito, gli studiosi distinguono tre livelli (non isolati fra loro) di accesso al linguaggio verbale; ad un primo approccio, gli infanti avrebbero accesso ad una implicita rappresentazione fonologica prelessicale del flusso acustico che contiene delle informazioni sugli schemi prosodici che lo interessano. Ciò vuol dire che nell'ambito di una proposizione è possibile individuare i sintagmi costitutivi attraverso tre importanti parametri: a) l'allungamento della durata della sua parte finale b) l'accentuazione prosodica del suo inizio e c) la discontinuità di frequenza fra due sintagmi contigui – che indicano rispettivamente la fine, l'inizio e in generale, il confine – degli stessi. Così sarebbe la demarcazione intonazionale all'interno del flusso fonico a delineare *ipso facto* i confini delle strutture sintagmatiche, come avviene ad esempio nella frase “the little boy is running fast”, suddivisibile per mezzo dei suddetti parametri prosodici in [“the little boy”] e [“is running fast”].

Quest'ultimo punto trae evidenza da due ulteriori studi che gli autori hanno effettuato tanto sugli adulti quanto sui neonati, in cui hanno dimostrato che i soggetti tendono a non identificare, ad esempio, una parola come “*paper*”, se compare in una frase come [the butler with highest **pay**] [**performs** the most]. In quest'esempio, infatti, la separazione fra le due

---

<sup>193</sup> Ovvero di parole che rientrano in categorie grammaticali quali: articoli, preposizione, ausiliari, congiunzioni, pronomi, particelle grammaticali, o morfemi. Si tratta di termini che non hanno senso compiuto in sè, ma che contribuiscono ad esprimere e completare le relazioni grammaticali tra i lessemi.

distinte parti fonologiche entro parentesi, marcata prosodicamente, cade a cavallo della stessa parola, impedendone la identificazione<sup>194</sup>. Ciò mostra come, in effetti, i confini marcati dalla modulazione prosodica della frase non cadano in punti casuali, nè sono volti primariamente alla pertinentizzazione delle unità lessemiche all'interno della frase, ma rispecchiano piuttosto lo schema sintattico in funzione del quale è possibile isolare i sintagmi. In tale scia si colloca coerentemente la ricerca sperimentale effettuata su soggetti adulti da Millotte *et al.*, “Phrasal prosody disambiguates syntax”<sup>195</sup>, in cui gli autori, focalizzando il fuoco

---

<sup>194</sup> Su tale scia ben si collocano i dati che emergono da uno studio realizzato da Trainor, L. J., Desjardin R. N., *Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels*, “Psychonomic Bulletin & Review”, 2002, 9, pp. 335-340 - in cui gli autori indagano circa gli specifici effetti legati rispettivamente alla maggiore altezza della frequenza fondamentale e alla esagerata modulazione del contorno prosodico. Da tale studio emerge che mentre entrambi i parametri hanno degli effetti significativi per quanto concerne la sfera emozionale condivisa tra i *caregiver* e l'infante (con le relative conseguenze, come abbiamo visto sul piano dell'attenzione/memoria) - la modulazione della frequenza facilita al bambino il riconoscimento delle categorie vocaliche, apportando un ulteriore vantaggio nella operazione di percezione e segmentazione del flusso fonico: “[...] High pitch is universally used in ID speech even though, as we have shown, it impairs vowel discrimination. [...] Such evidence is consistent with the idea that the vocal expression of emotion is phylogenetically older than the use of language. Such a conclusion, of course, in no way diminishes the importance of ID features that promote vowel learning but, rather, highlights the complex interplay of features and functions involved in ID speech. Of primary significance is the conclusion that ID pitch contours promote vowel discrimination, as well as serve an emotional communication function.”

Alla luce di tali importanti scoperte, rimane tuttavia aperta la domanda in merito alle specifiche conseguenze - sul piano della acquisizione del linguaggio - della capacità di discriminare con maggiore nitidezza fonica le vocali piuttosto che le consonanti. Alla luce dei dati descritti in Newport, E. L., *et al.*, *Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate*, cit. - potremmo speculativamente affermare che, nell'ambito del linguaggio naturale, tale capacità vanta degli impliciti effetti cognitivi sulla generale facoltà di discriminare pattern sintattici fra le parole. In effetti, le vocali rappresentano, in qualche modo, il luogo principale in cui si innesta la modulazione sovrasegmentale, ovvero la modulazione prosodica in grado di orientare la percezione degli schemi sintattici. Cfr. a tal proposito Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Ulla Sundberg, U., Lacerda, F., *Cross-Language Analysis of Phonetic Units in Language Addressed to Infants*, “Science”, 277, 1997, pp. 684-686, in cui l'autrice dimostra che lo spazio vocalico nei *baby talk* di tre diverse lingue (svedese, russo e americano) viene esagerata, con conseguenti effetti positivi sulla possibilità degli infanti di discriminare le categorie fonetiche cui appartengono le vocali della propria lingua.

<sup>195</sup> Millotte, S., Wales, R., Christophe, A., *Phrasal prosody disambiguates syntax*, “Language and Cognitive Processes”, 22, 2007, pp. 898-909. Su tale argomento, cfr. Johnson, E. K., *Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech*, “The Journal of the

della ricerca sui parametri prosodici hanno mostrato come questi giochino un ruolo determinante nella disambiguazione sintattica di una frase. In particolare, gli studiosi si sono serviti di frasi francesi in cui compaiono unità omofoniche con valori semantici, e, di conseguenza, con ruoli sintattici diversi. Ne sono un esempio le seguenti frasi, in cui i termini omofonici vengono enfatizzati in maiuscolo: [Ce président][ILLUSTRE la réussite sociale et professionnelle.] / [Ce présidentILLUSTRE][littrès peu de romans contemporains.] – oppure [On sait que][les hommes][PURENT s'échapper de ce camp de concentration .]] / [On saitque][les hommes PURS][sont très rares dans le milieu de la politique.]].

Nella prima fase dello studio, i ricercatori hanno registrato la lettura a voce alta (da parte di alcuni soggetti) di 22 frasi che, come quelle dell'esempio appena citato, presentavano termini ambigui, ovvero verbi o aggettivi omofoni. Dopo aver appurato quantitativamente la presenza di specifiche variazioni prosodiche entro i diversi sintagmi che occorreano nelle frasi (come l'allungamento della durata della parte finale o la modulazione in crescendo dei termini omofonici e delle unità ad essi adiacenti), gli autori hanno esposto auditivamente un nuovo insieme di soggetti alle stesse frasi registrate, artificialmente interrotte proprio nel punto successivo alla parola

---

Acoustical Society of America”, 123, 2008, pp. EL144–EL148; Soderstrom, M., Seidl, A., Nelson, D. G. K., Jusczyk, D. W., *The prosodic bootstrapping of phrases: Evidence from prelinguistic infants*, “Journal of Memory and Language”, 49, 2003, pp. 249–267.

ambigua, per poi chiedergli di completarle scrivendo su un foglio di carta.

Ne è risultato che le variazioni prosodiche all'interno del sintagma fonologico in cui figurano le parole omofoniche permettevano ai soggetti di riconoscerne lo schema sintattico, e di conseguenza fornire una continuazione della proposizione coerente con la funzione sintattica attribuita alla parola omofona. Così i ricercatori commentano tale evidenza empirica:

French listeners were thus able to correctly interpret two sentence beginnings that only differed in their syntactic and prosodic structures. They assigned different syntactic categories to the ambiguous words, depending on their prosodic context only: they gave more adjective than verb responses to adjective sentences, and more verb than adjective responses to verb sentences.

[...]

Our results suggest that phonological phrase boundaries guide syntactic analysis in French adults. Thus, it seems reasonable to assume that these prosodic units may help infants to perform a syntactic analysis of spoken sentences, and thus facilitate lexical and syntactic acquisition [...].<sup>196</sup>

Per ricapitolare, dunque, la processazione legata *in primis* alla struttura prosodica della proposizione, rappresenterebbe, secondo la prospettiva di Christophe *et al.*, un primo approccio al flusso sonoro da parte degli infanti<sup>197</sup>. Un secondo importante livello di analisi è determinato

---

<sup>196</sup> *Ivi*, pp. 905, 906-907. Cfr. Johnson, E. K., *Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech*, cit.

<sup>197</sup> È bene sottolineare, per amor di chiarezza, che il processamento sintattico di cui parlano gli autori avviene ad un livello implicito, ovvero della mera comprensione da parte degli infanti del significato dei termini (che consiste, come abbiamo visto, nel riconoscimento delle corrispettive connessioni cosali), dimensione in cui si innesta l'identificazione della stessa relazione logica fra le parti.



dall'uso delle "parole funzionali" - particelle del discorso spesso molto brevi, eppure dal ruolo essenziale all'interno del discorso. Klammer *et al.* ne forniscono una chiara descrizione nel seguente passo:

Consider the following sentences (1) and (2):

(1) The *winfy prunkilmonger* from the *glidgement mominkled* and *brangified* all his *levensers vederously*.

(2) *Glop* angry investigator *larm* *blonk* government harassed *gerfritz* infuriated *sutbor pumrog* listeners thoroughly.

In sentence (1) above, the content words have been changed into nonsense syllables but it is not difficult for one to posit that *winfy* is an adjective, *prunkilmonger*, *glidgement*, *levensers* as nouns, *mominkled*, *brangified* as verbs and *vederously* as an adverb based on clues like the derivational and inflectional morphemes. (The clue is in the suffixes: -y indicates adjectives such as "wintery"; -er, -ment and -ers indicates nouns such as "baker", "battlement" and "messengers"; -led and -fied suggests verbs such as "mingled" and "clarified"; and -ly is that of adverbs such as "vigorously"). Hence, even without lexical meaning, the sentence can be said to be rather "meaningful". However, when the reverse is done and the function words are being changed to nonsense syllables as in sentence (2), the result is a totally incomprehensible sentence as the grammatical meaning which is signaled by the structure words is not present. Hence, function words provide the grammatical relationships between the open class words and helps create meaning in sentences.<sup>198</sup>

Così, le parole funzionali, grazie alla alta frequenza distribuitiva, alla loro caratteristica brevità (spesso consistono infatti di monosillabe), e alla loro posizione nella frase (spesso si collocano infatti alle estemità di una frase, o di un termine) – ben si prestano ad una pronta estrazione

---

<sup>198</sup> Klammer, T., Schulz, M: R., Della Volpe, A., *Analyzing English Grammar*, Longman, 2009, in [http://en.wikipedia.org/wiki/Function\\_word#cite\\_ref-0](http://en.wikipedia.org/wiki/Function_word#cite_ref-0)

statisticamente regolata da parte degli infanti, facilitando allo stesso tempo l'operazione di identificazione a) dei vocaboli nuovi ad esse contigui, e b) del loro valore sintattico, così da determinare una comprensione del senso della frase, seppur vaga ed evocativa.

Per quanto riguarda il primo punto, ad esempio, da un esperimento condotto da Shi e Lepage<sup>199</sup> risulta che i bambini di otto mesi, se esposti a parole che occorrono con bassa frequenza nei discorsi ad essi rivolti come “prova” o “sella”, mostrano un maggiore effetto di familiarità verso di esse (indicato dalla quantità di tempo in cui guardavano verso l'altoparlante da cui proveniva lo stimolo acustico), se sono precedute da un articolo familiare piuttosto che da una parola nonsense. Ciò vuol dire che la presenza di una parola funzionale (in questo caso, di un articolo) facilita il processo di riconoscimento di un lessema, e nello specifico, di una parola di cui non si ha ancora una piena padronanza.

Tuttavia, oltre che alla identificazione di un nome poco comune, le parole funzionali sono utili alla identificazione (implicita) della categoria sintattica di appartenenza del termine. Ad esempio, come risulta dalla ricerca effettuata da Kedar *et al.*, dal titolo: “Getting there faster: 18- and 24-month-old infants’ use of function word to determine reference”<sup>200</sup>, i

---

<sup>199</sup> Shi, R., Lepage, M., *The effect of functional morphemes on word segmentation in preverbal infants*, “Developmental Science”, 11, 2008, pp. 407-413.

<sup>200</sup> Kedar, Y., Casasola, M., Lust, B., *Getting there faster: 18- and 24-month-old infants’ use of function word to determine reference*, “Child Development”, 77, 2006, pp. 325-338.

bambini sono in grado di distinguere un verbo da un nome in base alla presenza di una appropriata e contigua parola funzionale: un articolo nel caso dei nomi, e un pronome nel caso dei verbi. Così, la processazione del flusso linguistico incanalato dalla modulazione prosodica si interseca con una prima categorizzazione sintattica delle unità semantiche all'interno della proposizione, resa possibile dalla presenza di parole funzionali come articoli, pronomi o morfemi. La sincronizzazione di tali dimensioni sfocia nel terzo livello del modello proposto da Christophe *et al.*:

The model puts forward the hypothesis that infants may compute a preliminary syntactic structure by relying both on prosodic boundaries and function words: prosodic boundaries would give syntactic constituent boundaries, while function words would allow infants to label these constituents. To spell out the example from the model, in the sentence [The little boy] [is running fast], brackets are given by phrasal prosody. The first unit would be identified as a Noun Phrase because it starts with the determiner *the*, while the second unit would be identified as a Verb Phrase because it starts with the auxiliary *is*. Infants might thus hear this sentence as [The XXX]<sub>NP</sub>[is XXX]<sub>VP</sub>, where brackets are given by prosody, and labels by the function words *the* and *is*.<sup>201</sup>

In altri termini, la modulazione prosodica consentirebbe di suddividere la frase nei suoi sintagmi costitutivi, ed entro questi ultimi, la presenza di una parola funzionale permetterebbe di riconoscere la categoria logico-sintattica dei termini ad esso contigui, e di conseguenza, di attribuire

---

<sup>201</sup> Christophe, A., Millotte, S., Bernal, S., Lidz, J., *Bootstrapping Lexical and Syntactic Acquisition*, cit., p. 70.

al sintagma la precippua portata sintattica. A tale ipotesi, uno studio condotto dallo stesso *team* di ricerca di Anne Christophe, attribuisce una buona evidenza empirica, descritta in “Can prosodic cues and function words guide syntactic processing and acquisition”<sup>202</sup>. Tale studio presenta due diverse condizioni sperimentali, volte ad esplorare il ruolo delle parole funzionali e della prosodia nel facilitare il riconoscimento della categoria sintattica cui appartengono i termini contigui o prossimi alla stessa unità grammaticale. La peculiarità di tale studio consiste nell’aver utilizzato frasi in cui tutte le parole - tranne le particelle funzionali - vennero sostituite da termini inventati, ma che preservavano le medesime informazioni prosodiche. Vale la pena di riportare la prima condizione, chiamata “con particella funzionale”, che viene così descritta dagli autori:

a “with function word” condition: target words were directly preceded by a function word, as in “[une **bamoule**] [dri se froliter] [dagou]” (brackets indicate phonological phrase boundaries – the target word “bamoul” is a noun and it is preceded by an article – this sentence could be traduced in French by the following sentence, “[une **expo**] [doit se dérouler] [demain]” / “[an **exhibition**] [will take place] [tomorrow]”).

On the the other hand, we created a verb sentence, such as “[tu **bamoules**][saman ti] [à mon ada]” (where the verb target “bamoules” is preceded by a pronoun – its French equivalent could be “[tu **travaille**] [souvent trop] [à mon avis]” / “[you often **work**] [too hard] [in my opinion]”).<sup>203</sup>

---

<sup>202</sup> Wales, R., Dupoux, E., Christophe, A., *Can prosodic cues and function words guide syntactic processing and acquisition*, in R. Hofmann, H. Mixdorff (ed.), “Speech prosody: 3rd International Conference”, Dresden TUD Press, 2006.

<sup>203</sup> *Ivi*, p. 2.

L'idea, dunque, è quella di testare la capacità dei soggetti (adulti) di identificare i termini evidenziati in grassetto come nomi o verbi, sulla sola base della presenza di un articolo o di un pronome, ed in funzione delle informazioni prosodiche. Le risposte dei soggetti confermano l'ipotesi avanzata dagli studiosi: a discapito della mancanza di contenuti semantici specifici, essi erano infatti in grado di identificare correttamente il ruolo sintattico dei termini *target*.

Alla luce di tali importanti dati sperimentali, troviamo interessante, considerare le poesie “metasemantiche” di Fosco Maraini<sup>204</sup>, che affidano ai suoni di termini privi di senso, e alla loro combinazione il compito di caricare i significanti di un ruolo specifico all'interno della frase, che ne rende, in qualche maniera fruibile il senso. Una poesia di cui vale la pena riportare qualche verso è *Il Lonfo*: «Il Lonfo non vaterca né gluisce / e molto raramente barigatta, / ma quando soffia il bego a bisce bisce / sdilenca un oco e gnagio s'archipatta». È importante osservare il ruolo giocato dalla linea prosodica, che contribuisce alla determinazione dello schema sintattico, unitamente alla evocazione di specifiche informazioni di natura squisitamente emozionale.

---

<sup>204</sup> Maraini, F., *Gnosi delle fanfole*, Baldini e Castoldi, 1994.

## 8.5 Prosodia e contesti cross-situazionali

In ultimo, vorrei completare il presente capitolo tematizzando un aspetto fondamentale della acquisizione ontogenetica del linguaggio, la cui analisi sarà funzionale al guadagno di una prospettiva esaustiva sulle dinamiche evolutive della facoltà di linguaggio, intesa nel senso ampio che comprende tanto l'abilità di produzione che quella di comprensione delle articolazioni linguistiche vocali. Si tratta della capacità dei bambini pre-verbali di associare le unità lessemiche - individuate all'interno del flusso fonico del discorso ad essi rivolti - ad oggetti o eventi del contesto situazionale in cui si trovano. Cercheremo così di illustrare le prime fasi di acquisizione della capacità di comprendere un linguaggio, alla luce dei dati provenienti dalle ricerche sperimentali in tale ambito tematico.

Una questione in qualche modo connessa a tale tipo di analisi viene tematizzata nell'ambito del problema della "indeterminatezza della traduzione" in Quine<sup>205</sup>, il quale immagina la situazione in cui un linguista osserva il parlante nativo di una tribù indigena dire "gavagai" al passare di un coniglio in un campo vicino, ma non capisce quale aspetto del contesto (il coniglio, il campo, lo stesso correre del coniglio, il campo, l'esortazione

---

<sup>205</sup> Quine, W.V.O., *Word and Object*, Mit Press, Cambridge, 1960, pp. 26-45.

ad andare a caccia, etc.) la parola utilizzata dall'indigeno, assieme all'atto ostensivo che la accompagnava, vogliono *significare*. Come osserva il filosofo, l'indeterminatezza di una simile esperienza può venire progressivamente dissolta - anche se mai in maniera assolutamente definitiva - attraverso l'esperienza delle co-occorrenze del dato termine in diversi contesti situazionali, cosicché la varietà di questi ultimi ne possa evidenziare l'aspetto ad essi comune, ovvero il significato di cui il termine, utilizzato cross-contestualmente, si fa portatore.

Proviamo adesso a sostituire il personaggio del linguista in terra estranea con un infante il quale, esposto ad una serie di espressioni linguistiche (ciascuna contestualizzata in situazioni più o meno differenti) proferite dall'indigena del nostro caso - ovvero dalla madre - deve risalire all'oggetto o al fatto specifico cui questa si sta riferendo. Alla luce degli studi fin qui presi in esame, possiamo osservare che una delle operazioni fondamentali implicate in tale complicato, eppure spontaneo processo è la segmentazione delle parole all'interno del discorso. Allora, la prima domanda che è opportuno porsi è se le parti del flusso fonico a cui i bambini sono in grado di associare un significato siano effettivamente le unità lessicali - o meglio, la sequenza di suoni che discriminano in quanto tali - o non, piuttosto, a qualunque stringa di suoni che occorre all'interno del flusso vocale cui sono esposti, anche quelle i cui limiti non coincidono

con quelle delle parole ben formate. A tale quesito il gruppo di ricerca di Jenny Saffran ha dedicato uno studio dal titolo “Can Infants Map Meaning to Newly Segmented Words?”<sup>206</sup>. Riprendendo ancora una volta il paradigma adottato nello studio del 1996 sull’apprendimento statistico negli infanti di otto mesi, i ricercatori hanno ipotizzato che la strutturazione statisticamente regolata degli stimoli, dunque la presenza di unità ben formate all’interno del flusso fonico, faciliti ai bambini il compito di associare un significato alla data parola. Per dimostrare tale ipotesi, dopo aver familiarizzato gli infanti ad un linguaggio artificialmente costruito ne hanno indagato l’abilità di riconoscere l’associazione tra parole ben formate, non-parole (stringhe di suoni non facenti parte del minilinguaggio con cui avevano familiarizzato) e “parole parziali”, le cui sillabe occorreano nel minilinguaggi - con oggetti raffigurati su un’immagine. Utilizzando la tecnica della violazione delle aspettative, quindi della durata della fissazione verso immagini/suoni non familiari – gli autori hanno dimostrato che gli infanti preverbali mostravano una maggiore familiarità verso oggetti associati alle unità ben formate piuttosto che alle non-parole o alle “parole parziali”. Il fatto che gli infanti mostrassero minore familiarità con gli oggetti associati a queste ultime conduce gli autori alle seguenti

---

<sup>206</sup> Estes, K. G., Evans, J. L., Alibali, M. W., Saffran, J. R., *Can Infants Map Meaning to Newly Segmented Words? Statistical Segmentation and Word Learning?*, “Psychological science”, 18, 2007, pp. 254–260.



conclusioni:

Our results demonstrating that infants can use statistically segmented word forms as object labels support the claim that statistically segmented sound sequences are actually wordlike. A key aspect of these findings is that experience with sound sequences affects label learning, but not because of mere exposure and resulting familiarity. Instead, the internal structure of the sound sequences is crucial. Experiment 2 demonstrated that when two types of sound sequences occur with equal frequency in a speech stream, only sequences with strong internal structure are subsequently learned as object labels. These findings indicate that simply hearing sounds together in sequence is not enough; *the predictiveness of the sound sequence* is key for constituting a “good word.”<sup>207</sup>

Troviamo che tale studio, seppur assai significativo, manchi di render conto di un aspetto importante relativo ai contesti *reali* in cui i bambini sono esposti alla lingua materna, contesti altamente ambigui, in cui i singoli oggetti non sono isolati in immagini presentate su uno schermo, ma si presentano contestualizzati in ambienti estremamente ricchi di stimoli visivi e acustici. Dunque, l'operazione di capire il significato di un termine implica sì, da una parte, l'abilità di discriminare i termini ben formati all'interno del flusso del discorso, ma anche - dall'altra - l'abilità di discriminare, nell'ambito della complessa e variegata scena reale di cui il bambino è parte quegli oggetti/fatti cui il termine si riferisce. Si tratta, evidentemente di un'operazione non facile (seppure altamente spontanea

---

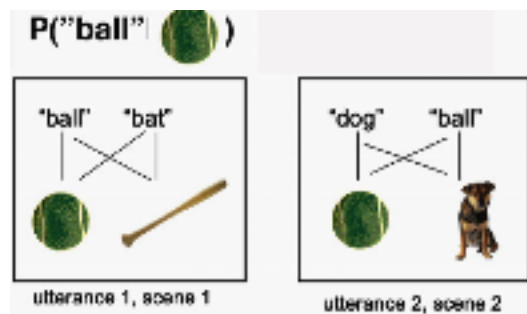
<sup>207</sup> *Ivi*, p. 259 (corsivo mio).

negli infanti preverbali), che fa capo all'abilità chiave di discriminare regolarità, e in particolare, ancora una volta, regolarità statisticamente calcolabili che interessano i diversi contesti in cui il dato termine occorre. Su tale scia sperimentale si colloca la ricerca effettuata da Smith e Yu, descritta in "Infants rapidly learn word-referent mappings via cross-situational statistics"<sup>208</sup>. I due studiosi presentano l'esempio raffigurato in figura 5, in cui un termine, "ball", viene proferito in due situazioni diverse, in cui occorrono diversi potenziali riferimenti, come ad esempio, una palla ed un cane nella prima scena, e una palla ed una mazza da baseball nell'altra. L'idea di base, evidente anche in questo esempio ultrasemplificato è che all'udire il termine "ball" in un solo contesto in cui figurano tanto la mazza quanto la palla, i bambini possono non essere in grado di distinguerne il riferimento, ma nel momento in cui rincontrano tale termine in un nuovo contesto, in cui la palla occorre ancora, ma stavolta accanto al cane, l'infante sarà nella condizione di associare la co-occorrenza del medesimo oggetto con la co-occorrenza della medesima sequenza di suoni attraverso scene contestuali diverse. In altri termini, gli infanti preverbali sono in grado di applicare una sorta di apprendimento

---

<sup>208</sup> Smith, L., Chen, Y., *Infants rapidly learn word-referent mappings via cross-situational statistics*, "Cognition", 106, 2008, pp. 1558-1568. Su tale argomento, cfr. Medinaa, T. N., Snedeker, J., Trueswella, J. C., & Gleitman, L. R., *How words can and cannot be learned by observation*, "Proceedings of the National Academy of Sciences", 108, 2011, pp. 9015-9019; Yu, C., Smith, L. B. *Rapid word learning under uncertainty via cross-situational statistics*, "Psychological science", 18, 2007, pp. 414-420.

statistico cross-situazionale che li mette nella condizione di imparare il riferimento delle parole.



**Fig. 5.** Immagine modificata da Smith *et al.* (2008). La co-occorrenza di un suono e il referente ad esso associato in scene situazionali diverse permette ai bambini di acquisire il significato delle parole.

Chiaramente, nelle situazioni reali le associazioni da stabilire si basano su statistiche molto più difficili, a causa dell'elevato numero di potenziali riferimenti presenti nelle scene di vita, e del numero nettamente maggiore di termini che i *caregivers* rivolgono ai bambini. Sulla base di tali considerazioni, al fine di sondare l'abilità di apprendimento statistico cross-contestuale gli studiosi hanno realizzato un esperimento, i cui stimoli consistevano in sei diversi oggetti - nello specifico, sei diverse figure con forme geometriche, e sei parole completamente inventate ad essi univocamente associate. Tali oggetti venivano presentati agli infanti due per volta in maniera random, in trenta diapositive, ciascuna delle quali prevedeva la realizzazione acustica delle parole corrispondenti alle

immagini viste. Tale tipo di design è una rappresentazione non troppo lontana dell'ambiguità che caratterizza i vari contesti situazionali dei proferimenti rivolti agli infanti, e il dato interessante che ne emerge è che nonostante la limitata esposizione agli stimoli, gli infanti hanno esibito familiarità con quattro parole su sei, mostrando, come commentano gli stessi autori - di possedere una spiccata abilità di discriminare le regolarità associative trasversalmente ai diversi contesti d'uso:

In sum, these results tell us that cross-situational statistical learning is in the repertoire of young word learners. Despite the ambiguity of word-referent mappings on any individual training trial, infants clearly accumulate information across trials and use that information to determine the underlying mappings. In less than four minutes, with six different word forms and six different objects, infants learned enough to systematically look longer at the objects more strongly associated with the forms than those more weakly associated.<sup>209</sup>

Per concludere, vale la pena di chiedersi se anche la modulazione prosodica (essenziale, come abbiamo visto nel paragrafo precedente, a richiamare l'attenzione dell'infante sulla linea del discorso che gli si rivolge, e sulla presenza di informazioni nuove), non ricopra un ruolo nel processo di associazione tra parole e riferimenti cross-contestualmente. Da uno studio condotto da Shukla *et al.*, dal titolo "Prosody guides the rapid

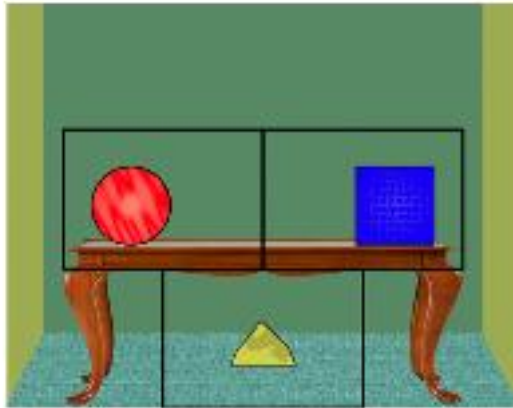
---

<sup>209</sup> Ivi, pp. 1564-1565.

mapping of auditory word forms onto visual objects in 6-mo-old infants<sup>210</sup>, risulta che in effetti essa influenzi positivamente tale operazione cognitiva già negli infanti di sei mesi di età. In particolare, gli autori hanno utilizzato una sequenza di suoni non-senso, dalla struttura x**A**Byz, dove AB (nella fattispecie costituita dalle sillabe *mu-ra*) rappresenta la parola *target* dell'esperimento. Lo stimolo uditivo cui i bambini sono stati esposti consisteva in un flusso fonico costruito in modo che a) la stringa AB si ripetesse più volte nel flusso fonico (rappresentando dunque la "parola" più frequente), mentre le altre sillabe *x*, *y*, *z*, variavano foneticamente, b) la sillaba B seguiva sempre A con probabilità 1 (ovvero, 100%), mentre la probabilità di transizione fra tutte le altre sillabe era inferiore. Una volta esposti a tale sequenza di suoni, i soggetti accedevano alla fase del test, il quale presentava due condizioni: una in cui la modulazione prosodica, ovvero la durata delle sillabe e la modulazione della frequenza confluiva nella stessa direzione dello schema strutturale ([x**A**] [yz]), ed una seconda condizione in cui andava in conflitto con essa ([x**A**] [**B**yz]). In entrambe le condizioni, oltre ad udire le sequenze di suoni, i bambini erano esposti ad una scena visiva animata, consistente in tre forme geometriche, di cui una (associata alla parola *target mu-ra*) si ingrandiva e tremava (figura 6).

---

<sup>210</sup> Shukla, M., White, K. S., Aslin, R. N., *Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants*, "Proceedings of the National Academy of Sciences", 108, 2011, pp. 6038–6043.



**Fig. 6.** Immagine modificata da Shukla *et al.* (2011).

1. CONDIZIONE A (prosodia e struttura statistica del flusso di suoni convergono verso la segmentazione di “mura” come parola ben formata): gli infanti associano la parola *target* al cerchio, il quale nella scena visiva cui venivano esposti si ingrandiva e tremava.
2. CONDIZIONE B (prosodia e struttura statistica divergono, suggerendo rispettivamente “mu” e “ra” come sillabe appartenenti a parole diverse, e come sillabe di un’unica parola ben formata): gli infanti associano la parola *target* agli elementi statici della scena visiva.

Contestualmente, *l’eye tracker*, un macchinario in grado di tracciare i loro movimenti oculari ne registrava la direzione dello sguardo all’udire delle parole del test, dunque l’oggetto-focus associato alla data unità sonora, nella fattispecie il cerchio rosso. Da tale esperimento è emerso che i bambini esposti alla condizione in cui prosodia e regolarità statistiche seguivano lo stesso pattern strutturale tendevano ad associare con successo la parola *target* al riferimento ad essa associato, ovvero al cerchio in movimento; al contrario, l’altro gruppo di infanti, esposti a sequenze di suoni la cui linea prosodica divergeva dalla struttura statisticamente

regolata, tendeva ad associare la sequenza AB agli altri elementi della scena. Così i ricercatori commentano tali risultati:

Our results suggest that the cognitive capacities of infants are appropriately constrained and that language acquisition is most rapid when the structure of the linguistic input is well matched to these constraints [...]. [...] The perception of prosodic phrases in speech has parallels to the perception of phrases in nonlinguistic domains, such as music [...], and appears to be related to basic perceptual grouping principles [...]. To the extent that the language faculty is built up from or relies on phylogenetically preexisting cognitive capacities such as these grouping principles or the nature of memory and attentional processes more generally, studying them can offer insight into the system that enables language acquisition and the structure of language itself.<sup>211</sup>

In sintesi, dunque, possiamo affermare che il processo ontogenetico di acquisizione del linguaggio posa su almeno due vincoli cognitivi. In primo luogo, come abbiamo visto essa fa capo alla processazione di regolarità sia all'interno del flusso del discorso (che ne rende possibile la segmentazione nelle parole costitutive), sia a livello delle co-occorrenze dei termini in corrispondenza dei medesimi riferimenti in scene diverse; in secondo luogo, tale importante processo si radica nella più generale abilità di percepire e seguire i pattern prosodici del discorso che, come abbiamo visto, operano all'unisono con la processazione della struttura sintattica della frase, orientando la segmentazione dei sintagmi, e allo stesso tempo,

---

<sup>211</sup> *Ivi*, p. 6041.

le loro relazioni sintattiche all'interno del periodo<sup>212</sup>.

Riteniamo, con gli autori dello studio, che la ricerca sui principi di categorizzazione delle unità, ovvero della determinazione delle loro relazioni interne meriti ulteriori esami sperimentali volti non soltanto allo studio dell'abilità di associare le singole unità lessemiche agli oggetti delle scene contestuali, a mo' di etichette – ma anche sui principi di associazione tra i pattern *sintattici* rivelati dagli schemi prosodici che interessano le parole all'interno delle frasi e le corrispettive connessioni cosali nella scena. Per dirla in altre parole, crediamo sia opportuno volgere l'analisi sull'abilità - tanto dei bambini quanto degli adulti - di associare frasi prive di contenuto semantico, ma prosodicamente modulate - non a singoli oggetti, ma a *fatti*, ovvero a relazioni fra oggetti - occorrenti nelle scene cui sono esposti. In secondo luogo, al fine di guadagnare importanti indizi sulla comprensione dei meccanismi cognitivi implicati nella facoltà di

---

<sup>212</sup> A tale proposito, vedi Hunyadi, L. *Cognitive grouping and recursion in prosody*, in H. van der Hulst, J. Koster, H. van Riemsdij (ed.), "Recursion and Human Language. Studies in Generative Grammar", Berlin/New York: De Gruyter Mouton, 2010, pp. 343-370. In tale studio, l'autore dimostra la modulazione prosodica condotta attraverso raggruppamenti a livello tonale e di pause rappresenta un indizio molto forte per il raggruppamento in schemi ricorsivi sia nel dominio verbale che a in quello della rappresentazione di figure visive; *ivi*, pp. 350-351: "We have shown that the computational mechanism of recursion can be identified in prosody in two forms: in tonal and pausal phrasing. As for tonal phrasing, the principle behind nested prosodic phrase generation is tonal continuity. Since tonal continuity involves a change (usually the lowering) of the tonal space of the embedded phrase, applied recursively it results in the recursive unidirectional change (continuous relative lowering) of the tonal space. It was shown that the computational difference between nested recursion and iteration has its prosodic correlate. Namely, recursion is characterized by the lack of downdrift, whereas iteration by the presence of downdrift. [...] That recursion in prosodic grouping is exactly the same mechanism as the one found in the grouping of less specific, more abstract, non-linguistic prosodic and visual elements suggest that recursion in prosody cannot be the effect of an interface relation between syntax and prosody, instead, it is the manifestation of a more general, more universal computational mechanism found beyond linguistic structure."



linguaggio (e nelle dinamiche che ne hanno interessato l'evoluzione) troviamo sia auspicabile connettere tale tipo di indagine con la ricerca intorno alla dominio-generalità e alla presenza in altre specie animali di tali meccanismi cognitivi. Domini non strettamente verbali come quello visivo<sup>213</sup> o quello musicale ben si prestano, infatti, alla esplorazione dei processi cognitivi di base implicati nella acquisizione del linguaggio. L'analisi dei parallelismi cognitivi tra la prosodia discorsiva e la musica (e le implicazioni di tali dati sulla riflessione intorno all'evoluzione del linguaggio verbale) sarà proprio l'oggetto di indagine del prossimo capitolo.

La ricerca sperimentale su questo argomento è ancora ai suoi esordi, ma tenendo entro un'unica cornice teorica aspetti cognitivi fondamentali per la processazione del linguaggio - come l'apprendimento statistico, la caratterizzazione prosodica delle frasi, e la loro capacità di ancorarsi ad una realtà oggettuale esterna - si presta bene ad ampliare la comprensione di un fenomeno tanto intricato quanto l'acquisizione ontogenetica del linguaggio. Inoltre, operando all'unisono con l'approccio comparativo sulla cognizione animale, tale paradigma può essere uno dei migliori candidati per la ricerca sulle dinamiche evolutive del linguaggio.

---

<sup>213</sup> Vedi Fabre-Thorpe, M., *Visual categorization: accessing abstraction in non-human primates*, "Philosophical Transactions of the Royal Society B: Biological Sciences", 358, 2003, pp. 1215-1223; Brady, T., Oliva, A, *Statistical learning using real-world scenes*, cit.

## **9. Prosodia, musica e facoltà di linguaggio: coincidenze evoluzionarie?**

La modulazione prosodica gioca, come abbiamo visto, un ruolo cardinale nel processo di acquisizione del linguaggio e nello sviluppo tanto cognitivo quanto emozionale degli infanti, agevolandone il processo di acquisizione del linguaggio. Ma in che modo l'indagine sullo sviluppo ontogenetico di tale fondamentale abilità può illustrarne il ruolo giocato nel percorso evoluzionario della facoltà di linguaggio verbale? Crediamo che un ottimo punto di riflessione su tale quesito sia desumibile dalle osservazioni di Wittgenstein in merito all'intrinseca fusione dell'aspetto musicale-sintattico e semantico nelle proposizioni articolate:

Il capire una proposizione è più affine di quanto non si creda al capire un brano musicale [...]. Non dimenticare come stanno le cose quando si legge una proposizione con la cadenza sbagliata, e perciò non la si capisce, - e poi ti viene in mente come si debba leggerla.<sup>214</sup>

Nell'uso di una parola si potrebbe distinguere una "grammatica superficiale" da una "grammatica profonda". Ciò che si imprime immediatamente in noi, dell'uso di una parola, è il suo modo di impiego nella *costruzione della proposizione*, la parte del suo uso - si potrebbe dire - che cogliamo con l'orecchio.<sup>215</sup>

Mettendo in luce come il suono delle proposizioni (*die Satzklang*), il loro ritmo costituiscano una componente cruciale, talvolta sufficiente per la loro stessa comprensione, l'autore osserva che esiste un aspetto "profondo" delle proposizioni che ne fonda l'aspetto familiare:

Ponendo la domanda circa la forma proposizionale generale teniamo conto che il linguaggio comune ha certamente un determinato ritmo proposizionale, ma non tutto ciò che ha questo ritmo è una proposizione. Cioè *suona come una proposizione*, ma non lo è. - Donde l'idea di *proposizione* dotata di senso o insensata.<sup>216</sup>

Dall'intreccio di tali riflessioni con la tesi darwiniana intorno alla natura musicale della comunicazione presso i nostri antenati ominidi, emerge l'ipotesi chiave del presente capitolo, ovvero che la modulazione musicale, centrale nella determinazione del significato di una proposizione, rappresenti una sorta di fossile del linguaggio, in grado di rivelarne le

---

<sup>214</sup> Wittgenstein L., *Grammatica filosofica*, ed. it. di M. Trinchero, La Nuova Italia, Firenze, 1990, § 4, cit. in Gargani A. G., *Wittgenstein. Musica, parola, gesto*, Raffaello Cortina, Milano, 2008, p. 127.

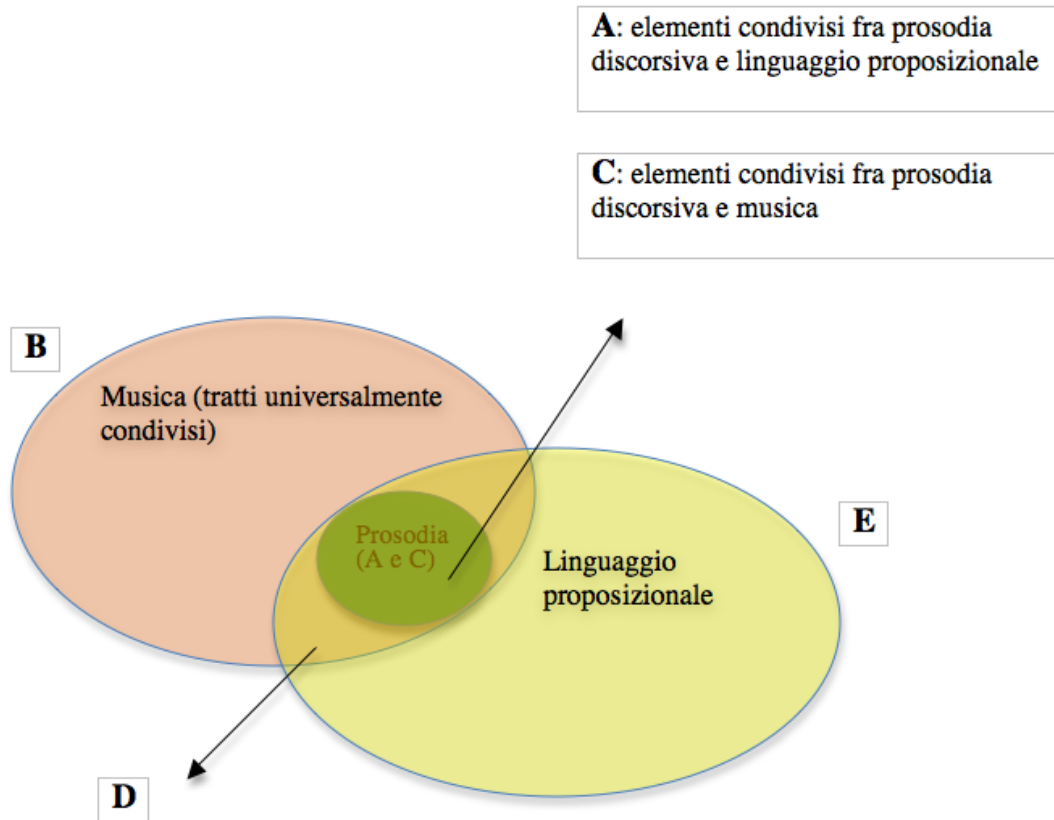
<sup>215</sup> Wittgenstein L., *Philosophische Untersuchungen*, cit., § 664.

<sup>216</sup> Wittgenstein L., *The Big Typescript*, cit., III, § 17, 2. Cfr. pp. 81-82.

forme originarie. L'idea che esploreremo nel presente capitolo è dunque che la musica, o più specificatamente la prosodia - rappresenti il terreno in cui l'evoluzione delle proposizioni articolate - e dei correlati neuronali dedicati alla loro processazione - affondano le proprie radici.

Al fine di illustrare il valore evolutivo della modulazione prosodica del linguaggio, è necessario centrare il fuoco dell'analisi sulla sua posizione intersecante tanto il linguaggio proposizionale quanto la musica (figura 1). In particolare, tale operazione richiede tre passaggi metodologici. In primo luogo, tenteremo di fornire una definizione del concetto di *musica* adottabile a livello cross-specifico (B, in figura 1). In secondo luogo, illustremo i punti di contatto tra musica e linguaggio proposizionale (D, in figura 1); infine, alla luce delle osservazioni argomentate nel precedente capitolo intorno alla portata cognitiva ed emozionale della prosodia nel processo di acquisizione ontogenetica del linguaggio, e dimostrata la presenza anche in specie animali non umane dell'abilità di riconoscere la linea prosodica nelle espressioni linguistiche - prenderemo in esame i tratti che la prosodia discorsiva condivide con la musica (C, in figura 1). Questi saranno i tre nuclei tematici che veneranno le prossime pagine della ricerca allo scopo di guadagnare una prospettiva più completa sul rapporto evolutivo tra linguaggio verbale, musica, e modulazione prosodica - illustrando allo stesso tempo i tratti cognitivi che,

distinguendo gli esseri umani dalle altre specie animali, ne rendono specifico il linguaggio.



**Fig. 1.**

### **9.1 Musica: somiglianze di famiglia (in specie diverse)**

Al fine di proseguire l'indagine intorno al valore cognitivo ed evolutivo della modulazione prosodica del linguaggio, è opportuno inserire tale analisi in un quadro di studio più ampio, che comprenda lo studio comparativo di un fenomeno particolarmente rilevante ai fini della

nostra ricerca: la musica. In tale direzione è anzitutto preliminarmente necessario fornire una definizione operativa di tale termine. Infatti, se consideriamo il canto di un uccello, i duetti dei gibboni, una canzone dei Beatles e il brano 4'33" di John Cage<sup>217</sup> possiamo osservare come, pur rientrando entro la medesima area semantica indicata dal termine ombrello "musica", costituiscono in realtà realizzazioni completamente differenti ed in qualche modo assai lontane fra loro. Per questo motivo, è opportuno considerare la musica come un'entità che comprende in sé diverse forme che presentano wittgensteiniane somiglianze di famiglia. In altri termini, le diverse forme di musica non presentano un unico tratto che le accomuna definendone l'essenza, ma - come in una corda sfilacciata i cui filamenti si intrecciano e sovrappongono - ciascuna di essa presenta tratti comuni ad altre forme, ma anche tratti totalmente diversi.

La musicologa Emilie Doolittle ha incentrato la sua ricerca di dottorato<sup>218</sup> sul tentativo (a nostro avviso ben riuscito) di fornire una definizione della capacità di fare musica individuando una serie di tratti che si presentano, se non in tutti, almeno nella maggiorparte delle realizzazioni musicali osservabili nella specie umana e in diverse specie animali, in particolare negli uccelli canori. La studiosa vi si riferisce con il termine

---

<sup>217</sup>Lo spartito del brano prevede che l'esecutore non suoni per tutta la durata del brano (4'33"), cosicché la composizione stessa, sebbene percepita come quattro minuti e trentatré secondi di silenzio, consiste dei suoni emessi dall'ambiente in cui viene eseguita e dall'uditorio che ne fruisce.

<sup>218</sup>Doolittle, E. L., *Other Species' Counterpoint. An investigation of the relationship between Human Music and Animal Song*, tesi di dottorato presentata alla University of Princeton, 2007.

“proto-musicali”, e riconosce in ciascuno di esso una precondizione necessaria e sufficiente alla nascita della capacità di fare musica negli esseri umani. Di seguito l’elenco di tali tratti:

- l’abilità di creare suoni a piacere,
- l’abilità di controllare la produzione di suoni,
- l’abilità di apprendimento vocale,
- l’abilità di imitare,
- l’abilità di creare nuove combinazioni di suoni,
- l’abilità di connettere i suoni ad un’attività fisica,
- l’abilità di connettere i suoni ad uno stato emozionale,
- l’abilità di usare i suoni per comunicare,
- l’abilità di sincronizzare i suoni con un ritmo esterno.
- 

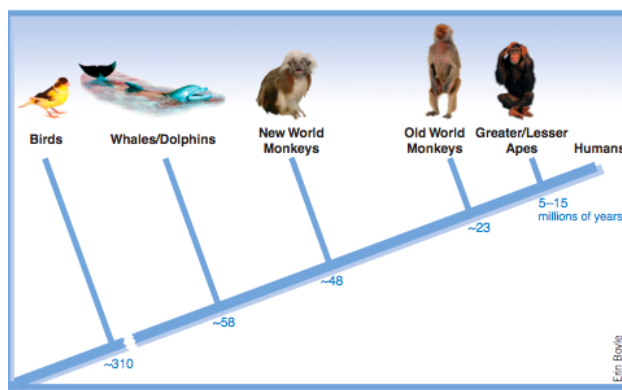
Proto-musical Traits	Mozart Requiem	Beatles “Love, love me do”	Shower singing	Cage 4’33”
<b>preconditions:</b>				
make sounds at will	yes	yes	yes	no (conspicuous for absence)
control sound production	yes	yes	yes	no (conspicuous for absence)
vocal learning	yes	yes	yes	no (conspicuous for absence)
imitation	yes	yes	maybe	no (conspicuous for absence)
new sound combinations	yes	yes	yes	yes (by chance)
connect to physical activity	no	yes	yes	no
connect to emotional state	yes	yes	yes	maybe
use sounds for communication	yes	yes	yes (to self)	maybe
coordinate with external rhythm	yes	yes	probably not	no

**Fig. 2.** Immagine modificata da Doolittle (2007)

Come illustrato nella figura 2, la studiosa descrive la distribuzione di tali tratti in diverse forme di realizzazioni musicali umane. La presenza di almeno uno di questi tratti, come ad esempio la capacità di creare nuove combinazioni di suoni in 4'33" è sufficiente a caratterizzare un'esecuzione in quanto musicale. Peraltro, è interessante notare che in questo caso particolare l'intenzionalità della funzione comunicativa, è attribuibile al compositore, il quale non si identifica nel diretto esecutore del pezzo, ovvero colui che crea la combinazione dei suoni. Questi infatti non è nè il compositore, nè il musicista, bensì l'uditorio e l'ambiente circostante, i quali (in maniera non direttamente intenzionale) danno luogo ad una combinazione non regolata di suoni. Una volta definiti i "tratti familiari" il cui intreccio determina il concetto stesso di musica presso la specie umana, è opportuno dirigere il fuoco dell'attenzione verso un orizzonte di ricerca comparativa funzionale al tentativo di individuare quegli elementi musicali che, universalmente diffusi nel regno animale rappresentano gli aspetti chiave su cui focalizzare l'indagine sul rapporto evolutivo tra musica e linguaggio umano. A tal proposito, è opportuno ribadire che la presenza dei medesimi tratti biocognitivi in specie filogeneticamente assai lontane è indice di una convergenza evolutiva dettata non direttamente dalla ereditabilità dei geni responsabili di tale tratto, ma dalla presenza di simili - se non identiche - pressioni selettive (figura 3). Così, i tratti legati alla



capacità di fare musica, la cui presenza è attestata in specie diverse – nell’indicarci le funzioni adattative che ne hanno determinato l’evoluzione – si prestano ad essere identificati in quel nucleo originario di elementi biocognitivi da cui ha preso avvio l’evoluzione della facoltà di linguaggio proposizionale.



**Fig. 3.** Immagine modificata da Hauser *et al.* (2003). Linea filogenetica dei gruppi tassonomici più importanti per lo studio dell’evoluzione della musica.

La stessa Doolittle ha indagato la presenza delle abilità sovra elencate (atte a definire il concetto di musica presso la nostra specie) in animali non umani, trovando che alcune specie di uccelli canori presentano la quasi totalità di esse, a differenza di altre specie come i gibboni o i lupi che mancano di aspetti cognitivi importanti come la capacità di apprendimento vocale o di imitazione (figura 4) .

Proto-musical Traits	Humans	European Blackbirds	Zebra finches	Humpback whales
<b>preconditions:</b>				
make sounds at will	yes	yes	yes	yes
control sound production	yes	yes	yes	yes
vocal learning	yes	yes	yes	yes
imitation	yes	<b>yes</b>	no	yes
new sound combinations	yes	yes	yes	yes
physical activity	yes	yes	yes	yes
connect to emotional state	yes	yes	yes	yes
communication	yes	yes	yes	yes
coordinate with external rhythm	yes	no	no	no

Proto-musical Traits	Humans	Wolves
<b>preconditions:</b>		
make sounds at will	yes	yes
control sound production	yes	yes
vocal learning	yes	no
imitation	yes	<b>no</b>
new sound combinations	yes	(by chance)
connect to physical activity	yes	yes
connect to emotional state	yes	yes
use sounds for communication	yes	yes
coordinate with external rhythm	yes	(no)

**Fig. 4.** Immagini modificate da Doolittle (2007)

Crediamo sia opportuno aggiungere a tali abilità cross-specifiche individuate da Doolittle ulteriori tratti bio-cognitivi universali nella specie umana e la cui presenza è stata inoltre recentemente attestata in alcune specie di scimmie ed uccelli: la generalizzazione dell'ottava, la percezione di note discrete, di scale, e di strutture ritmiche, e l'abilità di modulare la frequenza, il tempo e l'ampiezza dei suoni. Per quanto concerne il primo di tali tratti, studi etnomusicologici attestano che la maggior parte delle culture umane è in grado di riconoscere una stessa nota trasposta di

un'ottava, cosicché una canzone cantata in tonalità diverse verrà comunque percepita come “la stessa canzone”. In uno studio pubblicato nel 2000<sup>219</sup>, Anthony Wright ed i suoi collaboratori si sono chiesti se tale abilità è appannaggio dei primati non umani, testando in particolare il macaco rhesus. Ne è emerso quanto segue:

Rhesus monkeys showed octave generalization in five different experiments [...]. [...] These octave generalizations results, we believe, provide the first clear evidence of a nonhuman species perceiving musical passages in much the same way that human perceive them. This evidence is further bolstered by results showing a lack of generalization to random synthetic melodies, atonal melodies, half-octave transpositions, and single note (pitch) transpositions. [...] The results from these [...] experiments converge on the conclusion that musically strong tunes become a gestalt, a unit unto itself, for rhesus monkeys as they do form humans [...]. The melody can be transposed and moved on the frequency scale, with virtually undiminished recognition relative to the original, provided that the frequency change is some multiple or factor of two.<sup>220</sup>

Per quanto ne sappiamo ad oggi, fra i primati non umani i rhesus sono l'unica specie a presentare tale specifica sensibilità musicale. Tuttavia, ulteriori studi sul tema dimostrano che l'abilità di riconoscere la linea melodica a prescindere dalla tonalità è presente nei ratti bianchi e nei delfini<sup>221</sup>, ma non negli uccelli, i quali in generale sembrano codificare gli stimoli musicali in termini di note assolute e discrete, piuttosto che

---

<sup>219</sup> Wright, A., Rivera, J., Hulse, S., Neiworth, J. J., *Music perception and octave generalization in rhesus monkeys*, “Journal of Experimental Psychology”, 129, 2000, pp. 291-307.

<sup>220</sup> *Ivi*, p. 304.

<sup>221</sup> Blackwell, H. R., Schlosberg, H., *Octave generalization, pitch discrimination, and loudness thresholds in the white rat*, “Journal of Experimental Psychology”, 33, 1943, pp. 407-419; Ralston, J. V., *et al.*, *Melody Recognition by an Atlantic Bottlenose Dolphin*, “Journal of the Acoustical Society of America”, 1988 - in Doolittle, E. L., *op. cit.*

funzionalmente alla relazione strutturale delle note, quale si dà in una data scala. Infatti, come osservano McDermott e Hauser,

Relative pitch has some importance in birdsong recognition, at least in some species in which songs are defined by specific frequency ratios between the “notes” composing the song [...]. [...] To test whether birds extract relative pitch from melodies as humans do, starlings were trained to respond to falling but not rising melodies, a task they can master given enough training (Hulse, Cynx, & Humpal, 1984). The rising and falling melodies used were composed of pure tone notes and typically spanned an octave range of frequencies. [...] the birds were unable to generalize the discrimination to novel melodies transposed up or down by an octave and whose pitches, as a result, fell outside the range in which they had been trained [...]. This result has been replicated several times in various species of birds, including some nonsongbirds (Cynx, 1995). [...] This would appear to be at odds with the behavior of adult and infant humans, who readily recognize melodies across large transpositions, especially octave transpositions.<sup>222</sup>

In altri termini, possiamo affermare che sebbene alcune specie di uccelli usino una serie di note ascendenti e discendenti che almeno all’orecchio umano suonano simili ad una scala, e la cui relazione sembra essere regolata da specifici intervalli, essi non presentano la capacità di trasporre spontaneamente (ovvero senza lunghe ed intense e sessioni di training sperimentali) un medesimo pattern strutturale su suoni diversi. È opportuno, dunque, chiedersi se tali specie presentano dei vincoli percettivi diversi da quelli umani, che ne modificano l’esperienza stessa dei suoni. In tale direzione di ricerca, diversi studi sperimentali dimostrano che in effetti

---

<sup>222</sup> Vedi McDermott, J., Hauser, M. D., *The origins of music: Innateness, uniqueness, and evolution*, “Music Perception: An Interdisciplinary Journal”, 23, 2005, pp. 29-59.

esistono dei vincoli percettivi comuni (non meglio identificati), che permettono tanto agli uomini quanto agli uccelli e alle scimmie la distinzione di suoni consonanti da quelli dissonanti<sup>223</sup>

Un ulteriore aspetto musicale culturalmente universale su cui vale la pena porre l'attenzione è la struttura ritmica degli stimoli musicali. Come osservato nel capitolo precedente, diverse ricerche hanno attestato l'abilità di riconoscere strutture temporali in alcune specie non umane in stimoli dalla natura linguistica. Tuttavia, anche sul piano della produzione vocale, è riconoscibile la presenza di chiari pattern ritmici, in particolare nei canti di talune specie di uccelli canori. In effetti, sebbene non rispondano alle convenzioni metriche della musica umana, essi presentano un ritmo stabile e internamente coerente<sup>224</sup>. Per di più, specifici pattern temporali sono riscontrabili oltre che nei canti degli uccelli, anche nelle attività di percussione realizzate dagli scimpanzè in accompagnamento ai loro *pant hoots*<sup>225</sup>, e nei duetti performati dalle coppie di gibboni.

---

<sup>223</sup> Vedi Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriya, Y., Takeshita, H., Nakamura, T., Hashiya, K., *Preference for consonant music over dissonant music by an infant chimpanzee*, "Primates", 51, 2009, pp. 7-12; Izumi, A., *Japanese monkeys perceive sensory consonance of chords*, "The Journal of the Acoustical Society of America", 108, 2000, pp. 3073-3078; Hulse, S. H., Bernard D. J., Braten, R. F., 1995, *Auditory discrimination of chord-based spectral structures by European starlings (Sturnus vulgaris)*, "Journal of Experimental Psychology: General", 124, 1995, pp. 409-423.

<sup>224</sup> Interessante, tal proposito, lo studio condotto su Snowball, un esemplare di grande cacatua ciuffogiallo in grado di sincronizzare il movimenti del corpo al ritmo di un brano musicale: Patel, A., Iversen, J., Bregman, M., *Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal*, "Current Biology", 19, 2009, pp. 827-830.

<sup>225</sup> Arcadi, A., Robert, D., *Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness*, "Primates", 39, 1998, pp. 505-518.

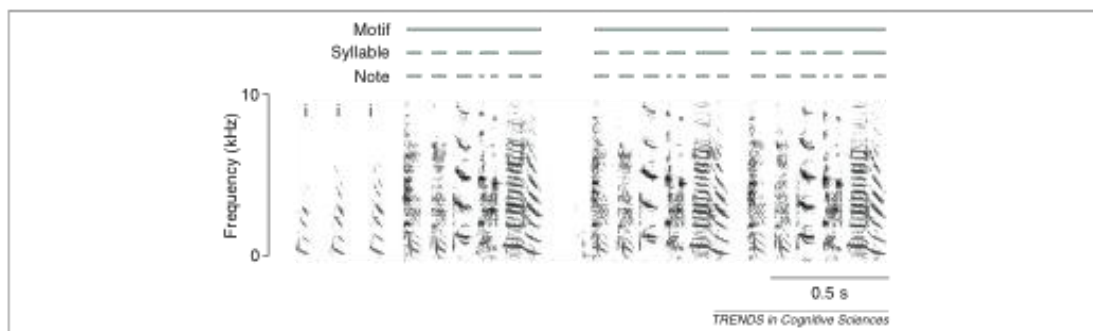
A tale insieme di aspetti musicali cross-specifici è opportuno aggiungere un ulteriore elemento in cui è possibile riconoscere un cruciale tratto di convergenza evolutiva: la strutturazione gerarchica degli elementi all'interno della vocalizzazione musicale. Infatti, così come il linguaggio proposizionale presenta una organizzazione gerarchica in funzione della quale i periodi si articolano in frasi, le frasi in sintagmi, i sintagmi in parole e così via – allo stesso modo, i canti degli uccelli e delle balene presentano una struttura tale che è possibile individuare sillabe, frasi, motivi e temierarchicamente articolati (figura 5). Tuttavia, come notano Doupe e Kuhl<sup>226</sup> a proposito di tale tipo di organizzazione nei canti degli uccelli, i principi di strutturazione dei pattern non sono individuabili in regole grammaticali come nel linguaggio umano, bensì semplicemente nella suddivisione temporale del segnale dall'unità più breve fino alle unità più lunghe, consistenti nella mera sequenza temporale degli elementi costitutivi:

Another feature that birdsong and language share is the conspicuous timing and ordering of components on a timescale longer than that of the syllable. Song syllables are usually grouped together to form phrases or “motifs” [...], which can be a series of identical or different syllables. Many songbirds sing several phrases in a fixed order as a unit, which constitutes the song, whereas other species such as mockingbirds and warblers produce groups of syllables in fixed or variable sequences. The timing and sequencing of syllables and phrases are rarely random but instead follow a set of rules particular to a species. In the songbird literature,

---

<sup>226</sup> Doupe, A. J., Kuhl, P. K., *Birdsong and human speech: common themes and mechanisms*, “Annual Review of Neuroscience”, 22, 1999, pp. 567-631.

the ordering of syllables and phrases in song is often called song syntax. The same word applied to human speech, however, implies grammar, i.e. rules for ordering words from various grammatical classes to convey meaning. Therefore, in this review, we avoid using the word syntax for song and simply use “order.” Thus, language and song share a dependence on timing on several timescales: a shorter timescale (on the order of tens of milliseconds), as in phonemes and syllables, and a longer one, up to many hundreds of milliseconds (as in syllable, phrase, and word ordering).<sup>227</sup>



**Fig. 5.** Immagine modificata da Berwick *et al.* (2011). Spettrogramma di un tipico canto dei diamanti mandarini (*Taeniopygia guttata*) dotati di una struttura gerarchica. Le unità minime dei canti sono le note, ovvero singole tracce sonore in cui tempo e frequenza sono in coerenza. Più note ininterrotte costituiscono una sillaba. I canti iniziano spesso con delle note introduttive (indicate nella figura dalla “i”), seguite da uno più motivi, ovvero da ripetizioni di sequenze di sillabe. Una continua riproduzione di più motivi costituisce a sua volta il tema del canto (*song bout*).

In ultimo, dunque, possiamo sintetizzare quanto esposto nel presente paragrafo includendo in un unico quadro d’insieme i tratti biocognitivi che concorrono a determinare una definizione operativa completa del concetto di musica (figura 6). Tali tratti rappresentano i punti chiave in funzione dei quali poter realizzare un’analisi comparativa volta alla comprensione del rapporto evolutivo tra musica, modulazione prosodica e linguaggio proposizionale.

<sup>227</sup> Ivi, pp. 572-573.

**DEFINIZIONE OPERATIVA DELLA CAPACITÀ  
DI CREARE MUSICA**

(PRECONDIZIONI NECESSARIE PRESENTI IN SPECIE NON UMANE)

1. abilità di creare suoni a piacere
2. abilità di controllare la produzione di suoni
3. abilità di apprendimento vocale
4. abilità di imitare
5. abilità di creare nuove combinazioni di suoni
6. abilità di connettere i suoni ad un'attività fisica
7. abilità di connettere i suoni ad uno stato emozionale
8. abilità di usare i suoni per comunicare
9. abilità di modulare la frequenza, la durata e l'ampiezza dei suoni\*
10. abilità di strutturazione ritmica\*
11. abilità di discernere uno stesso suono nell'ambito di scale trasposte in ottave diverse\*
12. abilità di percepire suoni consonanti/dissonanti\*
13. organizzazione strutturale delle unità musicali\*

**Fig. 6.** Aspetti che concorrono a determinare una definizione operativa del concetto di musica funzionale allo studio del rapporto evolutivo tra musica, modulazione prosodica e linguaggio verbale - (i tratti contrassegnati da un asterisco sono quelli aggiunti dall'autrice all'elenco stilato in Doolittle, 2007).



## 9.2 Prosodia: la musica dei discorsi

Se si presta ben attenzione alla tabella raffigurata in figura 7, tra le capacità biocognitive elencate ve ne sono alcune in cui è possibile identificare gli elementi definatori della prosodia discorsiva. Si tratta a) dell'abilità di controllo vocale, b) di modulare la frequenza, la durata e l'ampiezza dei suoni, c) di strutturazione ritmica, d) di connettere i suoni ad uno stato emozionale, ed infine e) di organizzare una sequenza di suoni in unità strutturali<sup>228</sup>. Da un punto di vista percettivo è però evidente che l'attività del parlare modulando prosodicamente le frasi non suona *ipso facto* come il cantare un brano musicale. In cosa risiede la differenza? Per rispondere a tale quesito, vale la pena di considerare le osservazioni di Ray Jackendoff:

Are the characteristics of tonal pitch spaces shared at all with language? Consider a number of possible parallels. First, prosodic contours in language tend to go down at the end, and so do melodies. But both probably inherit this from the form of human and mammalian calls (and possibly from the physiology that gives rise to calls, e.g., the drop in air pressure as the lungs are emptied). Thus this common characteristic could be the product of independent inheritance from a common ancestor. *An important difference between prosodic contours and melodies is that only melodies have discrete pitches, while prosodic contours usually involve a continuous rise and fall.* Of course there are mixtures. On the one hand, many vocal and instrumental traditions incorporate bending of pitches and sliding between them. [...] Thus, one might be able to think of intonation in language as fundamentally a two-pitch tonal system, modulated by continuous transitions between the anchoring pitches. However, the high and low pitches are not fixed in

---

<sup>228</sup> Ritengo quest'ultima abilità un sottocomponente della più complessa capacità di organizzare i suoni in strutture gerarchiche.

frequency throughout an utterance in the way the dominant and tonic are fixed in musical pitch space.<sup>229</sup>

Secondo lo studioso, dunque, la differenza saliente tra musica e prosodia consiste nel fatto che mentre la prima vanta una strutturazione in note discrete, la seconda consiste in un *continuum* intonazionale modulatorio con movimenti ascendenti e discendenti. In tale direzione, anche Aniruddh Patel, uno dei massimi neuroscienziati impegnati nella ricerca sul rapporto tra musica e linguaggio – individua nella mancanza di una stabile struttura negli intervalli fra le note l'elemento chiave nella divergenza tra musica e prosodia discorsiva. Tuttavia, illustrando dei punti salienti messi in luce dal paradigma di studio chiamato AM (*autosegmental-metrical*), spiega che:

[...] linguistic aspects of intonation are based on *categorical* distinctions involving sequences of discrete pitch events. Specifically, the continuous F0 [frequenza fondamentale] contours of speech are viewed as reflecting a sequence of distinct “pitch accents” and “edge tones” associated with certain points in the physical F0 contour. The rest of the intonation contour is considered a mere interpolation between these events, so that most of the F0 trajectory is not linguistically significant. That is, F0 contours in speech are seen as the result of moving between certain well-defined targets in pitch and time, a view quite analogous to the conception of melody in music [...].<sup>230</sup>

---

<sup>229</sup> Jackendoff, R., *Parallels and Nonparallels between Language and Music*, “*Music Perception*”, 26, 2009, p. 199 (corsivo mio).

<sup>230</sup> Patel, A. D., “*Music, Language, and the Brain*”, Oxford, Oxford University Press, 2008, p. 207.

Ciò vuol dire che sebbene le frequenze sonore della prosodia discorsiva manchino di una struttura regolare (ovvero di quegli intervalli regolari specifici della grammatica musicale) nell'ambito della quale contestualizzare il riconoscimento delle note, è nondimeno possibile isolare talune note discrete nel *continuum* sonoro della modulazione prosodica.

In termini evolutivi possiamo affermare che il controllo attivo della produzione dei suoni (resa possibile dall'instanziazione delle connessioni neuro-motorie) ha permesso, un po' come accade nel caso della maggior parte delle specie di uccelli canori finora studiati, la modulazione della frequenza, della durata dei suoni, dell'ampiezza e degli schemi ritmici. Ed è proprio in tali parametri acustici, unitamente alla presenza di note discrete che è possibile identificare nei discorsi che proferiamo uno scenario in qualche modo *musicale*, nonostante la mancanza di specifiche scale nell'ambito delle quali orientare la stessa percezione (e produzione) delle note.

Come illustrato nel capitolo precedente, l'impiego di costituenti acustici come la durata delle sillabe delle parole o la variazione della frequenza fondamentale conferisce all'espressione un significato di natura emozionale, e permette di organizzare il flusso discorsivo in unità significazionali, orientandone la connessione sintattica. È interessante notare che la modulazione di determinati parametri acustici sortisce effetti

emozionali molto simili nell'ambito della musica ed in quello più specifico della prosodia. In altre parole, le caratteristiche acustiche del segnale stesso (sia esso musicale o linguistico) sono in grado di veicolare una certa coloritura emozionale contribuendo alla determinazione della sua portata significazionale. A tal proposito, uno studio condotto da Fritz (studioso del medesimo gruppo di ricerca di Friederici e Koelsch), dimostra che emozioni basilari come gioia, tristezza, e paura – espresse nella musica sono individuabili tanto nella cultura occidentale, quanto in una popolazione nativa dell'Africa (Mafa) sulla base di parametri acustici come il tempo o il modo musicale:

In conclusion, both Mafa and Western listeners showed an ability to recognize the three basic emotional expressions tested in this study (happy, sad, and scared/fearful) from Western music above chance level. This indicates that these emotional expressions conveyed by the Western musical excerpts can be universally recognized, similar to the largely universal recognition of human emotional facial expression [...] and emotional prosody [...].<sup>231</sup>

In merito al colore emozionale dei discorsi prosodicamente modulati e di segnali musicali, Ilie e Thompson<sup>232</sup> hanno condotto uno studio di ricerca focalizzato su tre parametri in particolare (il volume, il ritmo e

---

<sup>231</sup> Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., Koelsch, S., *Universal Recognition of Three Basic Emotions in Music*, "Current Biology", 19, 2009, pp. 1–4.

<sup>232</sup> Ilie, G., Thompson, F., *A comparison of acoustic cues in music and speech for three dimensions of affect*, "Music Perception", 23, 2006, pp. 319-329.

l'altezza) e i loro effetti in termini di piacevolezza, arousal energetico e tensione. Da tale lavoro emerge che tanto nell'ambito della musica quanto della prosodia, un volume più alto viene giudicato meno piacevole, ma incrementa l'arousal e la tensione, così come la presenza di un ritmo sostenuto. In modo interessante, il parametro che segna la divergenza tra musica e prosodia relativamente alla sollecitazione emozionale è l'altezza della frequenza, che nello specifico, se alta, risulta meno piacevole in musica e più piacevole nell'ambito linguistico (figura 7). Così ne parlano le autrici:

Our findings suggest that intensity manipulations had similar affective consequences for all dimensions of affect, rate manipulations had similar consequences for judgments of energy, and pitch manipulations had different affective consequences in music and speech. Similarities between music and speech help to explain why listeners perceive music as expressive of emotion [...] and suggest that emotional processing in music and speech involves shared neural resources [...].<sup>233</sup>

Potremmo sostenere, speculativamente, che l'esagerazione della frequenza risulti maggiormente piacevole nella dimensione linguistica in virtù - oltre che del suo positivo valore emozionale - del suo intrinseco effetto nell'amplificare la capacità di discriminazione fonologica e lessicale dello stimolo acustico cui si è esposti.

---

<sup>233</sup> *Ivi*, p. 325.

TABLE 3. Significant main effects of intensity, rate, pitch height on valence, energy arousal, and tension arousal for music and speech.

Main finding by emotion dimension					
Stimulus properties			Valence	Energy arousal	Tension arousal
Intensity	Loud	Music	-	+	+
	Soft		+	-	-
	Loud	Speech	-	+	+
	Soft		+	-	-
Rate	Fast	Music	/	+	+
	Slow		/	-	-
	Fast	Speech	-	+	/
	Slow		+	-	/
Pitch height	High	Music	-	/	+
	Low		+	/	-
	High	Speech	+	+	/
	Low		-	-	/

Note. "+" indicates significantly higher ratings on the affect dimension represented, "-" indicates lower ratings on the affect dimension represented, "/" indicates no significant difference.

**Fig. 7.** Immagine modificata da Ilie *et al.* (2006).

- “+” indica la presenza di valutazioni significativamente più positive da parte dei soggetti esposti ai passi (musicali o verbali), relativamente alla dimensione affettiva.
- “-” indica che le valutazioni erano significativamente più negative, e
- “/” che non erano presenti differenze significative nelle valutazioni.

In tale ambito di indagine, sono essenziali le osservazioni emergenti dallo studio di Juslin e Laukka, relativamente ai particolari parametri acustici latori di informazioni emozionale in entrambi i canali comunicativi:

We argue that the hypothesis that there is an iconic similarity between vocal expression of emotion and musical expression of emotion applies only to certain acoustic features—primarily those features of the music that the performer can control (more or less freely) during his or her performance such as tempo, loudness, and timbre. However, the hypothesis does not apply to such features of a piece of music that are usually indicated in the notation of the piece (e.g., harmony, tonality, melodic progression), because these features reflect to a larger extent characteristics of music as a human art form that follows its own

intrinsic rules and that varies from one culture to another [...]. Neuropsychological research indicates that certain aspects of music (e.g., timbre) share the same neural resources as speech, whereas others (e.g., tonality) draw on resources that are unique to music [...]. Thus, we argue that musicians communicate emotions to listeners via their performances of music by using emotion-specific patterns of acoustic cues derived from vocal expression of emotion [...].<sup>234</sup>

Secondo gli studiosi, dunque, la portata emozionale dei segnali è resa, tanto nella musica quanto nella espressione vocale, dalla modulazione di parametri acustici specifici (tempo, timbro/frequenza, ampiezza) in grado di trasferire determinate emozioni di base come la rabbia, la gioia, la paura, la tristezza o la tenerezza. In modo interessante gli autori puntualizzano che proprio a partire da tali tratti acustici condivisi ha preso avvio l'evoluzione della musica verso la propria configurazione attuale, governata da leggi specifiche inerenti l'armonia, la tonalità o la progressione melodica.

---

<sup>234</sup> Juslin, P. N., Laukka, P., *Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?*, "Psychological Bulletin", 129, 2003, p. 774.

**TABLE 1.** *Emotional attributions significantly associated with acoustic parameters (reproduced from ref. 40, p. 339)*

Acoustic parameters of tone sequences	Direction of effect	Emotion rating scales listed in decreasing order of associative strength
Amplitude variation	Small Large	Happiness, pleasantness, activity Fear
Pitch variation	Small Large	Disgust, anger, fear, boredom Happiness, pleasantness, activity, surprise
Pitch contour	Down Up	Boredom, pleasantness, sadness Fear, surprise, anger, potency
Pitch level	Low High	Boredom, pleasantness, sadness Surprise, potency, anger, fear, activity
Tempo	Slow Fast	Sadness, boredom, disgust Activity, surprise, happiness, pleasantness, potency, fear, anger
Envelope	Round Sharp	Disgust, sadness, fear, boredom, potency Pleasantness, happiness, surprise, activity
Filtration cutoff	Intermediate (few)	Pleasantness, boredom, happiness, sadness
Level (number of harmonics)	High (many)	Potency, anger, disgust, fear, activity, surprise

**Fig. 8.** Immagine modificata da Scherer(1995)<sup>235</sup>

Alla luce di tali dati inerenti la portata emozionale della comunicazione vocale negli uomini, e considerata la caratteristica coloritura intrinsecamente emozionale delle comunicazioni delle specie non umane, troviamo altresì attendibile l'ipotesi che determinati pattern di modulazione in frequenza, durata, ritmo e ampiezza delle vocalizzazioni fossero già ampiamente in uso presso gli ominidi durante il processo di

<sup>235</sup> Scherer, K. L., Expression of emotion in voice and music, "Journal of Voice", 9, 1995, pp. 235-248.



acquisizione della capacità di modulare suoni fonetici. La nostra ipotesi è che questi ultimi, una volta collocati nella linea intonazionale già in uso, ne abbiano assorbito, mantendoli, gli schemi strutturali; in altri termini, crediamo che gli schemi logici delle protoespressioni verbali abbiano sviluppato una certa sistematizzazione logico-sintattica a partire dagli schemi strutturali dei suoni nella linea intonazionale... quegli stessi schemi prosodici in funzione dei quali ancora oggi è possibile stabilire non soltanto il valore pragmatico ed emozionale di una determinata espressione, ma anche il rapporto sintattico tra le unità nelle frasi. E probabilmente, come nel caso dei canti degli uccelli o dei duetti dei gibboni, che presentano un pattern sintattico variabile soltanto entro l'ambito di schemi flessibili – le vocalizzazioni emesse dai primi ominidi presentavano altresì un movimento prosodico relativamente regolare. Ancora oggi, in effetti, tali linee intonazionali e ritmiche permeano le lingue storico-naturali, consentendo (a prescindere dalle informazioni segmentali) di distinguere una domanda da un ordine o da un'imprecazione, la presenza di una proposizione accidentale nell'ambito di un periodo più ampio, o il valore pragmatico-emozionale dell'espressione.

Assumiamo, inoltre, che la stessa marcatura prosodica avrebbe favorito la processazione delle espressioni protolingustiche. Infatti, in un interessante lavoro di ricerca dal titolo "Songs as an aid for language

acquisition”<sup>236</sup>, Schön *et al.*, dimostrano che la modulazione in note di una sequenza di unità statisticamente definite<sup>237</sup> ne favorisce sia la segmentazione che la memorizzazione, confermando che:

Indeed, songs may contribute to language acquisition in several ways. First, the emotional aspects of a song may increase the level of arousal and attention. Second, from a perceptual point of view, the presence of pitch contours may enhance phonological discrimination, since syllable change is often accompanied by a change in pitch. Third, the consistent mapping of musical and linguistic structure may optimize the operation of learning mechanisms.<sup>238</sup>

In conclusione, possiamo affermare che la linea prosodica vanta degli schemi di modulazione “a maglia larga”, ovvero pattern flessibili che, seppur non regolati da principi di ordine matematico (come nel caso della musica *strico sensu*), sono nondimeno in grado di amplificare la capacità di processazione delle unità segmentali, orientandone le interconnessioni, e conferendovi un significato in funzione della portata emozionale e pragmatica della stessa *melodia* discorsiva. Pertanto, crediamo che ciò legittimi a riconoscere nella prosodia una sorta di fossile primordiale del linguaggio, nonchè l’enzima che ne ha catalizzato l’evoluzione della

---

<sup>236</sup> Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R., *Songs as an aid for language acquisition*, “Cognition”, 106, 2008, pp. 975–983. Cfr. Nakata, T., *Infants' responsiveness to maternal speech and singing*, “Infant Behavior and Development”, 27, 2004, pp. 455-464.

<sup>237</sup> Vedi Saffran J.R., Aslin R. N., Newport E. L., ‘Statistical Learning by 8-Month-old Infants’, *Science*, 274, 1996, 1926-1928.

<sup>238</sup> *Ivi*, p. 976. Cfr. Fernald, A., *Meaningful melodies in mothers' speech to infants*, in H. Papousek, U. Jurgens, M. Papousek (ed.), “Nonverbal vocal communication: comparative and developmental approaches”, Cambridge University Press, Cambridge, 1992, pp. 262-282.

grammatica morfosintattica. Rimane tuttavia da chiedersi se la modulazione prosodica non sia al contempo identificabile con il fossile della musica. Probabilmente non giungeremo mai ad ottenere una risposta esatta e definitiva a quest'ultimo quesito. Tuttavia indizi assai significativi possono essere guadagnati a partire da un'analisi comparativa tra linguaggio verbale e musica. È ciò di cui ci occuperemo nel prossimo paragrafo.

### **9.3 Linguaggio e musica: affinità selettive**

Il rapporto evolutivo e cognitivo tra il linguaggio proposizionale e la musica è oggi un tema assai ricercato, e specialmente nell'ultimo decennio ha ottenuto una crescente comprensione grazie al connubio tra le teorie musicale e linguistica e le nuove tecnologie di *neuroimaging*. Un'idea essenziale in tale ambito di ricerca è che la musica e il linguaggio condividano specifiche risorse neurali, idea che converge con un dato incontrovertibile: tutte le lingue storico-naturali fanno uso della modulazione prosodica per trasferire informazioni di natura diversa (emozionale, pragmatica, ed in modo assai importante per l'economia del nostro studio, sintattica) - e circa la metà di esse usa la tonalità come mezzo per marcare una distinzione lessicale. Ad esempio, la sillaba /ma/ in cinese

mandarino assume un significato diverso (“madre”, o “intorpidito”) a seconda che venga pronunciata con un tono alto, o in crescendo. Inoltre, come emerge dalla ricerca condotta da Wong *et al.*, l’esperienza musicale favorisce una migliore processazione dei segnali linguistici stessi, in particolare delle informazioni legate alla frequenza musicale in cui le parole vengono proferite, al timbro, o alla durata dei fonemi:

In sum, we found more robust and faithful encoding of linguistic pitch information by musicians. Such encoding, arguably associated with increased musical pitch usage, may reflect a positive side effect of context-general corticofugal tuning of the afferent system, implying that long-term music-making may shape basic sensory circuitry. These results complement our existing knowledge of the brainstem’s role in encoding speech and frequency modulation by demonstrating the interplay between music and speech, subcortical and cortical structures, and the impact of long-term auditory experiences.<sup>239</sup>

Uno dei maggiori neuroscienziati contemporanei dedicati alla ricerca su tale tema è, come già accennato, Aniruddh Patel. In “Language, Music, syntax, and the brain”<sup>240</sup>, egli presenta una teoria assai interessante ai fini della nostra ricerca. Specificatamente, la sua ipotesi (che va sotto il nome di “shared syntactic integration resource hypothesis” - SSIRH) è che i processi implicati nella rappresentazione delle singole unità e quelli responsabili della elaborazione sintattica degli elementi, ovvero della loro

---

<sup>239</sup> Wong, P., Skoe, E., & Russo, N., *Musical experience shapes human brainstem encoding of linguistic pitch patterns*, “Nature Neuroscience”, 2007, pp. 1-2.

<sup>240</sup> Patel, A., *Language, music, syntax and the brain*, “Nature Neuroscience”, 6, 2003, pp. 674-681.

integrazione in un contesto fraseologico - risiedono in aree cerebrali che seppur operanti all'unisono, sono separate e distinte. Il punto chiave all'interno di tale teoria è che l'area cerebrale dedicata alla integrazione sintattica delle unità è responsabile della processazione tanto delle unità musicali quanto di quelle lessicali. Ciò significa che il meccanismo cognitivo alla base dei processi sintattici condiviso da entrambi i domini è quello detto della "integrazione strutturale" (*structural integration*):

[...] Structural integration is a key part of syntactic processing; that is, mentally connecting each incoming element X to another element Y in the evolving structure [...] Overlap in the syntactic processing of language can thus be conceived of as overlap in the neuronal areas and operations which provide the resources of integration.<sup>241</sup>

Tale area neurale condivisa sarebbe, secondo tale ipotesi, implicata nell'integrazione di elementi (le parole, per le lingue, e le note e gli accordi nella musica) nel corso dell'elaborazione-comprensione di strutture più ampie, ovvero delle proposizioni articolate delle lingue, e delle sequenze armoniche nelle melodie. Patel riconosce infatti che, prescindendo dalla specificità di ciascuna, è possibile ravvisare in entrambi dei principi sintattici, combinatori. Meccanismi che nella lingua, ad esempio,

---

<sup>241</sup> *Ivi*, p. 678.

esprimono la relazione “chi fece cosa a chi”, la cui comprensione si realizza connettendo, o meglio *integrando* ogni parola che andiamo percependo a una parola già percepita, da cui dipende, in funzione della struttura della frase. Ciò è evidente nei seguenti due enunciati, in cui occorrono esattamente le stesse parole, ma il cui ordine differente esprime una diversa connessità sintattica, dunque un diverso significato: (1) “Il ragazzo incontrò Maria che comprò dei fiori”, e (2) “Il ragazzo che incontrò Maria comprò dei fiori”<sup>242</sup>. Nella musica, tale processo di integrazione trova espressione nella percezione delle tonalità e degli accordi, tra i quali occorrono delle relazioni conformi ad una geometria particolare, paragonabile a quella che interessa le parole “dipendenti dalla struttura” delle proposizioni. Così, ciascun accordo (come le parole all’interno di una proposizione) è associato ad un altro in base ad un preciso rapporto strutturale; per dirla con le parole dello stesso Patel, “the processing of chords is influenced by their distance from one another in a structured cognitive space of pitch classes, chords and keys”<sup>243</sup>. Tale tesi viene trasversalmente confermata diverse ricerche condotte sul tema. Attraverso lo studio dell’attivazione di indici neurofisiologici detti “potenziali correlati ad eventi cerebrali” (ERP: *Event-Related Brain*

---

<sup>242</sup> *Ibidem.*

<sup>243</sup> *Ibidem.* Per una rassegna sulle differenze nella elaborazione percettiva degli stimoli linguistici da quelli musicali, cfr. Zatorre, R. J., Belin, P., & Penhune, V. B., *Structure and function of auditory cortex: music and speech*, “Trends in Cognitive Sciences”, 6, 2002, pp. 37–46.

*Potentials*), Besson ed i suoi collaboratori hanno dimostrato, di fatti, che tanto la musica quanto il linguaggio verbale attivano i medesimi indici di fronte alla violazione delle rispettive strutture temporali, o di fronte alla presenza di unità *dissonanti* rispetto al contesto rispettivamente musicale e fraseologico:

Taken together, results have shown that the semantic computations required to access the meaning of words, and their integration within a linguistic context, seem to be specific to language. Indeed, whereas unexpected words within a sentence context are associated with the occurrence of an N400 component, unexpected notes or chords within musical phrases elicit a P600 component. By contrast, words that are unexpected on the basis of the syntactic structure of the sentence, and chords that are unexpected as a function of the harmonic structure of the musical sequence, elicit similar effects in both cases, namely, a P600 component. [...] Finally, violations of temporal structure within language and music also elicit similar effects, a biphasic negative–positive complex, the emitted potential. The occurrence of the emitted potential shows that in both language and music, words and notes or chords are expected at specific moments in time.<sup>244</sup>

L'idea secondo la quale la musica e il linguaggio condividano le risorse neurali dedicate alla processazione strutturale confluisce verso la nostra ipotesi di ricerca, ovvero verso l'assunto che si possa a buon diritto identificare nella musica una tappa evolutiva successiva alla abilità di modulazione della linea prosodica dei suoni, tratto analogo alla abilità di strutturazione sintattica dei canti negli uccelli e nei primati non umani.

---

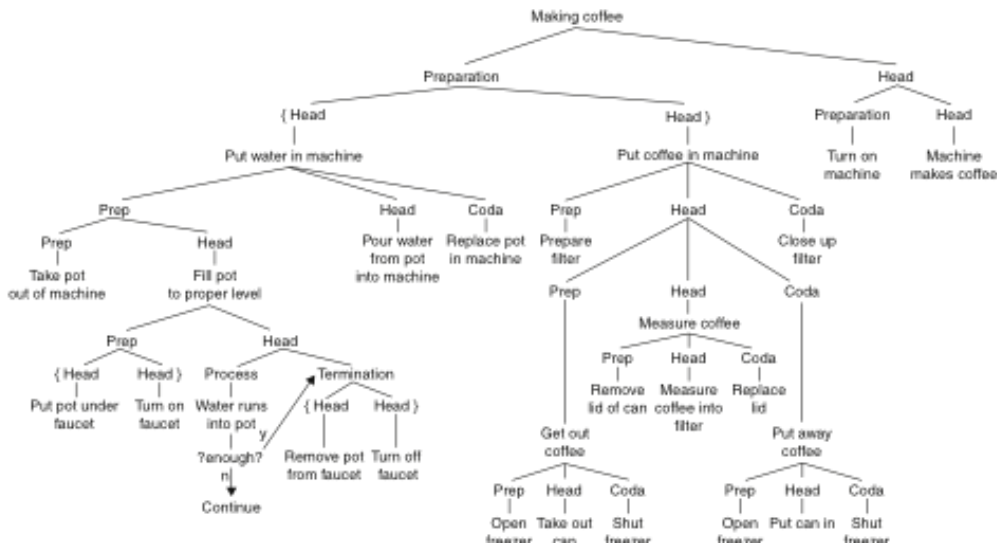
<sup>244</sup> *Ivi*, p. 254.

Pertanto, in tale prospettiva di studio, è bene ribadire che la cellula da cui ha preso avvio l'evoluzione della facoltà di linguaggio è l'abilità generale di *sintassi*, intesa come la capacità di combinare le date unità uditive (o visuali) in base a regole strutturali. Così, tale abilità, come abbiamo visto ravvisabile anche in specie non umane, sarebbe sfociata nella capacità di modulare prosodicamente suoni fonetici, e da lì nell'emergenza di strutture sintattiche più complesse, sistematicamente combinatoriali, ovvero la cui organizzazione sintattica si dispiega gerarchicamente su più livelli. I meccanismi cognitivi implicati in tale capacità vantano una natura dominio-generale, e trovano applicazione in domini distinti - ma evidentemente non separati - quali quello visivo, musicale, linguistico o motorio. In tale direzione si collocano le riflessioni di Jackendoff, secondo il quale la pianificazione di una azione motoria come ad esempio fare il caffè (figura 9), ben rappresenta un parallelo cognitivo della facoltà di linguaggio e di creare musica, in quanto organizzata gerarchicamente e tale che ogni nodo principale è a sua volta ampliabile attraverso l'inclusione di ulteriori azioni entro strutture dette "*prolongational trees*":

Patel [...] presents experimental evidence that the hierarchical structures of language and music, although formally distinct, are integrated by the same part of the brain, roughly Broca's area. If so, I would hazard a guess that complex action structures are too. That this area is usually considered premotor would add some plausibility to this speculation. In fact, the integration and execution of complex action might be a strong candidate for a more general, evolutionarily



older function that could be appropriated by both language and music, quite possibly independently.<sup>245</sup>



**Fig. 9.** Immagine modificata da Jackendoff (2009). Struttura di un'azione complessa come quella del fare il caffè.

Assumiamo che, a partire da tali precondizioni cognitive, il linguaggio proposizionale abbia desunto la sua specifica fisionomia in virtù dell'innesto dell'abilità di sintassi sulla capacità di connettere intenzionalmente le unità foniche a stati cosali reali o meno, ovvero sul valore semantico delle stesse unità (questo innesto è tale che la connessione sintattica degli elementi dipende dal valore semantico degli stessi, e viceversa); tuttavia, è opportuno riconoscere che la connessione delle unità verbali ad un *significato* può prescindere dalla loro organizzazione

<sup>245</sup> Jackendoff, R., *Parallels and Nonparallels between Language and Music*, cit., p. 202.

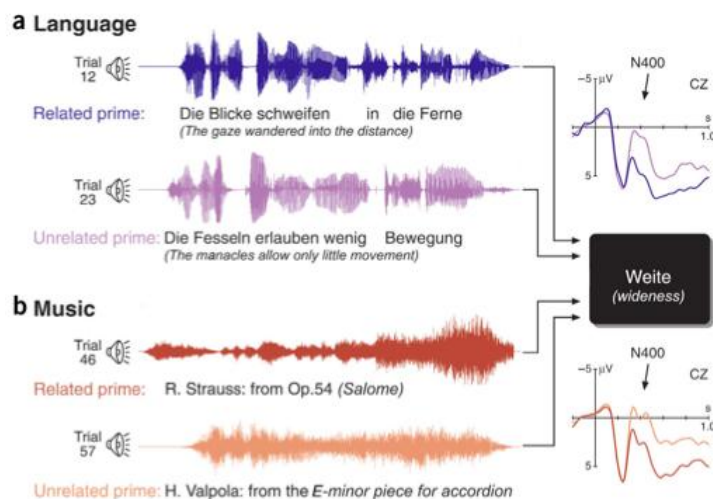
sintattica, giocando sulla portata pragmatica ed emozionale. Detto ciò, vale la pena di notare che lo studio dell'intreccio evolutivo e neuronale tra musica e linguaggio rivela altresì interessanti aspetti sulla capacità di significazione negli esseri umani. Nel paragrafo precedente abbiamo visto come la musica e la prosodia siano in grado di trasferire informazioni di natura emozionale. Ma non è tutto. Da una ricerca condotta da un gruppo di ricercatori guidati da Angela Friederici<sup>246</sup>, emerge che la musica è in grado di trasferire altresì informazioni di natura semantica. Nello specifico, gli studiosi hanno sfruttato un effetto cognitivo detto del *priming* semantico, che consiste nel fatto che la processazione di parole connesse ad un dato contesto semantico - ad esempio "sono stata al ristorante" e "pizza" - è *favorita* a livello neurofisiologico, rispetto a termini semanticamente lontani come potrebbe essere, nel nostro esempio, il termine "ruspa". Il "vantaggio" neurofisiologico è misurato attraverso l'attivazione di un indice fisiologico correlato ad eventi cerebrali, e gli stimoli dell'esperimento consistevano in parole connesse allo stesso tempo ad alcuni contesti fraseologici - in funzione di legami semantici - e a dei brani musicali, in funzione di somiglianze a livello fonosimbolico (ad esempio "uccello" ad un brano i cui suoni ne ricordavano il canto, o "scale" per un

---

<sup>246</sup> Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D., *Music, language and meaning: brain signatures of semantic processing*, "Nature Neuroscience", 7, 2004, pp. 302–307.

brano dalle note ascendenti). Il dato interessante che emerge da tale esperimento è che tanto le frasi linguistiche quanto i brani musicali attivano il medesimo indice fisiologico, l’N400, per la processazione di termini incongrui al contesto semantico. Così gli autori commentano tale risultato:

Our behavioral data indicated that three independent groups of subjects with no special musical training who are naively presented with particular musical excerpts similarly associate the musical excerpts with particular words. The correspondent ERP results indicated that the N400 priming effect in the music domain did not differ from that in the language domain [...]. Thus, the present data show that music cannot only influence the processing of words, but it can also prime representations of meaningful concepts, be they abstract or concrete, independent of the emotional content of these concepts. [...] The present results indicate that music transfers considerably more semantic information than previously believed.<sup>247</sup>



**Fig. 10.** Immagine modificata da Koelsch *et al.* (2004). Dopo essere stati esposti a dei passi verbali (condizione 1) o musicali (condizione 2), i soggetti udivano una parola connessa in qualche modo al passo. In entrambi i casi, venivano attivati i medesimi indici neurofisiologici.

<sup>247</sup> Ivi, pp. 305-306.

Tali ricerche - unitamente agli studi sulla capacità della musica e della prosodia di trasferire informazioni emozionali - mettono in luce la non indifferente portata significazionale dell'espressione, determinata dalle qualità acustiche del segnale stesso - al di là della sua struttura *sintattica*. Tuttavia, è il caso di precisare che nonostante gli aspetti cognitivi condivisi, la musica e il linguaggio verbale sottendono differenze cruciali sul piano semantico-morfologico. Nelle melodie, infatti, dalla natura combinatoria degli elementi costitutivi non risulta, come nel caso delle parole, un significato (riferimento) ben individuabile<sup>248</sup>. Inoltre, in esse manca la proprietà della doppia articolazione, la possibilità di far riferimento in modo univoco ad eventi o oggetti non dispiegati nell'immediato *hic et nunc*, ed infine, l'arbitrarietà della connessione fra significanti e significati, caratteristiche invece distintive delle parole.

Resta così aperta la questione intorno al modo in cui la portata emozionale e pragmatica del segnale stesso possa avere influito sull'evoluzione della struttura sintattica in strutture morfo-sintattiche in grado di trasferire non soltanto emozioni, ma anche, con esse, messaggi semanticamente discreti. Si tratta di un passaggio evolutivo importante, che ci ha permesso di approdare alla nostra forma di vita: un mondo in cui è possibile parlare di fatti reali o meno, persone, oggetti concreti o astratti;

---

<sup>248</sup> Mi riferisco ai principi di composizionalità e di contesto di Gottlob Frege, ben illustrati in Carapezza M., *Segno e simbolo in Wittgenstein*, Bonanno, Acireale-Roma, 2005, pp. 99-114.

ma anche di compiere delle azioni non soltanto rese possibili dallo stesso linguaggio verbale, ma anche comprensibili (se ci va bene!) soltanto all'interno di tale *forma* cognitiva – talvolta potendo persino prescindere dall'uso effettivo ed esplicito delle parole. Tuttavia, resta da chiedersi quale possa essere stata la spinta propulsiva in grado di trainare l'evoluzione di sequenze di suoni prosodicamente modulati verso l'attuale fenomeno del linguaggio proposizionale. Una risposta a tale fondamentale quesito può essere guadagnata guardando ad una caratteristica biologica fondamentale nell'uomo: l'inclinazione a cooperare, condividere emozioni ed informazioni intenzionalmente all'interno di un gruppo sociale. In particolare, la nostra ipotesi è che l'evoluzione di tale intrinseco senso di socialità di cui sipregna il linguaggio proposizionale, abbia acquisito la sua linfa originaria proprio dalla dimensione *prosodico-musicale* del linguaggio stesso. Riteniamo, infatti, che quest'ultima, nel trasmettere informazioni di natura primariamente emozionale e pragmatica a livello gruppale ne abbia alimentato non soltanto la motivazione a cooperare, ma anche la capacità di *parlare*,tenendo in considerazione gli stati mentali altrui. Tratteremo in maniera più estesa queste ultime abilità nel capitolo 10, dopo aver illustrato il ruolo evolutivo della musica nell'ambito dell'emergenza della mente sociale (e del linguaggio).

#### **9.4 Co-evoluzione tra mente sociale e linguaggio: il ruolo della modulazione musicale**

L'universalità della musica, ovvero la sua pervasività nella vita e nelle diverse culture umane la rende un'ottima candidata per lo studio sull'evoluzione della cognizione umana. In particolare, una volta fornita una definizione operativa del concetto di musica sulla base dei tratti musicali condivisi a livello cross-specifico, è opportuno centrare il fuoco dell'analisi su quegli aspetti che, rappresentando dei fondamentali elementi di divergenza dalle forme di musica percepibili e prodotte nelle specie non umane, ben si prestano ad ampliare la comprensione della mente umana. Nel paragrafo precedente, illustrando i punti di analogia cognitiva tra linguaggio e musica, abbiamo messo a tema la capacità degli uomini di riconoscere, ma soprattutto di produrre unità entro strutture sintattiche regolate, abilità che in musica si traduce nella produzione/percezione di suoni a frequenze specifiche, entro scale musicali differenti - ovvero nell'ambito di intervalli determinati (in qualche modo analoga all'abilità di connettere e comprendere le unità *nelle* strutture logico-sintattiche delle proposizioni). Ma non è tutto. Come suggerisce Sandra Trehub in una sua recente pubblicazione sul tema<sup>249</sup>, vi sono due ulteriori aspetti cognitivi

---

<sup>249</sup>Trehub, S. E., *The developmental origins of musicality*, "Nature Neuroscience", 6, 2003, pp. 669-673.

propri della musica che rappresentano importanti aspetti di divergenza cognitiva tra l'uomo e le altre specie animali. Si tratta della capacità di tradurre ed alimentare l'espressività emozionale in contesti sociali (basti pensare, ad esempio al suo uso generalizzato in cerimonie pubbliche, rituali, o feste), e dell'abilità specie-tipica degli umani di sincronizzare il tempo con uno o più soggetti, come avviene ad esempio quando si fa musica o si balla con qualcun'altro. Tali osservazioni sulla specie-specificità delle attività musicali nell'uomo sono un chiaro indice della intrinseca connessione tra musica e socialità, o per essere più precisi, tra l'abilità di sincronizzazione, e la espressione/condivisione di emozioni da una parte, e la loro realizzazione in contesti sociali.

La mia ipotesi è che l'abilità di *sincronizzare* il ritmo con dei soggetti esterni, assieme alla capacità di creare prosodia (ovvero quell'area musicale in cui i suoni, seppur non strutturati sulla base di specifiche scale musicali presentano una modulazione di frequenza, tempo ed intensità), trasferendo informazioni di natura primariamente emozionale e pragmatica - rappresentino il nucleo cognitivo in grado di avviare l'evoluzione della musica, del linguaggio proposizionale, e in modo importante, di alimentare la natura sociale degli esseri umani, determinandone la specie-specificità. Nel seguito del paragrafo prenderemo in esame alcune salienti teorie in

---

merito al valore evolutivo della capacità di fare musica in gruppo. Tuttavia, prima di procedere nella trattazione, per amor di chiarezza metodologica, tengo a precisare preliminarmente che gli studi sperimentali presentati di seguito impiegano stimoli musicali in senso stretto, caratterizzati cioè dall'uso di melodie le cui note sono arrangiate in base a (o violano) specifiche leggi armoniche, e riconoscibili in quanto tali. Così, pur gettando luce su importanti aspetti evolutivi, essi individuano nella capacità del fare “musica” *stricto sensu* uno dei tratti cognitivi in grado di promuovere o segnalare la capacità di cooperare a livello grupale. Tuttavia, riteniamo necessario esser cauti nell'identificare nella “musica” in sé il sistema in grado di catalizzare l'evoluzione della cognizione umana. Piuttosto, troviamo che un sistema valido in tal senso possa essere stata la modulazione prosodica dei suoni, ovvero una modulazione regolata, ma che - come nel caso dei canti degli uccelli - non fa specificamente capo alla grammatica musicale – (tale idea rimane suscettibile di test sperimentali volti ad attestarne l'effettiva plausibilità).

Ciò detto, è opportuno prendere in esame quanto illustrato da tali studi in merito ai vantaggi evolutivi dell'instanziazione della abilità di fare *musica*. Una spiegazione interessante emerge da una ricerca realizzata da Hagen e Bryant dal titolo “Music and Dance as a Coalition System”<sup>250</sup>. Gli

---

<sup>250</sup> Hagen, E., *Music and Dance as a Coalition System*, “Human Nature”, 14, 2003, pp. 21-50.



studiosi hanno avanzato l'ipotesi che la musica e la danza siano nate come strategie evolutive per segnalare il livello di coesione interna di un gruppo, e di conseguenza l'abilità dello stesso gruppo di difendere il proprio territorio e segnalare la propria identità<sup>251</sup>. Per testare tale assunto, gli studiosi hanno utilizzato come stimoli dell'esperimento tre tipi di musica, in ciascuna delle quali hanno arrangiato le parti di quattro diversi strumenti musicali. I pezzi musicali presentavano livelli di sincronia diversi, ed i soggetti erano invitati a giudicare la qualità della coalizione fra i musicisti sulla base del mero ascolto del brano musicale. Così gli autori commentano tale risultato:

A connection also exists between the coalition quality hypothesis and the "social cohesion" hypothesis of music function. Although we are proposing that music and dance signal social cohesion rather than cause or promote social cohesion, members of coalitions who perceive that continued membership in the coalition will yield significant benefits should be motivated to practice singing and dancing together. Coalition members should enjoy singing and dancing, even when other coalitions are not present, because they need to develop and master their musical repertoire. Further, once individuals have demonstrated, by other means, their ability to provide valuable benefits or that they share important interests with other group members, their subsequent costly effort to master a group's musical repertoire is a credible signal of their loyalty to the group because it significantly increases the costs of switching groups, enhancing group stability [...].<sup>252</sup>

---

<sup>251</sup> *Ivi*, p. 27: "We believe that for humans and human ancestors, musical displays may have similarly functioned, in part, to defend territory (and perhaps also to signal group identity), and that these displays may have formed the evolutionary basis for the musical abilities of modern humans. [...] Territorial defense and alliance formation both require communicating credible information about group capabilities to non-group members, information that would deter intruders but attract allies. We will now outline the hypothesis that music and dance evolved to credibly communicate such information."

<sup>252</sup> *Ivi*, p. 41.

L'abilità di sincronizzare la produzione dei suoni, ancora oggi tratto distintivo dell'abilità di fare musica nella specie umana, si sarebbe evoluta, secondo tale prospettiva, a fronte della necessità di difendere il territorio segnalando l'abilità di agire collettivamente, e motivando allo stesso tempo i membri a non lasciare il gruppo. Tuttavia, è bene notare che oltre a segnalare la presenza di una efficace coalizione, la produzione della musica può essere la stessa causa della propensione alla cooperazione. Tale assunto è quanto emerge da una ricerca recentemente condotta dal gruppo di Micheal Tomasello, illustrata in un articolo dal titolo "Joint music making promotes prosocial behavior in 4-year-old children"<sup>253</sup>. Nello specifico i ricercatori, hanno condotto un esperimento su bambini dell'età di quattro anni, dimostrando che la partecipazione ad un gioco collettivo in cui sono chiamati ad eseguire azioni ritmicamente sincronizzate e al contempo cantare ritornelli melodici in coro promuove l'inclinazione a cooperare con i compagni, perseguendo un fine comune. Gli autori descrivono tale studio in un passo che vale la pena di proporre nella sua lunghezza:

Thus, we hypothesize that joint music making, as in our study, not only creates a higher level of coordination compared to other non-musical group activities, but also encourages the participants to maintain a constant audiovisual representation of the collective intention and shared goal of vocalizing and moving together in time, thereby strengthening their sense of acting together as a unit. The intrinsic desire to share emotions, experiences and activities with others clearly separates

---

<sup>253</sup> Kirschner, S., Tomasello, M., *Joint music making promotes prosocial behavior in 4-year-old children*, "Evolution and Human Behavior", 31, 2010, pp. 354–364.

humans from other primates. It appears very early in human ontogeny and enables children to participate in cooperative activities involving joint intentions and joint attention from early infancy onwards [...]. Understanding music as a collectively intended activity — with specially designed features that satisfy this human desire to share emotions, experiences and activities with others — might explain why the children in our study felt a stronger commitment after joint music making and so spontaneously helped or cooperated with each other. Compared to language which is most effective in mobilizing joint intentionality for goal-directed behavior, music might be more efficacious in mobilizing joint intentionality per se — in the sense of feeling a “we” unit, thus getting people to experience each other as co-active, similar and cooperative members of a group.<sup>254</sup>

Estendendo l'applicazione di tale studio sulla nostra teoria, possiamo ipotizzare che nell'ambito della generale abilità di produrre musica, il ruolo chiave nell'emergenza del linguaggio sia stato giocato dall'abilità di sincronizzare il ritmo e la modulazione delle frequenze in un contesto grupppale, favorendo in qualche modo l'istanziamento di un contesto emozionale condiviso. Ciò avrebbe creato la condizione adatta alla cooperazione finalizzata ad uno scopo comune, in altri termini, alla esecuzione di azioni in un contesto di intenzionalità condivisa. Tali dati si collocano sulla medesima linea teorica della ipotesi avanzata da Livingstone e Thompson in “The Emergence of music from the Theory of Mind”<sup>255</sup>, secondo i quali la stessa abilità di fare musica trova le sue radici evolutive nelle predisposizioni di origine biologica dell'empatia e della teoria della mente. In particolare, gli autori sostengono che a partire da una

---

<sup>254</sup> *Ivi*, p. 366.

<sup>255</sup> Livingstone, R. S., Thompson, W. F., *The emergence of music from the Theory of Mind*, “*Musicae Scientiae*”, 2009, pp. 83-115.

sorta di risonanza emozionale ed affettiva occorrente a livello neurale (definita *affective engagement* dagli autori), e amplificata dall'impiego di canali multimodali (visivo, motorio, auditivo) – in altre parole, a partire dall'abilità del co-sentire, l'uomo ha sviluppato l'abilità di fare musica in un contesto collettivo, mentre l'evoluzione della capacità del rispondere intenzionalmente a qualcuno, ovvero di saperne interpretare e modificare gli stati mentali ed emozionali avrebbe gettato le basi per la sua propagazione e diversificazione a livello culturale:

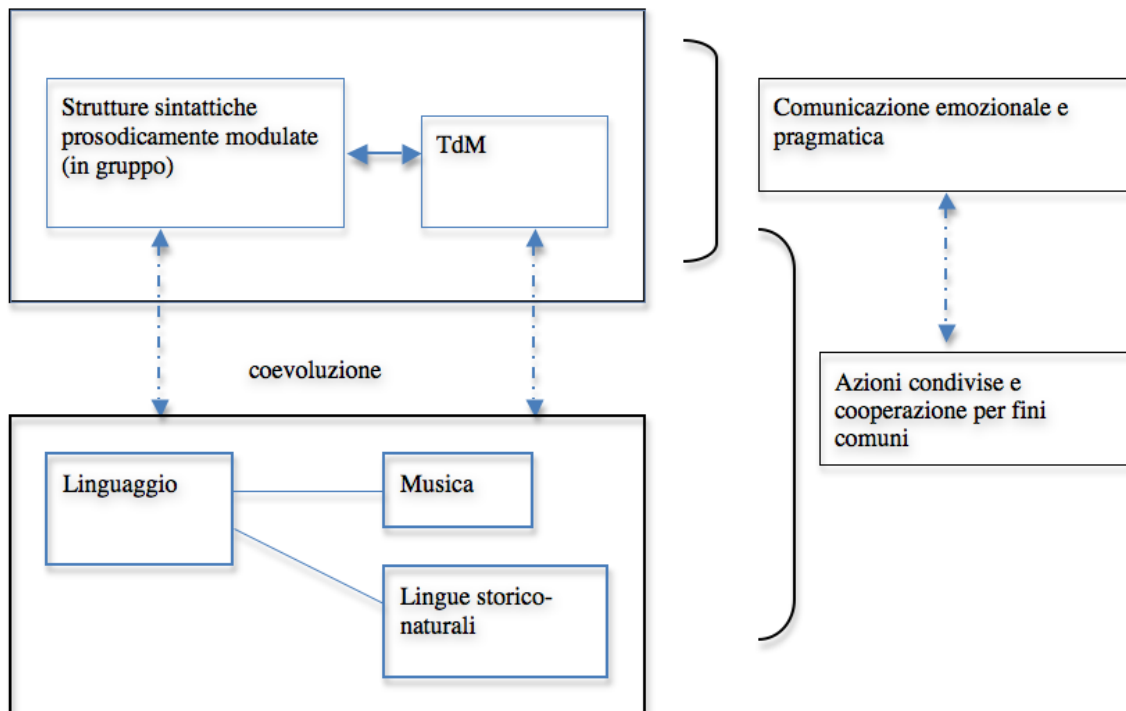
ToM allows an individual to recognise the mental and emotional state of conspecifics, and is pivotal in the cultural transmission of knowledge. We propose that a specific form of ToM, Affective Engagement, provides the foundation for the emergence of music. Underpinned by the mirror neuron system of empathy and imitation, music achieves engagement by drawing from pre-existing functions across multiple modalities. As a multimodal phenomenon, music generates an emotional experience through the broadened activation of channels that are to be empathically matched by the audio-visual mirror neuron system.<sup>256</sup>

Sebbene trovi assai rilevante quanto sollevato dagli autori in merito al ruolo del sistema biologico alla base dell'*affective engagement*, crediamo che tale ipotesi meriti ulteriori approfondimenti volti a stabilire con maggiore precisione la natura della relazione evolutiva tra musica, o più precisamente, della modulazione prosodica dei suoni e la teoria della mente. Crediamo, infatti, che non sia del tutto verosimile uno scenario evolutivo causalmente lineare in cui una volta poste in essere le basi

---

<sup>256</sup> *Ivi*, p. 83.

biologiche del co-sentire e la teoria della mente sia emersa *ipso facto* la capacità di creare e trasmettere musica. La nostra idea è infatti che la teoria della mente e l'empatia da un lato e la facoltà di modulare suoni in gruppo stiano piuttosto in un rapporto di coevoluzione, un rapporto, cioè, in cui l'evoluzione dell'uno risuonava sull'altro, trainandone la progressione verso l'attuale configurazione. In altri termini, l'abilità di modulare sequenze di suoni in gruppo e secondo pattern sintattici specifici avrebbe fondato l'emergenza e l'evoluzione da una parte, di principi sintatticologici in funzione dei quali strutturare sistematicamente il segnale - e dall'altra di ordini sempre più elevati di teoria della mente, risultando in domini in qualche modo analoghi ma diversi, quali la musica e il linguaggio proposizionale. Entrambi i domini infatti sottendono delle grammatiche sintattiche in funzione delle quali organizzare sistematicamente le unità costitutive: le leggi armoniche nel caso della musica, e le regole morfosintattiche nel caso delle lingue storico-naturali (figura 11).



**Fig. 11.** L'abilità di modulare prosodicamente pattern di suoni in gruppo si è evoluta secondo un rapporto di reciproca retroazione a) nella capacità di modulare elementi secondo una vera e propria grammatica (linguaggio), b) in ordini sempre più complessi di teoria della mente.

La modulazione prosodica dei suoni intensificava altresì la coloritura emozionale e pragmatica della comunicazione, co-evolvendo sull'abilità di cooperare per fini condivisi (legata al possesso di una complessa teoria della mente). Da ciò, è scaturita l'emergenza delle lingue storico-naturali e della musica.

E proprio nell'abilità di acquisire combinazioni *sintattiche* gerarchicamente organizzate identifico la capacità di linguaggio specie-tipica umana, un'abilità i cui diversi strati, ravvisabili anche in specie non umane, nell'uomo costituiscono un'unica entità cognitiva evolutasi a fronte di una necessità vitale che ancora oggi ne determina la divergenza dalle altre specie animali: la ipertrofica necessità di condividere con gli altri pensieri, sentimenti, o esperienze, comunicandole nel modo più efficace

possibile. Su tale scia teorica si collocano coerentemente le riflessioni di Rousseau in merito all'origine delle lingue:

Da che cosa può dunque derivare tale origine? Dai bisogni morali, dalle passioni. Tutte le passioni avvicinano gli uomini che la necessità di cercare i mezzi di sussistenza obbliga a sfuggirsi. Non è stata la fame, neppure la sete, ma l'amore, l'odio, la pietà, la collera, che hanno sempre strappato loro le prime voci. I frutti non si sottraggono alle nostre mani, possiamo nutrircene senza parlare; possiamo inseguire in silenzio la preda di cui vogliamo cibarci: ma per suscitare emozioni in un giovane cuore, per respingere un aggressore ingiusto, la natura detta accenti, grida, lamenti. In questo modo furono inventate le più antiche parole, ed è per questo motivo che i primi linguaggi furono sonori [*chantantes*] e dettati dalla passione prima di essere semplici e metodici. [...]

La melodia, imitando le inflessioni della voce, esprime i pianti, le grida di dolore e di gioia, le minacce, i gemiti; tutti i segni delle passioni ne sono la molla. Essa imita gli accenti dei linguaggi [...]; *essanon imita soltanto, parla.*<sup>257</sup>

La melodia, ovvero la combinazione di suoni privi di un riferimento preciso ed univoco, ma ciò nondimeno dotata di un certo potere espressivo, sarebbe stata, dunque, secondo lo studioso, la prima forma di linguaggio articolato adottato dagli uomini, nato direttamente dall'intima esigenza di tradurre agli altri (e a se stessi) le proprie emozioni, di dividerle pubblicamente, e permetterne un più preciso riconoscimento. E nel soddisfare tale esigenza espressiva, la modulazione prosodica delle

---

<sup>257</sup> Rousseau J. J., *Essai sur l'origine des langues*, Gallimard, Paris; trad. it. di G. Gentile, *Saggio sull'origine delle lingue*, Guida, Napoli, 1984, pp. 52, 107 (corsivo mio).

espressioni avrebbe *ipso facto* alimentato anche la coesione all'interno dei gruppi.

È opportuno sottolineare altresì che la necessità di comunicazione, o meglio di comunione emozionale, oltre a determinare l'evoluzione del linguaggio sotteso alle lingue e alla musica a partire dal nucleo cognitivo della prosodia, ne veniva al contempo alimentata, in un rapporto di causalità circolare.

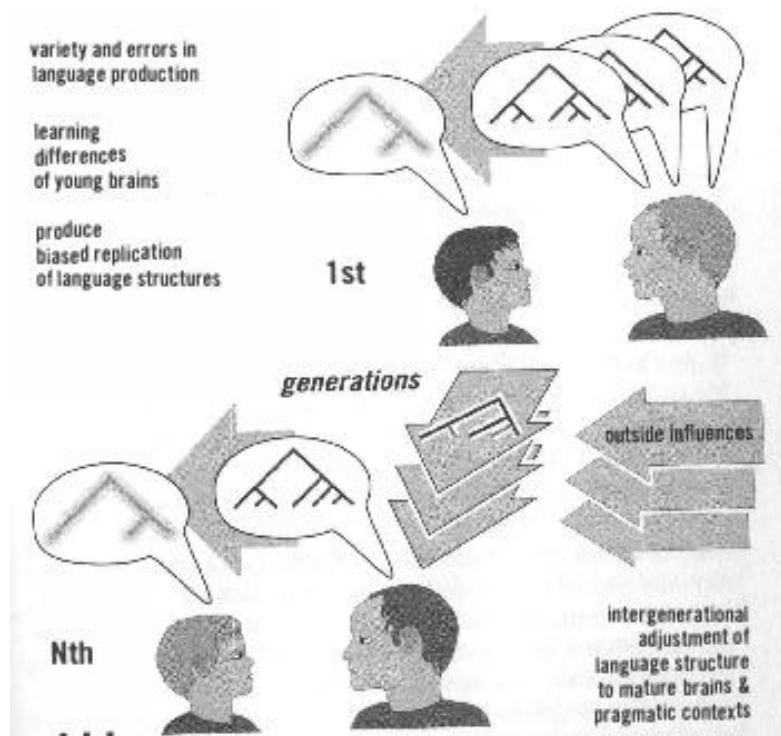
Tale ipotesi si colloca coerentemente nella prospettiva di studio delineata da Terrace Deacon, secondo il quale il *linguaggio* sotteso alla diversità delle lingue storico-naturali – è emerso come entità di natura primariamente sociale, la cui evoluzione, cioè, è stata dettata non da leggi fisiche o matematiche, ma dalla possibilità di essere riprodotte e trasmesse con crescente facilità (figura 12). E, riprendendo le idee baldwiniane, lo studioso sostiene che proprio tale specifica esigenza cognitiva avrebbe guidato la selezione non soltanto di pattern sintattici efficaci in tal senso, ma anche la riorganizzazione delle reti neurali nel nostro cervello, secondo una complessa dinamica in cui l'evoluzione dei geni e dei comportamenti sociali si rinforzano mutuamente, selezionando quei tratti in grado di garantire maggiori opportunità di sopravvivenza alla specie (figura 13).

The world's languages evolved spontaneously. They were not designed. If we conceive of them as though they were invented systems of rules and symbols,



intentionally assembled to form logical systems, then we are apt either to assign utility and purpose where there is none, or else to interpret as idiosyncratic or inelegant that for which we cannot recognize a design principle. But languages are far more like living organisms than like mathematical proofs. The most basic principle guiding their design is not communicative utility but reproduction-theirs and ours. So, the proper tool for analyzing language structure may not be to discover how best to model them as axiomatic rule systems but rather to study them the way we study organism structure: in evolutionary terms. Languages are social and cultural entities that have evolved with respect to the forces of selection imposed by human users.

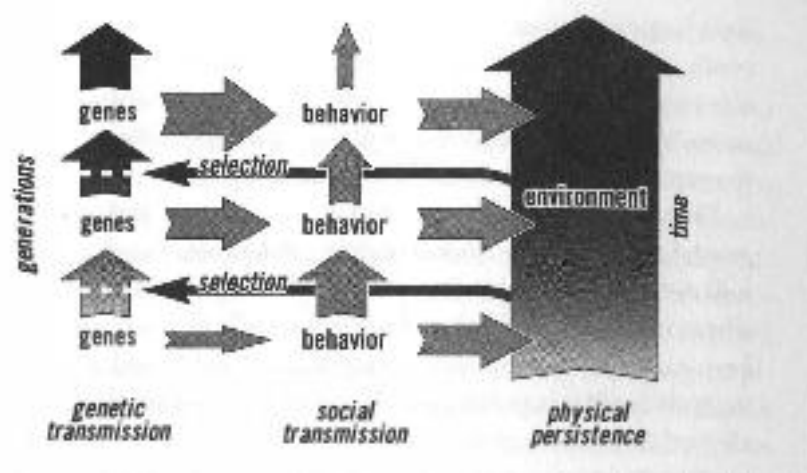
[...] the computational demands of symbolization not only are the major source of the selection pressures that could have produced the peculiar restructuring of our brains, they are likely also the indirect source for the selection pressures that initiated and drove the prolonged evolution of an entire suite of capacities and propensities that now constitute our language "instinct."<sup>258</sup>



**Fig. 12.** Immagine modificata da Deacon (1997). Specifici vincoli cognitivi determinati dall'apprendimento per associazione e dai limiti imposti dalla memoria a breve termine costituiscono una sorta di "filtro" naturale a cui le lingue

<sup>258</sup> Deacon T., W., *The Symbolic Specie. The coevolution of Language and the Brain*, Norton & Company, New York, 1997, pp. 110, 340.

sono soggette. Ciò fa sì che i bambini percepiscano selettivamente certe strutture in favore di altre, spingendo la selezione di quei pattern “*child-friendly*” (facili da ricordare).



**Fig. 13.** Immagine modificata da Deacon (1997). Schema del processo alla base della selezione baldwiniana.

- Le frecce verso l'alto indicano tre simultanei processi di trasmissione: genetico, sociale (per apprendimento), per i cambiamenti fisici nell'ambiente prodotti dai cambiamenti comportamentali.
- Le frecce a destra indicano l'influenza dei geni sui comportamenti.
- Le frecce a sinistra indicano gli effetti della selezione sui geni.
- Le frecce relative alla selezione sociale diventano sempre più sottili, ad indicare il ruolo sempre più tenue dell'apprendimento in favore di un crescente influsso genetico sul comportamento stesso (segnalato dalle frecce sempre più marcate dai geni ai comportamenti)

Nel prossimo capitolo tratteremo più estesamente la tesi dell'origine primariamente sociale del linguaggio. Tuttavia, prima di affrontare tale argomento è opportuno illustrare con maggiore precisione un tema assai dibattuto nell'attuale studio sull'evoluzione del linguaggio, inerente l'eventuale valore adattativo della musica per la specie umana.

## 9.5 Musica: adattamento o dessert?

Pur essendo una delle pratiche più diffuse, e nonostante il suo ruolo centrale nelle culture umane, ad oggi le origini e la funzione evolutiva della musica è oggetto di intensi dibattiti che vedono contrapporsi posizioni assai diverse fra loro. Molti studiosi sostengono che non sia corretto identificare nella musica un adattamento biologico in grado di favorire la sopravvivenza della specie umana, ma piuttosto un sorta di effetto collaterale dell'evoluzione delle componenti cognitive del linguaggio stesso, che al pari della percezione del buon gusto di un dessert manca di ogni vantaggio biologicamente adattativo. Così ne parla Pinker, uno dei massimi sostenitori di tale prospettiva:

We enjoy strawberry cheesecake, but not because we evolved a taste for it. We evolved circuits that gave us trickles of enjoyment from the sweet taste of ripe fruit, the creamy mouth feel of fats and oils from nuts and meat, and the coolness of fresh water. Cheesecake packs a sensual wallop unlike anything in the natural world because it is a brew of megadoses of agreeable stimuli which we concocted for the express purpose of pressing our pleasure buttons.

But if music confers no survival advantage, where does it come from and why does it work? I suspect that music is auditory cheesecake, an exquisite confection crafted to tickle the sensitive spots of at least six of our mental faculties.<sup>259</sup>

Fra le sei facoltà cui l'autore si riferisce vale la pena di menzionare la

---

<sup>259</sup> Pinker, S., *How the Mind works*, Penguin Books, New York, 1997, pp. 524-525, 534.

componente prosodica del linguaggio, il colore emozionale delle espressioni, l'abilità di controllo motorio, il movimento ritmico.

All'altro capo della discussione si collocano i difensori della tesi secondo la quale la musica, al contrario, vanta un fondamentale ruolo adattativo. Tale prospettiva si suddivide a sua volta in tre ordini diversi di spiegazione a seconda che si intenda la musica come tratto evolutosi sulla base a) della selezione sessuale, come indicatore di buone qualità cognitive quali memoria, capacità di acquisire la padronanza di strutture complesse<sup>260</sup>, o di coordinazione motoria, b) della selezione parentale: la modulazione prosodica delle espressioni avrebbe apportato dei vantaggi selettivi in funzione dei suoi effetti emozionali e cognitivi nell'ambito delle interazioni fra gli infanti e i *caregiver*<sup>261</sup> e fra individui geneticamente

---

<sup>260</sup> Vedi Fitch, W. T., *On the Biology and Evolution of Music*, "Music Perception", 24, 2006, p. 87: "Thus, whether we call it "animal music" or not, an important lesson from other species possessing complex, learned vocalization is that attempts to separate music into "cultural" and "biological" components will be futile. These are hopelessly entangled aspects of human biology. It is precisely the human ability to acquire complex novel aspects of our environment (acoustic and otherwise) that constitutes a core adaptation of our species."

<sup>261</sup> A tale proposito, vedi Falk, D., *Prelinguistic evolution in early hominins: Whence motherese?*, "Behavioral and Brain Sciences", 27, 2004, p. 492: "Around the time of the australopithecine/early Homo transition, maternal pelvises that had been modified to accommodate bipedalism became subject to an emerging trend for increasingly large brains [...] which eventually caused a selective shift toward females that gave birth to relatively helpless infants [...]. Consequently, the ability of babies to cling actively to their mothers was lost in hominins [...]. Similar to some anthropoid mothers that live under difficult foraging circumstances [...], these mothers are hypothesized to have adopted postnatal foraging-related changes in maternal care, which included periodically putting their infants down beside them in order to obtain and process food. As a result, the incidence of distal mother-infant gestural communications increased [...] and prosodic (affective) vocalizations became ubiquitous to compensate for the reduction in sustained mother-infant physical contact."; vedi anche: Fitch, W. T., *Evolving Meaning: The roles of Kin Selection, Allomothering, and Paternal Care in Language Evolution*, in C. Lyon, C. L. Nehaniv, A. Cangelosi (ed), "Emergence of Communication and Language", Springer-Verlag, 2007, pp. 29-52; Fitch, W. T., *The evolution of language: a comparative review*, "Biology and Philosophy", 20, 2005, pp. 193-230; Fitch, T. W., *Kin selection and "Mother Tongues": A Neglected Component in Language Evolution*, in D. K. Oller, U. Griebel (ed.), "Evolution of Communication Systems: A Comparative Approach", MIT Press, Cambridge, Massachusetts, 2004, pp. 276-96.

imparentati<sup>262</sup> - o c) della selezione di gruppo, in quanto segnale della presenza di una buona coalizione, funzionale dunque ad una buona difesa del territorio.

Infine, una terza prospettiva tra le teorie sul valore evolutivo della musica che vale la pena di illustrare è quella avanzata da Patel, secondo il quale la musica non è un'abilità evolutasi *ad hoc* per favorire la sopravvivenza dell'uomo, alla stregua di una cheesecake, un tratto secondario derivante dalla presenza di altre componenti evolutivamente salienti. Secondo l'autore, la musica è un'abilità in cui al pari della capacità del leggere o dello scrivere - non è identificabile un tratto adattativo, ma che costituisce, ciononostante, un'abilità-invenzione che ha letteralmente rivoluzionato la mente dell'uomo e le sue capacità, in funzione della sua portata emozionale, sociale e cognitiva:

Neuroscientific research suggests that music is an invention that builds on diverse, pre-existing brain functions, rather than a trait that originated via processes of natural selection. This is consistent with Pinker's thesis. However, growing evidence from neuroscience also suggests that music is biologically powerful, meaning that it can have lasting effects on nonmusical abilities (such as language or attention) during the lifetime of individual humans. Importantly, these effects can be observed not only in trained musicians but also in ordinary individuals who engage regularly with music. Thus I believe that music should

---

<sup>262</sup> Teniamo presente che è possibile garantire la riproduzione dei geni adatti alla sopravvivenza in maniera diretta, per mezzo della riproduzione sessuale, ed indirettamente, favorendo la sopravvivenza, aiutando cioè quegli individui che portano geni simili: i parenti, e la prole dei parenti. Ora, posto ciò, l'idea è che tale collaborazione fra consanguinei finalizzata alla sopravvivenza, e alla perpetuazione dei geni comuni, è potenziata dalla possibilità di comunicare circa la presenza di cibo o di pericoli.

be regarded as a biologically powerful human invention or “transformative technology of the mind.” [...]

[...] exaptation is not a specific enough term to capture the idea of a transformative technology. This is because exaptation (a term coined before our modern understanding of neural plasticity) does not connote the power of a novel trait to shape the biological systems from which it arose [...]. Furthermore, exaptation allows the notion of secondary adaptation (as in the feather example above), whereas TTM [transformative technology of the mind] theory holds that that there has been no evolutionary modification of the brain aimed at supporting musical behavior.<sup>263</sup>

Crediamo che la diversità di tali posizioni teoriche possa essere in qualche modo risolta entro un quadro di spiegazione coerente soltanto se si ricorre ad una definizione unitaria del concetto di musica. La nozione guadagnata nel presente lavoro di ricerca è particolarmente funzionale in tal senso, soprattutto per il fatto che sottende un approccio comparativo operante a livello sia cross-culturale che cross-specifico. Così, sulla base delle riflessioni fin qui esposte, assumiamo che la musica intesa in senso stretto non sia un adattamento biologico, ma semplicemente l'espressione della più generale abilità biologicamente adattativa di acquisire e produrre espressioni gerarchicamente organizzati su più piani ed in base a specifiche regole grammaticali. Tuttavia, con Patel, vi identifico un'abilità cognitiva che, al pari della lettura o della scrittura, è in grado di rivoluzionare l'assetto neurofunzionale della mente stessa, pur non facendo capo ad aree

---

<sup>263</sup> Patel, A., *Music, biological evolution, and the brain*, in M. Bailar (Ed.), “Emerging Disciplines”, Rice University Press, Houston, 2010, pp. 92, 130.

del cervello specificamente evolute per l'attuazione di comportamenti musicali. Consideriamo invece l'abilità di modulare e riconoscere la linea intonazionale di un'espressione musicale o linguistica, che comprende in sé informazioni legate al ritmo, alla frequenza, alla intensità del segnale - unitamente all'aspetto socio-emozionale di cui tali espressioni si pregnano. Tale tratto cognitivo, al contrario della musica intesa in senso stretto, presenta un alto valore evolutivo in grado, a nostro avviso, di far fronte alle pressioni selettive sia in ambito sessuale, che parentale e di gruppo. Per quanto riguarda il potenziale di attrazione sessuale, infatti, i dati comparativi sulle specie non umane - *in primis* sugli uccelli canori - rivelano che l'abilità di modulare specifiche frequenze seguendo pattern sintattici è un saliente tratto sessualmente selettivo. In secondo luogo, l'abilità di modulare prosodicamente il segnale, come abbiamo ampiamente illustrato nel corso degli ultimi due capitoli ha dei significativi effetti sul piano della acquisizione del linguaggio e dello sviluppo della interazione sul piano emozionale e pragmatico, effetto che non si annulla se si estende la prospettiva dalla diade infante-*caregiver* alla cerchia parentale o del gruppo sociale di appartenenza. Pertanto, la mia assunzione di fondo è che questi tre livelli di spiegazione, in conflitto se riferiti al concetto di musica in senso stretto, possono invece guadagnare una collocazione armonica entro il medesimo paradigma di spiegazione se se ne individua l'oggetto di

riferimento nella modulazione prosodica del segnale. Tale operazione consentirebbe inoltre di render conto di taluni problemi teorici altrimenti emergenti dall'assunzione di una delle teorie di spiegazione a discapito delle altre; ad esempio, l'adozione della tesi della selezione sessuale non renderebbe pienamente conto della natura dimorfica dell'abilità del fare musica, non necessariamente legata alla maturità sessuale o ad attività esclusivamente legate al corteggiamento, o alla difesa del territorio o della coppia stessa da attacchi esterni. Al contrario, la tesi della selezione parentale o di gruppo ne spiegherebbe l'adozione in attività di carattere emotivo, istruttivo, interrelazionale, o ludiche, ma lascerebbe irrisolta la sua applicazione in contesti decisivi per la propagazione della specie, ovvero nel contesto della conquista del partner sessuale. Identifico pertanto nella capacità di modulare prosodicamente i segnali e le azioni motorie la componente cognitiva che, sussumibile nei tre diversi ambiti di spiegazione evolutivi - ha in qualche modo catalizzato l'evoluzione del linguaggio. Tale processo, è bene sottolinearlo, vede nella dimensione grupale la sede naturale di quella necessità che ci rende specifici: la necessità del condividere informazioni, pensieri, emozioni o esperienze, comunicandole. Sarà questo l'oggetto di studio del prossimo capitolo.



### SEZIONE III:

## LA MATRICE SOCIALE DEL LINGUAGGIO

Per questo 'seguire la regola' è una prassi. E *credere* di seguire la regola non è seguire la regola. E perciò non si può seguire una regola '*privatim*': altrimenti credere di seguire la regola sarebbe la stessa cosa che seguire la regola.

Il modo di comportarsi comune agli uomini è il sistema di riferimento mediante il quale interpretiamo una lingua che ci è sconosciuta.

L. WITTGENSTEIN

## **10. Le origini sociali della cognizione umana**

Nel tentativo di individuare quei processi cognitivi che nel fondare l'origine del linguaggio proposizionale determinano l'unicità della specie umana, nel corso del presente lavoro di ricerca abbiamo messo a tema due processi cognitivi di rilevanza strutturale, su cui il linguaggio stesso si impernia: la sintassi e la modulazione prosodica. Tuttavia, basta assistere alla più semplice e comune delle scene di vita quotidiana, come versare da bere ad un amico, o il semplice salutare per capire che la cognizione umana non si risolve in nessuna delle suddette abilità. Un po' come l'efficienza degli organi di un corpo dipende dalle attività vitali di base come il respiro e il battito cardiaco, infatti, la stessa funzionalità di tali tratti si regge sull'aspetto chiave della cognizione umana: l'intrinseca abilità di scambiare atti comunicativi, tenendo conto degli stati mentali altrui, abilità pressoché

assente nel resto del mondo animale, eccezion fatta per le scimmie antropomorfe, in cui si presenta in forme assai semplificate. Come vedremo, nell'uomo tale aspetto cognitivo si traduce in distintive facoltà specie-specifiche come ad esempio quella del cooperare, di attuare rituali, o del fare gossip, in altri termini, in attività prettamente legate alla vita di gruppo. Nei primati non umani, al contrario, lo scenario cognitivo presenta tratti assai differenti. Nel prossimo paragrafo vedremo più da vicino in che senso.

## **10.1 La comunicazione “intenzionale” dei primati**

Nonostante la condivisione di gran parte del patrimonio genetico, l'uomo e i primati non umani presentano facoltà cognitive assai diverse. Uno dei maggiori tratti di divergenza consiste nella abilità di percepire gli stati mentali degli altri, in termini di scopi, desideri, credenze o volizioni. Come osserva Dennett (relativamente alla cognizione dei cercopitechi), “una ragione [...] di essere interessati all'interpretazione intenzionale [...] è che essa promette di rispondere – o almeno di aiutare a rispondere – alle domande: Questo comportamento è veramente linguistico? Stanno

veramente comunicando?”<sup>264</sup>. Nel tentativo di trovare una risposta a tali quesiti, vale la pena di chiedersi anzitutto cosa accade nella mente delle scimmie quando osservano qualcuno compiere un’azione. Sono in grado di comprenderne lo scopo ad esso soggiacente, o si limitano alla semplice determinazione degli effetti consequenziali, prescindendo dalla comprensione dello scopo dell’agente? Per rispondere a tale quesito, riportiamo un esperimento condotto da Nagell, Olguin, e lo stesso Tomasello (1993), che coinvolge nello specifico scimpanzé e bambini:

[...] degli scimpanzé e dei bambini di due anni erano posti davanti a un attrezzo simile a un rastrello e a un oggetto fuori della loro portata. L’attrezzo poteva essere usato in due modi che permettevano a entrambi di raggiungere l’oggetto. Per ciascuna delle due specie, un gruppo di soggetti osservava un dimostratore usare l’attrezzo nel primo modo (meno efficiente), mentre un altro gruppo di soggetti osservava il secondo modo (più efficiente) di usare l’attrezzo. Il risultato era che mentre i bambini in genere copiavano il metodo del dimostratore in ciascuna delle due condizioni di osservazione (apprendimento per imitazione), gli scimpanzé facevano tutta una serie di tentativi per raggiungere l’oggetto e facevano lo stesso tipo di cose quale che fosse la dimostrazione osservata (apprendimento per emulazione).<sup>265</sup>

Tale esperimento porta alla luce un aspetto importante della cognizione dei primati non umani, i quali, di fronte alla scena dell’esperimento isolano due eventi legati da un rapporto di

---

<sup>264</sup> Dennett, D. C., *The Intentional Stance*, MIT press, Cambridge, 1987; trad. it. Di E. Bassato, *L’atteggiamento Intenzionale*, Il Mulino, Bologna, 1993, p. 328.

<sup>265</sup> Tomasello, M., *The cultural origins of human cognition*, Harvard University Press, Cambridge, Mass, 1999; trad. it. di M. Ricucci, *Le origini culturali della cognizione umana*, Il Mulino, Bologna, 2005, pp. 48-49.

consequenzialità: un dimostratore muove un rastrello nelle vicinanze di un oggetto; l'oggetto cade. La loro percezione non è sottile come quella dei bambini, i quali, della medesima scena distinguono molti più elementi di non immediata percezione. Questi ultimi, difatti, sono in grado di connettere quei movimenti che provocano la caduta dell'oggetto ad un'intenzione specifica del dimostratore, ovvero ad uno scopo. Tale comprensione riorganizza la percezione della scena, quindi la sua stessa riproduzione da parte del bambino che, associando quei gesti ad uno specifico scopo, è in grado di imitare il dimostratore *nell'*ottenere quel determinato risultato.

In altri termini, la comprensione del senso intenzionale di quelle azioni ne determina la possibilità di imitazione e ciò rinvia *ipso facto* alla capacità di isolare lo scopo dalle diverse strategie comportamentali che, finalizzate ad esso, lo realizzano – quindi, di converso, alla possibilità di attuare diverse modalità d'azione per raggiungere il medesimo risultato. Al contrario, le grandi scimmie *in questo genere di situazioni* non sono in grado di afferrare l'intenzione del dimostratore (intesa come “forza mediatrice”), e di conseguenza si limitano a riprodurre, ovvero ad emulare dei movimenti *casuali* con il rastrello, nell'aspettativa che l'oggetto cada. Le grandi scimmie, in un certo senso, non “vedono” che il dimostratore sta tentando di afferrare un oggetto, ma semplicemente che sta compiendo

alcuni movimenti particolari, sortendo un dato effetto nell'ambiente.

Tomasello, infatti, afferma:

Nel caso degli esseri umani, lo scopo o l'intenzione del dimostratore è parte essenziale di quel che viene percepito, e in effetti lo scopo appare come qualcosa di distinto dai mezzi comportamentali che possono essere usati per raggiungerlo. [...] Incapaci di separare nelle azioni degli altri lo scopo dai mezzi comportamentali, gli scimpanzé si concentrano sui mutamenti di stato (compresi i mutamenti di posizione spaziale) degli oggetti implicati nella dimostrazione, dove le azioni del dimostratore sono solo un movimento fisico fra i tanti. *Gli stati intenzionali del dimostratore e, di conseguenza, i suoi metodi in quanto entità comportamentali a sé stanti, non sono parte della loro esperienza.*<sup>266</sup>

Ma fino a che punto è corretto affermare che gli scimpanzé non comprendono gli stati intenzionali del dimostratore? Non sarebbe, piuttosto, opportuno chiedersi se essi, in realtà, non siano in possesso di una specie-specifica modalità di percezione degli stati intenzionali altrui, che dall'interno del medesimo paradigma di interpretazione adottato per la lettura delle azioni sociali umani (quale è il caso del suddetto esperimento) rischia di non render giustizia alla effettiva configurazione cognitiva della loro mente? In effetti, come riconosce lo stesso Tomasello, sarebbe estremamente riduttivo relegare la percezione delle grandi scimmie all'ambito di una *totale* cecità relativa agli stati intenzionali degli altri. E allora, come conciliare la tesi secondo la quale i primati non umani riconoscono “in qualche modo” gli stati intenzionali degli altri con le

---

<sup>266</sup> Tomasello, M., Op. cit., p. 49 (corsivo mio).

osservazioni emerse dall'esperimento in questione intorno alla mancata esperienza - da parte delle grandi scimmie - degli stati intenzionali del dimostratore? Il passo seguente potrebbe fornirci la soluzione del paradosso:

Poiché i fini e le percezioni altrui non sono immediatamente osservabili, e dunque richiedono inferenze, io stesso pensavo qualche tempo fa che solo gli umani fossero capaci di comprenderne la natura e il modo in cui operano insieme nell'azione diretta a fini. Ma le ricerche recenti [...] hanno cambiato tutto. Le grandi scimmie comprendono che gli altri hanno fini e percezioni, e che questi sono correlati nell'azione intenzionale (e forse addirittura nell'azione razionale). [...] Credo piuttosto che mentre le grandi scimmie comprendono l'azione dell'altro in quanto agente intenzionale individuale, esse non hanno le capacità e le motivazioni per stabilire insieme agli altri fini congiunti e attenzione congiunta, o per partecipare in qualunque altro modo alle relazioni tipiche dell'intenzionalità condivisa.<sup>267</sup>

Il *punctum saliens* in grado di districare il nostro nodo teorico è, dunque, che le grandi scimmie, in effetti, sono in grado di comprendere gli altri in quanto agenti intenzionali senzienti dotati di percezioni e di fini, ovvero di rappresentazioni interne del mondo che in qualche modo vogliono realizzare, ma tale tipo di comprensione è radicalmente diversa nelle scimmie, rispetto agli umani. Le prime, infatti, comprendono l'azione dell'altro in quanto agente intenzionale *individuale*, che agisce, cioè, solo ed esclusivamente per fini individualistici. Al contrario, quando un umano

---

<sup>267</sup> Tomasello, M., *Origins of Human Communication*, Massachusetts Institute of Technology, 2008; trad. it. di S. Romano, *Le origini della comunicazione umana*, Raffaello Cortina, Milano, 2009, p. 155.

indica qualcosa ad un altro suo conspecifico, il ricevente, in maniera spontanea si rivolge verso la cosa indicata, implicitamente consapevole che ciò sarà utile per sé, e che dunque il comunicatore ha agito in suo beneficio, o per motivazioni cooperative:

Così, i bambini sanno che se un adulto indica un secchiello nel contesto di una attività di ricerca, ciò, probabilmente sarà in qualche modo rilevante per il loro fine congiunto di trovare il giocattolo. Ma le grandi scimmie non sanno, né possono, assumere che l'altro stia indicando a loro beneficio, e così non si chiedono: "Perché lui pensa che questo sia rilevante per me?". Invece, vogliono sapere cosa l'altro possa desiderare *per sé* (poiché quando sono loro a indicare è sempre per fini egoistici. [...] Dunque le grandi scimmie semplicemente *non vedono* il gesto di indicazione come qualcosa di rilevante ai loro fini.<sup>268</sup>

In una simile cornice altamente individualistica di comprensione degli stati intenzionali altrui propria dei primati, l'evoluzione della facoltà di apprendere imitando non ha trovato terreno fertile. In altre parole, le grandi scimmie, come illustrato dal suddetto esperimento, non sono in grado di imitare gli altri, ma non perché non ne comprendano fini e percezioni *tout court*:

La conclusione generale è dunque che le grandi scimmie comprendono gli altri nei termini dei loro fini e delle loro percezioni e di come questi e quelle operino per determinare decisioni comportamentali;cioè, esse comprendono gli altri come agenti intenzionali, forse addirittura razionali. Basandosi su questa comprensione, sono in grado di intraprendere quei generi di ragionamento pratico che soggiacciono all'interazione e alla comunicazione sociale flessibile e

---

<sup>268</sup> *Id.*, p. 56.



strategica – per esempio, determinando quel che vuole l'altro, perché lo vuole e cosa è probabile che faccia in seguito.<sup>269</sup>

Così, il modo cognitivo specie-tipico di rappresentarsi il mondo, presso le scimmie antropomorfe, è profondamente intriso di un individualismo tale da venare ogni azione in termini competitivi o di conquista individuale. Per tale ragione, dunque, è scorretto affermare che le scimmie antropomorfe non fanno esperienza degli stati intenzionali di chi compie un'azione. Esse non ne fanno esperienza *in quanto* azioni suscettibili di imitazione, e per ciò non le riproducono, al contrario di quanto accade presso la forma di vita degli esseri umani, caratterizzata, come vedremo meglio nei paragrafi successivi, da cornici attenzionali congiunte e da fini condivisi, ed in cui tale possibilità cognitiva si è rivelata cruciale per l'evoluzione non solo del linguaggio, ma anche dell'intero edificio della cultura umana<sup>270</sup>.

Un aspetto cognitivo fondamentale, legato alla abilità di tenere in considerazione gli stati mentali degli altri è l'abilità di influenzarne i comportamenti. Tale tratto trova dei significativi correlati presso le grandi scimmie. In *Origins of Human Communication*, Tomasello presenta delle

---

<sup>269</sup> *Id.*, p. 53. Su tale tema, vedi Premack, D., *Does the chimpanzee have a theory of mind?*, "Behavioral and Brain Sciences", 4, 1978, pp. 515-526; Tomasello, M. *Does the chimpanzee have a theory of mind? 30 years later*, "Trends in Cognitive Sciences", 2008, pp. 187-192.

<sup>270</sup> Per una rassegna sui diversi livelli di intenzionalità in specie non umane vedi Fitch, W. T., Bugnyar, T., Huber, L., *Cognition and the Evolution of Language: Constructing Cognitive Phylogenies*, "Neuron", 65, 2010, pp. 795-814.

idee assai significative in merito a tale tema, distinguendo due tipi di segnali in grado di sortire degli effetti comportamentali negli altri: le “esibizioni comunicative”, ovvero tratti tipici geneticamente determinati e su cui l’individuo non ha alcun controllo volontario – e i “segnali comunicativi”, espressioni prodotte in modo flessibile e strategico, al fine intenzionale di influenzare il comportamento degli altri. Secondo la prospettiva teorica di Tomasello, esempi del primo tipo di segnali sono l’esibizione delle penne della coda dei pavoni, o i richiami vocali delle scimmie. A proposito di questi ultimi, infatti, egli afferma che i primati non umani non imparano a produrre i loro richiami vocali, né sono in grado di esercitare su di essi un forte controllo volontario. Essi sono geneticamente fissati, e, in effetti, non esistono variazioni individuali, ma tutti posseggono il medesimo specie-specifico repertorio vocale di base. Inoltre, esiste una rigida relazione tra il segnale vocale e l’emozione ad esso associato, o alla situazione scatenante, tant’è che emettono il richiamo a prescindere dalla presenza di un conspecifico:

Un’altra importante caratteristica delle vocalizzazioni dei primati, derivante dalla fisica del canale acustico, è che esse vengono trasmesse indiscriminatamente a chiunque si trovi nelle vicinanze. Questo è un ovvio vantaggio in situazioni altamente emotive e urgenti, ma da un punto di vista psicologico significa che l’emittente non ha bisogno di prestare nessuna attenzione al/ai ricevente/i, e anzi gli risulta difficile indirizzare richiami vocali a individui selezionati, escludendone altri. [...] Nel complesso, le vocalizzazioni

dei primati sembrerebbero espressioni emotive essenzialmente individualistiche, non atti indirizzati a un ricevente.<sup>271</sup>

Ciò vuol dire che i primati non umani, nell'udire un richiamo emesso da un conspecifico, non lo *comprendono* in quanto atto comunicativo rivolto a loro per fini cooperativi, ma al contrario, lo percepiscono semplicemente come un segnale che rinvia in modo deterministico alla presenza di un riferimento ad esso associato, sia esso un pericolo o una situazione emotiva, secondo la logica che, come abbiamo visto nel primo paragrafo, incardina la loro cognitività.

Tuttavia, dallo studio di Tomasello, emerge che i primati non umani siano altresì in grado di *produrre* dei segnali gestuali in maniera controllata, realizzando ciò che classifica come veri e propri “segnali comunicativi”, volti ad influenzare il comportamento degli altri. Le grandi scimmie avrebbero addirittura due tipi di intenzionalità alla base dei loro segnali: l'intenzione sociale, che emana dal fatto che il comunicatore voglia che il suo destinatario faccia qualcosa - e l'intenzione referenziale, che consiste, invece, nel richiamare l'attenzione del ricevente su qualcosa, in modo che questi, guardando nella direzione indicata faccia qualcosa. È importante, qui, sottolineare che i comunicatori indirizzano l'attenzione

---

<sup>271</sup> Tomasello, M., *Origins of Human Communication*, cit., p. 29.

*inducendo* i loro interlocutori a fare cose utili al conseguimento dei propri scopi personali, influenzandone, cioè le scelte comportamentali.

In particolare, in base alla loro funzione comunicativa l'autore distingue i segnali gestuali in “movimenti di intenzione”, e i “richiami dell'attenzione”. I primi sono sostanzialmente il primo passo di una normale sequenza comportamentale, che, una volta ritualizzato, consente di sortire lo stesso effetto della sequenza intera. Per intenderci, un esempio di movimento d'intenzione è alzare il braccio, per invitare il ricevente al gioco, che rappresenta l'abbreviazione della sequenza in cui un individuo con l'idea di avviare un gioco di lotta solleva il braccio per colpire il suo compagno, per poi saltargli addosso e cominciare a giocare. In questo caso, dunque, è chiara l'intenzione sociale del gesto. I richiami dell'attenzione, invece, presentano un livello di complessità maggiore, tanto che Tomasello vi vede una sorta di anello mancante<sup>272</sup> tra la comunicazione umana e quella dei primati. Essi, in effetti, presentano un'intenzionalità a due livelli, poiché consistono in gesti che *nel* dirigere l'attenzione del ricevente su qualcosa (intenzione referenziale), lo induce a compiere l'azione desiderata (intenzione sociale). Un esempio, in tal senso, è dato dall'offrire la schiena: le scimmie, in questi casi, piazzano con insistenza il dorso di fronte al ricevente con la chiara intenzione di indurlo all'attività di *grooming*.

---

<sup>272</sup> Vedi Ivi, p. 38.

È altresì interessante notare che le scimmie non segnalano né in modo dichiarativo (al fine di condividere, cioè, con altri l'attenzione verso qualcosa) né in modo informativo (per far notare qualcosa di rilevante e utile per gli altri). I segnali gestuali delle grandi scimmie sono, infatti, per la quasi totalità imperativi e “non bidirezionali”, nel senso che i partecipanti alla comunicazione non sono in grado di invertire i propri ruoli – cosicché il comunicatore, ad esempio, non riconoscerebbe il gesto che egli stesso attua ritualmente, se fosse qualcun altro ad attuarlo nei suoi confronti. Tali aspetti qualificativi si collocano coerentemente in una cornice cognitiva altamente individualistica e non cooperativa, quale è, appunto, quella specie-tipica delle scimmie antropomorfe.

In ultima analisi, dunque possiamo affermare, con Tomasello, che le grandi scimmie sanno che gli agenti hanno determinati fini (cioè *vogliono* che qualcosa accada), che *vedono* il mondo, anche in modi diversi dal proprio, e che *fanno* cose in modo da realizzare lo stato finale auspicato. Essi, in altri termini, comprendono la logica fondamentale del ragionamento pratico soggiacente all'azione intenzionale (individuale); ciò gli consente di comprendere e predire le reazioni degli altri a determinati comportamenti, ovvero, di influenzarne le scelte comportamentali, inducendoli a fare ciò che vogliono che essi facciano, ma per fini prettamente egoistici.

Il comunicatore si aspetta che se il ricevente *vedrà* il suo gesto, allora *farà* la cosa desiderata, perché è ciò che ha sempre fatto in passato [...]. Da parte sua, vedendo il movimento di intenzione del comunicatore, il ricevente sa che il comunicatore vuole che lui faccia qualcosa di particolare (basandosi sulle sue abilità di lettura delle intenzioni e la sua esperienza passata in interazioni simili).<sup>273</sup>

È opportuno, a questo punto, chiedersi quali possano essere state le relazioni logiche fra gli eventi in grado di condurre l'evoluzione del linguaggio da un dominio intriso di individualismo e competizione quale è quello delle nostre antenate scimmie antropomorfe, all'ambito della cognizione sociale specie-tipica dell'*Homo*. Come già illustrato nel capitolo precedente, il ruolo catalizzante di tale evoluzione è attribuibile all'abilità di modulare prosodicamente le espressioni vocali, le azioni, e di sincronizzare tali modulazioni in gruppo. Tuttavia, rimangono da spiegare nello specifico i meccanismi cognitivi di base che, coevolvendo con le abilità *musicali*, hanno fondato la natura specie-tipicamente sociale dell'uomo. Sarà questo l'oggetto di analisi del prossimo paragrafo.

## **10.2 Le origini dell'intenzionalità condivisa**

Sulla scia del pensiero di Tomasello, sosteniamo che l'anello evolutivo che congiunge la cognizione delle grandi scimmie a quella

---

<sup>273</sup> Tomasello, M., *Origins of Human Communication*, cit., p. 54.

sociale umana sia la capacità di influenzare i comportamenti dei propri conspecifici, facendo richieste. Tale abilità fa capo ad una complessa rete di prerequisiti cognitivi. Per chiedere a qualcun altro di fare qualcosa, infatti, devo preliminarmente essere in grado di riconoscere nell'altro un soggetto non solo in grado di aiutarmi, ma interessato a farlo. Nel paradigma cognitivo dei nostri giorni, ciò sembra un requisito banale, quasi ovvio, in quanto tale tipo di assunzioni praticamente pervade la nostra quotidianità. Ogni giorno, infatti facciamo delle richieste rivolgendoci a chi conosciamo, o, talvolta persino a degli sconosciuti, con l'aspettativa che vengano soddisfatte, implicitamente consapevoli, cioè, che gli altri possano volere aiutarci. Ma per i primi ominidi non era così.

Come abbiamo visto, infatti, la capacità di vedere nell'altro un soggetto che può soddisfare *altruisticamente* le nostre richieste (a prescindere dal livello della gerarchia sociale di appartenenza) è un tratto quasi del tutto assente nel patrimonio cognitivo in comune con le nostre antenate scimmie. Esse eccellono nella facoltà di prevedere le conseguenze di un comportamento, o di un'azione. Così, se imparano ad associare dei suoni (o dei gesti) ad una loro esigenza e, ad esempio, i ricercatori che stanno a contatto con loro la soddisfano puntualmente, è verosimile che possano cominciare a produrre tali suoni al fine di ottenere ciò che vogliono. Ma non per via del riconoscimento di una forma di sudditanza

sociale nei loro confronti da parte dello studioso, né in funzione dell'effettiva comprensione del fatto che questi possa aiutarli per fini non individualistici.

Per questo, ad uno sguardo attento ci accorgeremmo che questa non è una vera e propria richiesta nel senso umano del termine. Non lo è per il fatto che una richiesta, per esser tale, deve presupporre la comprensione dell'altro in quanto agente intenzionale dotato di fini altruistici. Come osserva Tomasello, infatti, quel che gli scimpanzé comprendono quando percepiscono un segnale gestuale è semplicemente che l'altro individuo sta facendo un gesto, di cui immaginano le conseguenze. Ma solo riconoscendo gli altri in quanto soggetti in linea di principio interessati ad aiutarci è possibile guadagnare la capacità di chiedere loro di fare qualcosa a proprio beneficio. Così, l'abilità di influenzare i comportamenti degli altri, tipica delle grandi scimmie, dev'essere sfociata, in qualche modo, nell'abilità di comprendere gli altri in quanto agenti dotati di fini non esclusivamente individualistici e, nel percorso evolutivo, tale abilità, a sua volta, deve avere avviato il motore di un aspetto cruciale della comunicazione: la reciprocità, ovvero la capacità non più soltanto di *fare* richieste, ma anche, di converso, di soddisfare quelle degli altri. Operazione che verosimilmente ha fondato l'abilità cognitiva di vedere in me un agente che come me gli altri può soddisfare delle richieste di aiuto,



e, simmetricamente, di comprendere gli altri in quanto agenti dotati di fini e atteggiamenti mentali, in altre parole, in quanto agenti intenzionali al pari del Sé.

È così, dunque, che prese avvio uno dei motori più importanti dell'evoluzione della cognizione umana, ovvero l'abilità (intrinsecamente sociale) dell'intenzionalità condivisa, in funzione della quale la cognizione umana ha potuto rivoluzionare il patrimonio cognitivo ereditato dall'ultimo antenato comune alle scimmie antropomorfe e all'*Homo*. Tomasello ne parla così:

L'ipotesi è che gli esseri umani posseggano effettivamente un adattamento cognitivo specie-specifico, un adattamento sotto molti aspetti particolarmente potente, giacché cambia in profondità il *processo* dell'evoluzione cognitiva. Questo adattamento è sorto in un particolare momento dell'evoluzione umana, forse in tempi piuttosto recenti, presumibilmente per effetto di eventi genetici e per l'azione della selezione naturale. Esso consiste nella capacità degli individui di identificarsi con i conspecifici (e nella tendenza a farlo) in modi che permettano loro di comprendere quei conspecifici analoghi al Sé, ciascuno in possesso dei propri desideri e della proprie credenze.<sup>274</sup>

In effetti, la capacità di identificarsi con i conspecifici ha cambiato il processo dell'evoluzione della cognizione umana, ridisegnanone in qualche modo la fisionomia, poiché il riconoscimento e la comprensione reciproca di tali atteggiamenti intenzionali condussero verosimilmente alla cardinale possibilità cognitiva di trovare un piano di condivisione di stati

---

<sup>18</sup> Tomasello, M., *The cultural origins of human cognition*, cit., p. 84.

mentali e di riferimenti percettivi, entro un *format* condiviso. È in tal modo, dunque, che presero piede quelle che Tomasello definisce “cornici di attenzione congiunta”, o “terreno comune”, elementi fondanti la possibilità stessa della comunicazione:

Il terreno comune è necessario perché il ricevente possa determinare *ciò* cui il comunicatore vuole che egli rivolga l’attenzione (l’intenzione referenziale), sia il *perché* lo sta facendo (l’intenzione sociale). [...] E anche mantenendo esattamente lo stesso referente, l’intenzione sociale potrebbe differire a seconda del tipo di terreno comune.<sup>275</sup>

Il “terreno comune”, determina ciò che è *rilevante* per l’interazione sociale, ciò che i partecipanti sanno essere tale per gli altri e per sé. Un tale sfondo reciprocamente noto di pertinenza fonda la possibilità stessa di un’intesa comunicativa, in cui il ricevente comprenda sia l’intenzione referenziale che quella sociale del comunicatore. Così, a seconda del tipo di terreno comune, ovvero del modo condiviso di dirigere l’attenzione in funzione dei medesimi criteri di rilevanza, si realizza la possibilità della comprensione dell’atto comunicativo, che si può intendere come un modo per realizzare *insieme* un fine condiviso, all’interno di una medesima cornice concettuale che - in funzione di ciò che è rilevante per *noi* - permette di intendersi su ciò che si dice. Un “noi”, la cui dimensione

---

<sup>275</sup> Tomasello, M., *Origins of Human Communication*, cit., p. 88.

sociale e cognitiva - è bene sottolinearlo ancora una volta - veniva alimentata dalla modulazione prosodica delle comunicazioni, e delle azioni di gruppo.

In tale prospettiva, condizione fondante ogni azione collettiva in generale e - nello specifico, di ogni conversazione è, dunque, la percezione degli altri agenti in quanto individui in grado di rivolgere *altruisticamente* l'attenzione degli altri soggetti su riferimenti pertinenti e utili ai loro fini, ovvero, per dirla con Tomasello, in quanto “candidati all'azione cooperativa”. Questo è un punto che vale la pena enfatizzare, in quanto, d'accordo con l'autore, sosteniamo che un dominio di cooperazione, o di “mutuo soccorso” possa aver rappresentato la piattaforma necessaria alla evoluzione della comunicazione umana, che in effetti può esser intesa come una forma *sui generis* di cooperazione sociale:

La mutua collaborazione è dunque il luogo di nascita del terreno comune concettuale necessario per la comunicazione cooperativa inferenzialmente ricca tipica del genere umano. In termini specificamente comunicativi, se stiamo elaborando il nostro fine comune in una mutua collaborazione ciascuno si avvantaggia aiutando l'altro – ed è anche probabile che noi comprendiamo i tentativi di chiedere e offrire aiuto comunicativamente, visto che ci troviamo sul terreno comune dell'attività cooperativa.<sup>276</sup>

Dalla contestualizzazione di un atto comunicativo all'interno di una cornice di mutua cooperazione, intrisa dalla facoltà di intenzionalità

---

<sup>276</sup> *Ivi*, pp. 169-70.

condivisa, deriva un ulteriore passo in avanti nell'evoluzione della cognizione linguistica umana: oltre al contenuto esplicito del gesto comunicativo, infatti, comincerà ad emergere anche la reciproca *consapevolezza*, da parte dei parlanti, di un soggiacente “ragionamento cooperativo”<sup>277</sup> (tipico - secondo Tomasello – della comunicazione umana). In altre parole, l'intenzionalità condivisa, tradotta da un contesto di “mutua collaborazione”, ad un contesto di “mutua cooperazione *comunicativa*”, darà luogo all'intenzionalità comunicativa, ovvero alla consapevolezza reciproca di quelli che adesso sono caratterizzabili come fini comunicativi: proferimenti o gesti intenzionalmente eseguiti allo scopo di sortire effetti pragmatici, in una cornice attenzionale congiunta, all'interno della quale gli interlocutori in linea di principio tengono reciprocamente in considerazione gli stati attentivi degli altri. Vale la pena riportare, a tal proposito, le parole dell'autore stesso:

Il fatto che i comunicatori operino con queste motivazioni cooperative e che i riceventi tendano a rispondere in modo appropriato (a parità di tutte le altre condizioni) fa parte del terreno comune tra comunicatori umani. Anzi, è ciò che li motiva fin dall'inizio a cooperare per assicurare una corretta trasmissione del messaggio – entrambi assumono in comune che fare ciò sarà utile per loro, personalmente e collettivamente. Poiché il comunicatore lo sa, si assicura che il ricevente sappia a sua volta che il primo sta cercando di comunicare, come a dire: “Devi sapere questo” (cioè, che io ti chiederò qualcosa, che ho qualcosa di cui informarti, che ho un atteggiamento che voglio condividere). Questo livello aggiuntivo di intenzionalità – “Voglio che tu sappia che io voglio qualcosa da te” – è assolutamente cruciale per il

---

<sup>277</sup> *Ivi*, p. 88.

processo e viene comunemente definito “intenzione comunicativa” (griceana).<sup>278</sup>

Poiché un atto soggettivo intriso di intenzionalità comunicativa (griceana) emana da uno sfondo di intenzionalità condivisa, le cui azioni trovano il proprio agente in un “noi” radicale, nel momento stesso della sua realizzazione, entra a far parte della dimensione intersoggettivamente condivisa. Ciò vuol dire che il fatto che io ti stia *intenzionalmente* esibendo un mio atteggiamento, o un mio bisogno, emana dalla dimensione ancora più originaria della tendenza condivisa al “mutuo soccorso” e alla fiducia reciproca proprie di una dimensione sociale in cui ciascuno è implicitamente portato a rispondere positivamente alle aspettative dell’altro, che riconosce in funzione di uno sfondo di stati di cose reciprocamente noto. In tale direzione si colloca il guadagno di una proprietà cardine della cognitività umana, che, secondo il quadro teorico di Tomasello ha permesso di sortire effetti decisivi nella trasmissione della cultura umana. Si tratta della capacità di imitazione con inversione di ruolo,

---

<sup>278</sup> *Ivi*, pp. 83-84. A tal proposito, vedi Tomasello, M., *The cultural origins of human cognition*, cit., p. 128: “[...] La comprensione di un’intenzione comunicativa è un caso speciale della comprensione di un’intenzione; si tratta di comprendere l’intenzione di un’altra persona riguardo al mio stato attentivo. Questa è una forma di comprensione chiaramente più complessa rispetto alla mera comprensione dell’intenzione di un’altra persona. Per comprendere che l’intenzione di un’altra persona è dare un calcio alla palla, devo semplicemente determinare il suo scopo in relazione alla palla. Ma per comprendere che cosa intende un’altra persona quando emette il suono “Palla!” rivolgendosi a me, devo determinare il suo scopo in relazione ai miei stati intenzionali/attentivi verso una certa entità terza.”

che Tomasello descrive contestualmente all'acquisizione ontogenetica del linguaggio:

Per imparare a usare un simbolo comunicativo in maniera convenzionalmente appropriata, il bambino deve impegnarsi in quella che ho chiamato *imitazione per inversione di ruolo* [Tomasello 2001]. Il bambino deve cioè imparare a usare un simbolo nei confronti dell'adulto nello stesso modo in cui l'adulto lo ha usato nei suoi confronti. Questo evidentemente è un processo di apprendimento imitativo nel quale il bambino si conforma all'adulto quanto allo scopo e al mezzo per raggiungere tale scopo; tuttavia, in questo caso, il bambino non deve soltanto sostituirsi all'adulto come attore (che è quel che accade in tutte le forme di apprendimento culturale), ma anche sostituire l'adulto a se stesso come bersaglio dell'atto intenzionale (cioè lo scopo non deve essere il proprio stato attentivo ma, al posto di questo, lo stato attentivo dell'adulto).<sup>279</sup>

È opportuno ribadire che agli stadi primitivi dell'evoluzione del linguaggio, il processo che qui coinvolge l'adulto e il bambino interessava alla stessa stregua tutti gli individui, che in quella fase erano in procinto di dar luogo ad un fenomeno che ne avrebbe radicalmente ristrutturato la mente: il simbolo linguistico. Il risultato di un simile processo di imitazione per inversione di ruoli non può che essere, infatti, uno strumento comunicativo *intenzionalmente* condiviso, in cui chi produce un dato segnale simbolico è in grado di assumere che, a sua volta, anche chi lo ascolta può servirsene, in una cornice di mutua cooperazione comunicativa.

---

<sup>279</sup> Tomasello, M., *The cultural origins of human cognition*, cit., pp. 130-31.

È bene notare che lo stesso non vale per quanto concerne il processo di comprensione dei segnali comunicativi presso i primati non umani, in cui ciascuno dei partecipanti comprende il proprio ruolo solo da un punto di vista interno. Per dirla con Tomasello,

[...] gli infanti umani comprendono l'attività congiunta da un punto di vista esterno, "a volo d'uccello", con il fine congiunto e i ruoli complementari ricompresi in un unico "format rappresentazionale" – il che permette loro di invertire i ruoli quando serve. Di contro, gli scimpanzé comprendono la propria azione da una prospettiva stretta di prima persona e quella del partner da un punto di vista di terza persona, ma sono incapaci di avere una vista "dall'alto" dell'interazione: quindi per loro, in senso stretto, non esistono ruoli, e di conseguenza neanche la possibilità di invertirli nell'ambito della "stessa" attività.<sup>280</sup>

È bene altresì notare che tale contesto intenzionalmente (e comunicativamente) condiviso, il cui agente intenzionale primario è il soggetto plurale "noi" - rappresenta lo sfondo generativo dei segnali simbolici sociali, in funzione dei quali gli oggetti, le persone, e gli eventi che occorrono al suo interno vengono trasformati, acquisendo proprietà in qualche modo speciali. Così, immersi in uno spazio cognitivo siffatto, un semplice pezzo di metallo può trasformarsi in moneta, un individuo fra tanti può divenire capo della tribù, e due persone possono diventare marito e moglie.

---

<sup>280</sup> Tomasello, M., *Origins of Human Communication*, cit., p. 157.

Nel prossimo paragrafo, tenteremo di innestare il contributo dei maggiori studiosi intorno all'evoluzione del linguaggio, nella prospettiva teorica fin qui esposta, al fine di guadagnare una comprensione più profonda delle dinamiche evolutivamente selettive che ci hanno consentito di approdare, conquistandolo, allo spazio cognitivo del linguaggio umano.

### **10.3 Una questione di vita o di morte: la selezione di gruppo**

Seguendo la prospettiva teorica proposta da Tomasello, sosteniamo dunque che la molla che diede avvio all'evoluzione della comunicazione umana è stata la tendenza sociale a rendere noti agli altri i propri fini (assumendo che essi ci aiutino a realizzarli), le informazioni e i propri stati mentali - da cui derivano, rispettivamente, le possibilità cognitive di fare richieste, informare, e condividere emozioni o atteggiamenti. Tale tesi si incontra con quanto esposto da Id Ulbaek, in *The origin of language and cognition*:

Language is cognitive, whereas animal communication is not. Cognitive intelligence is an earlier and more widely spread property of mind than language, because evolution selects for effective information gathering. Language's proper function is to communicate, *which here means sharing information.*<sup>281</sup>

---

<sup>281</sup> Ulbaek, I., *The origin of language and cognition*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy (ed.), "Approaches to the Evolution of Language", Cambridge University Press, MA, 1998, pp. 40-41 (corsivo mio).



Nel suddetto articolo, identificando la cognizione umana con la capacità di condividere (altruisticamente) informazioni, Ulbaek sostiene che la comunicazione animale non vanti la medesima natura cognitiva. Ciò nondimeno, egli riconosce che il linguaggio umano, o meglio l'intelligenza, intesa come la facoltà di condividere e scambiare informazioni, si sia evoluta proprio a partire dai meccanismi cognitivi già in uso presso le altre specie animali, in particolar modo da quelli che le nostre antenate scimmie antropomorfe impiega(va)no nel dominio della mentalizzazione, della navigazione spaziale, nell'uso degli strumenti. Tutte facoltà, queste, che nell'uomo desumono una fisionomia radicalmente diversa, generata dalla costitutiva inclinazione alla *intenzionale* condivisione sociale.

Chiarito tale punto essenziale, è opportuno dirigere il fuoco dell'indagine intorno al valore significazionale delle prime informazioni stesse. Nei capitoli precedenti abbiamo argomentato l'ipotesi secondo la quale l'impronta significazionale del protolinguaggio fosse di natura essenzialmente emozionale e pragmatica. Tuttavia, a tale fase è sicuramente seguito il graduale innesto della facoltà biologicamente adattativa di *significare* oggetti o fatti esterni. È opportuno chiedersi, dunque, quali possano essere stati i primi riferimenti (dalla natura non esclusivamente emozionale o pragmatica) scambiati dai nostri antenati

---

ominidi. In altri termini, ci chiediamo quali aspetti della realtà esercitavano un'influenza tale da incentivare negli uomini la facoltà di farne il primo oggetto di comunicazione. Secondo la nostra prospettiva, giacchè gli aspetti più significativi della realtà dei primi soggetti della specie *Homo* non potevano che avere a che fare con gli elementi più utili alla sopravvivenza, le prime informazioni degne di essere scambiate dovevano vertere anzitutto sulla presenza di cibo, o di pericoli esterni imminenti. Ora, se proviamo ad inserire quest'osservazione all'interno del paradigma teorico di Tomasello, possiamo inferire che le informazioni scambiate dai primi ominidi dovevano riguardare, dunque, un fine condiviso utile alla sopravvivenza. Steven Pinker, a tal proposito, propone un'idea che vale la pena di mettere in luce. In *An adaptation to the cognitive niche*, egli tematizza la questione intorno al ruolo adattativo del linguaggio, cercando una soluzione in grado di render conto della complessità strutturale del linguaggio, e delle modalità in cui le pressioni selettive ne hanno determinato la forma. L'autore osserva che le funzioni soddisfatte delle prime forme di linguaggio, in grado di sospingerne l'evoluzione verso una complessa fisionomia interna stabile, e culturalmente apprendibile, dovevano far capo alla possibilità di codificare informazioni essenziali alla preservazione della vita stessa:

What is the machinery of language trying to accomplish? The system appears to have been put together to encode propositional information – who did what to whom, what is true of what, when, where and why – into a signal that can be conveyed from one person to another. [...] The structures of grammar are well suited to conveying information about technology, such as which two things can be put together to produce a third thing; about the local environment, such as where things are; about the social environment, such as who did things to whom, when where and why; and about one's own intentions, such as *If you do this, I will do that*, allowing people to convey promises and threats that undergird relations of exchange and dominance.<sup>282</sup>

Pinker trova che il linguaggio si sia evoluto in prima istanza come adattamento alla loro “nicchia cognitiva”, ovvero un ambiente che chiamava i propri abitanti a ragionare in termini di causa-effetto, soprattutto in risposta ai pericoli rappresentati da piante ed animali. Tale tipologia di pensiero si sarebbe in seguito estesa su diversi domini cognitivi dell'uomo, traducendosi nell'emergente capacità di organizzare le azioni in sequenze più complesse. Da ciò è derivata, ad esempio, la possibilità di creare le prime forme di utensili, o, quel che qui ci importa sottolineare, l'evoluzione delle prime forme comunicative volte a gettar luce su chi ha fatto cosa a chi, sulle intenzioni altruistiche, e così via.

In questa scia si collocano, altresì, le osservazioni di Dereck Bickerton, il quale sostiene che le prime forme di linguaggio dovevano avere a che fare anzitutto con la ricerca di cibo. Egli identifica una condizione necessaria per l'evoluzione linguistica dei primi esemplari di

---

<sup>282</sup> Pinker, S., *Language as an Adaptation to the Cognitive Niche*, in M. H. Christiansen M. H., S. Kirby S. (ed.), “Language evolution”, Oxford University Press, 2004, p. 27.

uomo nel possesso di una intelligenza sociale che consentisse una qualche forma di intenzionalità condivisa fra i membri di uno stesso gruppo. Tuttavia rimarca che, a prescindere dal ruolo giocato dalle competenze di ordine sociale, la prima spinta sostanziale verso la conquista del linguaggio è ascrivibile alle esigenze vitali strettamente legate al contesto ambientale di sussistenza. L'autore, infatti, afferma:

The foraging considerations suggest that an origin for protolanguage stemming directly from social intelligence and its various uses is relatively unlikely. More probable is an origin deriving from various interactions with the environment – exchange of information gleaned in foraging, interpretation of natural signs, warning of the young against dangers. [...] None of the above in any way implies a reduction of the role that social intelligence undoubtedly played as soon as protolanguage became established, or the vastly important contributions that such use made to the development of both protolanguage and full human language; [...] Moreover, only in a species with a richly developed social intelligence, and with at least some beginnings of a theory of mind and a capacity for intentional communication between group members, could language or even protolanguage ever have developed.<sup>283</sup>

Come Bickerton, sosteniamo che per i primi ominidi le informazioni più “urgenti” dal punto di vista della sopravvivenza, quelle insomma, in grado favorire la selezione naturale della specie (nonché l’evoluzione del linguaggio), dovevano poggiare su una solida cornice sociale all’interno della quale poter essere scambiate. Riteniamo, altresì, che la spinta d’avvio per la comparsa del linguaggio siano provenute dalla esigenza di

---

<sup>283</sup> Bickerton, D., *Foraging versus Social Intelligence in the Evolution of protolanguage*, in A. Wray (ed.), “The Transition to Language”, Oxford University Press, 2002, p. 221.

rintracciare cibo, pericoli, e conquistare partner sessuali. Ma, considerata l'importanza della cognizione sociale in tali condizioni di vita, crediamo che la spinta più forte e decisiva doveva arrivare anzitutto dalla viscerale esigenza dei primi ominidi di preservare ed alimentare la coesione interna, quindi la stabilità del gruppo sociale di appartenenza.

Nella linea evolutiva umana, infatti, il linguaggio ha trovato le condizioni favorevoli alla sua evoluzione, in quanto poggiava su un sistema sociale i cui membri, se da una parte scoprivano una favorevole inclinazione a condividere con gli altri fini, informazioni utili, ed emozioni in uno sfondo di attenzioni congiunte – dall'altra escogitavano stratagemmi utili a mantenere e preservare una solida coesione interna, atta a preservare le condizioni adatte alla realizzazione di simili atteggiamenti altruistici. Inoltre, gli individui erano (piacevolmente) rassicurati nel fatto che appartenere ad un gruppo solido e socialmente efficiente gli avrebbe garantito le condizioni più adatte per far fronte a situazioni estreme di sopravvivenza, cosa che per loro, in effetti sarebbe stata selettivamente vantaggiosa. In tale prospettiva è opportuno sottolineare, ancora una volta con Tomasello, l'importanza della capacità specie-specifica umana dell'imitazione, capacità che poggia su quell'adattamento cognitivo peculiare dell'uomo che gli permette di comprendere gli altri in quanto esseri intenzionali al pari del Sé, e appartenenti al medesimo sfondo

originariamente cooperativo di attenzione e fini congiunti: l'intenzionalità comunicativa. Come abbiamo visto, l'autore osserva che solo in una cornice sociale in cui si è in grado di discernere gli scopi dai mezzi per attuarli, quindi, di *vedere* le azioni eseguite dagli altri venute dallo scopo a cui mirano (e di cui, di fatto, sono intrise) si dà la possibilità di imitare dei movimenti o dei suoni significativi, riproducendoli in quanto mezzi finalizzati ad un determinato effetto fisico, o socio-pragmatico, su cui gli interlocutori si intendono. E, cosa più importante nel dominio cognitivo umano, è che si dia la capacità di usare quei mezzi pragmatici scambiandosi i ruoli all'interno dell'interazione.

Tuttavia, una dimensione spesso non opportunamente evidenziata in merito al ruolo dell'imitazione nella cognizione umana è la sua funzione sociale. L'imitazione, infatti, consente agli individui di un medesimo gruppo di allineare i propri stili comportamentali, uniformando atteggiamenti, esperienze ed emozioni, e favorendo, di conseguenza la coesione del gruppo stesso. E ciò è rilevante, da una parte perché fa capo alla originaria motivazione comunicativa della condivisione, e dall'altra perché sorprendentemente esercita allo stesso tempo una sorta di pressione sui singoli affinché si conformino al resto del gruppo, saturandone le aspettative, pressione che può sfociare nella costituzione di vere e proprie norme sociali. Una simile coercizione sociale, infatti, favorisce la

“selezione culturale di gruppo”, massimizzando le affinità di stile e i comportamenti intra-gruppali, e le differenze con gli altri gruppi sociali. A tale proposito, è importante rimarcare, con Tomasello, che:

Una funzione di rilievo dell'intenzione comunicativa griceana è dunque quella di collocare il mio atto comunicativo nell'arena pubblica, dove tutte le norme si applicano. Se mi rivolgo a voi, e voi ne prendete atto, allora dovrete impegnarvi comunicativamente con me.<sup>284</sup>

Un aspetto interessante di tale dimensione comunicativa è che, sottendendo assunzioni di “fiducia reciproca”, spiana la strada alla possibilità che tale fiducia venga disattesa, ovvero alla possibilità di mentire. Il gruppo sociale cerca di sopprimere tale possibilità creando, nelle modalità che abbiamo visto nel paragrafo precedente, uno sfondo pubblico di norme in grado di screditare certe azioni indesiderate, e chi le compie. «Pertanto, mentre le grandi scimmie possono nascondere cose agli altri, non c'è evidenza che esse siano capaci di fuorviare attivamente i loro simili o dir loro bugie – poiché non comunicano cooperativamente con la mutua aspettativa che entrambi i partecipanti stiano cercando di essere d'aiuto, e siano quindi anche sinceri»<sup>285</sup>. Proprio alla luce della consapevolezza della possibilità che taluni individui mentano, o realizzino atti in grado di minare

---

<sup>284</sup> Tomasello, M., *Origins of Human Communication*, cit., p. 186.

<sup>285</sup> *Ivi.*, p. 187.

l'armonia all'interno del gruppo, l'intera collettività – è bene ribadirlo - si impegna a promuovere comportamenti utili in tale direzione. La maggioranza degli individui, infatti, avrebbe tanto da perdere se la comunità non si mantenesse integra ed unita. Così, conseguenza naturale di una simile inclinazione a preservare la forza e la coesione interna del gruppo, fu una nascente pressione sociale tale da favorire lo scambio di informazioni efficaci nello scoraggiare ogni tentativo di inganno (anche attraverso punizioni morali e fisiche).

È opportuno adesso, alla luce dei concetti fin qui esposti, esporre una tesi corroborata dai più recenti studi di psicologia evolutiva. Si tratta della tesi sul “reciproco altruismo”, fenomeno che, all'interno di uno sfondo sociale cooperativo, giocava un ruolo essenziale nel garantire atteggiamenti favorevoli alla preservazione della coesione interna della comunità di appartenenza, quindi, in ultima analisi, alla sua selezione naturale. A tal proposito, è necessario conciliare la tesi secondo la quale nella linea evolutiva umana vennero selezionati atteggiamenti altruistici con gli standard concettuali della teoria darwiniana, secondo i quali sarebbe svantaggioso, ai fini della selezione individuale, condividere con gli altri informazioni utili o le prede di caccia. In effetti, la cognizione umana si è evoluta proprio a partire da ciò che a primo acchito sembrerebbe andare contro ogni vantaggio selettivo: la condivisione di informazioni utili anche



laddove questa non apportava alcun vantaggio immediato al benefattore. Allora, è il caso adesso di guadagnare, alla luce di alcune teorie recenti, una valida giustificazione di tale dinamica evolutiva, in grado di inserirsi organicamente nella sinossi teorica sulle origini della cognizione umana che stiamo delineando. È ciò di cui ci occuperemo nel prossimo paragrafo.

#### **14.4 L'altruismo adattivo**

Spesso, in biologia, con il termine “altruismo” si intende un comportamento costoso per il soggetto e vantaggioso per il destinatario, e per questa ragione, si tende a vedervi un fenomeno selettivamente non adattivo. Ma uno sguardo più attento del fenomeno può rivelarci che anche gli atteggiamenti non direttamente volti all'esclusivo beneficio personale possono nondimeno risultare vantaggiosi. De Waal a tal proposito afferma che «l'evoluzione favorisce gli animali che si aiutano reciprocamente se in questo modo raggiungono vantaggi a lungo termine di maggiore entità rispetto a quelli che gli deriverebbero se facessero da soli e se entrassero in competizione con gli altri»<sup>286</sup>. Ciò vuol dire che il soggetto altruista è nelle condizioni di ricevere dei benefici, seppur in un momento non simultaneo, o immediato rispetto a quando si è fatto carico dell'azione dispendiosa.

---

<sup>286</sup> De Waal, F., *Primates and Philosophers*, Princeton University Press, 2006; trad. it. di F. Conte, *Primati e filosofi. Evoluzione e moralità*, Garzanti, Milano, 2008, pp. 33-34.

Chiaramente, la possibilità stessa della propagazione di atteggiamenti cooperativi e generosi poggia sull'aspettativa del ritorno di un beneficio, cosicché i ruoli di beneficiario e ricevente siano in un continuo alternarsi. Ne deriva, dunque, che l'evoluzione delle tendenze all'altruismo possa effettivamente convenire agli interessi personali, quindi, in ultima analisi, all'evoluzione della specie.

In tale direzione si collocano le osservazioni di Tomasello, secondo cui uno degli interessi individuali maggiori nell'essere cooperativi consiste nel guadagno di una buona nomea nel gruppo. Talvolta, infatti, gli umani offrono aiuto in maniera indiretta, collaborano, cioè, con gli altri al fine di guadagnare una buona reputazione di cooperatività, anche in mancanza di un immediato beneficio diretto. Così un individuo può offrire aiuto agli altri, consapevole che i suoi atti possono contribuire a migliorare la propria reputazione, quindi il suo *status* e che ciò, in futuro, porterà gli altri a voler collaborare con lui. Inoltre, è importante sottolineare che da simili contesti "superservizievoli" (in cui è implicita una volontà di mutua collaborazione) emerge l'intenzione comunicativa griceana: il ricevente comprende l'intenzione comunicativa del parlante e questi, a sua volta, si aspetta che il primo risponda in maniera cooperativa a tale atto comunicativo.

Tale tesi incontra il sostegno delle argomentazioni di Jean-Louis Dessalles che in *Altruism, status and the origin of relevance* afferma:

People feel especially concerned by anything connected with individuals they regard as high-ranking. This makes the coupling between status and relevance symmetrical: by being relevant, one gets status in return, and conversely one is more likely to be relevant if one has high status or is talking about high-ranking people.<sup>287</sup>

Così, in un contesto di reciproco altruismo, gli individui in grado di condividere informazioni rilevanti - che secondo il pensiero di Dessalles sono quelle in grado di esibire in qualche maniera la fatticità di situazioni desiderabili, improbabili o imprevedute<sup>288</sup> (quindi di risolvere l'eventuale problematicità di una situazione) - sono coloro che avranno più possibilità di migliorare la propria reputazione agli occhi della collettività, con implicazioni positive anche per quanto concerne il potenziale di seduzione sessuale.

### **14.5 Strategie di coesione: gossip, promesse, e rituali.**

Come abbiamo visto, alle origini il linguaggio poggiava su una struttura sociale cooperativa, e all'interno della quale agli individui premeva per lo più rinsaldare i legami che occorreivano al suo interno, quindi, in generale, la forza e l'unione della propria comunità.

---

<sup>287</sup> Dessalles, J., *Altruism, status, and the origin of relevance*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy, (ed.), "Approaches to the Evolution of Language", Cambridge University Press, 1998, p. 143.

<sup>147</sup> *Ivi*, p. 133.

L'appartenenza ad un dominio pragmatico sociale più stabile e vantaggioso rispetto a quello strettamente individualistico garantisce, infatti, migliori condizioni in termini emozionali, e di sopravvivenza. Inoltre, come si può facilmente intuire, quanto più è largo il gruppo, tanto maggiori saranno i vantaggi in tal senso. A tali idee si connette la tesi di Robin Dunbar, il quale osserva:

Per mantenere uniti i loro gruppi, i primati si servono soprattutto delle operazioni di *grooming*, cioè di pulizia del corpo cui si sottopongono reciprocamente. Non siamo veramente in grado di capire perché il *grooming* riesca a tenere unito il gruppo, ma il fatto è che il tempo speso nel pulirsi reciprocamente, nei vari gruppi di scimmie, antropomorfe e no, è direttamente proporzionale alle dimensioni tipiche delle comunità in cui essi vivono.<sup>289</sup>

Secondo alcune ricerche sperimentali condotte su piccole scimmie non antropomorfe, il *grooming* è efficace nel creare un legame fra due individui. Infatti, quando una scimmia si appresta a pulire il corpo di un'altra, e ad effettuare lo "spulciamento", provoca il rilascio - nella beneficiaria di queste cure - di endorfine, antidolorifici naturali appartenenti alla famiglia degli oppioidi, prodotti dal cervello stesso. In altri termini, durante questa attività, il primate si rilassa ed acquista un senso di benessere tale da fare del *grooming* un fenomeno di coesione nei gruppi sociali, in quanto consente ai partner di sentirsi bene quando stanno

---

<sup>289</sup> Dunbar, R. I. M., *The Human Story. A New History of Mankind's Evolution*, Faber and Faber Limited, London, 2004; trad. it. di D. Giusti, *La scimmia pensante. Storia dell'evoluzione umana*, Il Mulino, Bologna, 2009, p. 131.

insieme, e tale sensazione si traduce nella volontà di sostenersi reciprocamente in casi di conflitto.

Dunque, è in questo meccanismo che Dunbar ravvisa l'elemento in grado mantenere coeso un gruppo sociale. E tanto più è grande la comunità in cui i primati vivono, tanto maggiore risulta il tempo speso nella pulizia reciproca. Ora, è bene notare, con l'autore stesso, che nel caso degli umani, poiché i cui gruppi talvolta superano i 150 membri, la realizzazione di simili cure fisiche reciproche con tutti (come avviene nei primati) sarebbe proibitiva, in quanto porterebbe via troppo tempo utile ad altre attività indispensabili alla sopravvivenza. Ed è a questo punto, secondo l'autore, che la cognizione umana avrebbe preso le mosse da quella degli altri primati:

Parola e linguaggio si sono evoluti per consentirci di tenere insieme raggruppamenti sociali che stavano divenendo troppo grandi perché il tradizionale *grooming* dei primati bastasse a consolidarli. Tuttavia, ancora oggi noi facciamo ricorso a questo comportamento, essenzialmente con i medesimi scopi. Inoltre, per garantire maggiore efficacia all'azione del linguaggio, dobbiamo attingere con abbondanza a risorse non verbali (come la capacità di ridere e di produrre musica), che ci riconducono direttamente agli stessi processi chimici attivati dal *grooming*.<sup>290</sup>

---

<sup>290</sup> *Ivi*, p. 155. Vedi Dunbar, R. I. M., *The Origin and Subsequent Evolution of Language*, in M. H. Christiansen, S. Kirby (ed.), "Language evolution", Oxford University Press, 2004, p. 220: "Hunting is often best done in silence, and tool-making is best done by demonstration, rather than instruction. In addition, the size of human social groups gives rise to a serious problem: grooming is the mechanism that is used to bond social groups among primates, but human groups are so large that it would be impossible to invest enough time in grooming to bond groups of this size effectively. The alternative suggestion, then, is that language evolved as a device for bonding large social groups – in other words, as a form of grooming-at-a-distance."

In altri termini, il tempo speso per le interazioni con gli altri varia in proporzione al numero dei membri del gruppo, cosicché oltre una certa soglia numerica il *grooming* ha smesso di essere l'operazione più idonea alle relazioni sociali, e venne sostituito da un mezzo in grado di coinvolgere più persone contemporaneamente, sortendo effetti parimenti piacevoli. A prova delle aumentate dimensioni dei gruppi, Dunbar porta la tesi intorno all'esistenza di una precisa correlazione tra le dimensioni della neocorteccia e quelle del gruppo in cui gli animali vivono. Coerentemente con le sue idee, in base a quel rapporto, la massa cerebrale di un essere umano prospetta un gruppo di 150 individui, e non solo: posto che nel corso dell'evoluzione dei primati, l'espansione del cervello è avvenuta verso la fronte, provocando un'espansione vistosa del lobo frontale, è bene tenere presente un dato cruciale ai fini del presente studio, ovvero la grande abbondanza di dati clinici a sostegno dell'ipotesi che il lobo frontale del cervello abbia un ruolo cruciale nell'attività di leggere una mente altrui<sup>291</sup>. Ne segue che il cervello degli *Homo* si sia espanso in correlazione alla crescente realizzazione di azioni socialmente sempre più sofisticate, legate in particolare a competenze come la capacità di lettura della mente, o di coltivare relazioni amichevoli. In altri termini, l'espansione del cervello e l'evoluzione delle competenze sociali facevano capo alla crescente

---

<sup>291</sup> Vedi Dunbar, R. I. M., *The Human Story. A New History of Mankind's Evolution*, cit., pp. 84-89.

necessità di vivere in mondo sociale sempre più complesso, un mondo in grado di tradurre le emozioni e il benessere in una nuova dimensione. Cosa che, verosimilmente, ha generato anche una nuova specie: l'animale linguistico.

È bene notare che l'autore non è in possesso di una spiegazione causale più precisa intorno alle dinamiche che hanno portato i gruppi umani a raggiungere dimensioni tali da richiedere l'installazione di comportamenti nuovi, in grado di mantenere un certo grado di coesione interna. Egli riconosce nel linguaggio (questa sorta di “*grooming* a distanza”) il *risultato* di un simile allargamento dei gruppi sociali, piuttosto che la causa. Nondimeno, esso gioca un ruolo efficientissimo nel mantenerne l'identità e la coesione interna, per mezzo di operazioni di cui ancora oggi viene riconosciuto il fascino:

I suggest, then, that the principal function of language was (and still is!) to enable the exchange of social information ('gossip') in order to facilitate bonding in larger, more dispersal groups. Additional complementary benefits can, of course, be seen in terms of the ability to monitor and control social cheats [...] and the spread of malicious propaganda [...].<sup>292</sup>

In altri termini, secondo questa prospettiva, il linguaggio - fin dagli esordi - consentiva lo scambio *cooperativo*, se vogliamo – di gossip,

---

<sup>292</sup> Dunbar, R. I. M., *Theory of mind and the evolution of language*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy, “Approaches to the Evolution of Language”, Cambridge University Press, MA, 1998, p. 98.

ovvero informazioni concernenti le relazioni sociali entro la comunità di appartenenza, e la possibilità di instaurare stabili legami interni. Infatti, questa nuova tipologia vocale di *grooming*, da una parte, rappresentava un modo per render manifeste le caratteristiche degli individui, consentendo di scegliere con maggiore cognizione di causa gli amici o i partner all'interno delle reti sociali in cui si era inseriti - dall'altra, forniva la possibilità di controllare e scovare le strategie machiavelliche messe in atto dagli individui meno leali all'interno del gruppo, favorendo in tal modo l'assunzione di atteggiamenti amichevoli, e fungendo, di conseguenza, da dispositivo di coesione per i gruppi sociali. Eppure, il potenziale dei gossip può non limitarsi soltanto a questo.

Come nota Terrence Deacon, i gossip diffusi dai membri del gruppo sociale cui si appartiene possono dissuadere gli individui da atti di infedeltà sessuale verso i propri partner, o al contrario, migliorare la reputazione degli individui. L'autore identifica proprio in tale aspetto la molla d'avvio dell'evoluzione del linguaggio. Vediamo meglio in che senso:

Il problema è organizzare il comportamento di gruppo intorno a un'essenza intangibile come un codice di comportamento desiderato per il futuro. Ogni individuo deve essere a conoscenza di queste aspettative; ma soprattutto, ognuno deve sapersi affidare al sostegno di altri membri del gruppo, per prevenire le violazioni di tali modelli di comportamento.<sup>293</sup>

---

<sup>293</sup> Deacon T. W., *The Symbolic Specie. The Coevolution of Language and the Brain*, cit., pp. 386-87.



In altri termini, secondo l'autore, presso le primitive società di ominidi, si avvertì l'esigenza di creare dei modelli di comportamento, ovvero delle sorte di norme pubbliche in grado di conformare le azioni degli individui entro un unico paradigma di correttezza e onestà in grado di preservarne i legami interni. Considerando una caratteristica cruciale delle prime società di ominidi, ovvero il fatto che i maschi delle società, per impegni di caccia, erano costretti ad allontanarsi dalle rispettive partner (e di converso, che anche le femmine restavano separate dai propri compagni), l'autore mette in risalto una precipua esigenza da parte dei congiunti: quella di poter esser sicuri della fedeltà sessuale del proprio partner. Quale assicurazione migliore degli occhi dell'intera collettività?

A tal fine, bisognava stabilire un modo per render pubblicamente nota l'unione fra i prescelti, una sorta di "promessa" in grado di radicare l'unione nella rete sociale di appartenenza: un dominio speciale, che ne potesse sancire l'indissolubilità. Si trattava, dunque, di una promessa *simbolica* in grado di trarre la sua forza deterrente dalla stessa struttura sociale, legittimando gli altri membri della società a denunciare eventuali tradimenti sessuali, o a evitarli. In altre parole, si trattava del matrimonio.

Il matrimonio, nella sua incredibile varietà, è la regolamentazione di relazioni riproduttive con mezzi simbolici e, nelle culture umane, è pressoché universale.

Si tratta preminentemente di una relazione simbolica, assente nel resto del regno animale. Allora, dove voglio arrivare? A sostenere che una forma correlata di regolamentazione delle relazioni riproduttive con mezzi simbolici fu essenziale affinché gli ominidi primitivi traessero vantaggio da una strategia di sussistenza per approvvigionamento basato sulla caccia.<sup>294</sup>

Bisogna riconoscere, altresì, che per i primi ominidi, dal cervello simile a quello di uno scimpanzé, o di un bonobo, non deve essere stato di certo facile approdare ad una simile conquista cognitiva, ovvero alla possibilità di fissare una relazione simbolica (dal valore sociale). E allora, ci chiediamo con Deacon, come sia nato spontaneamente un contesto sociale in grado di far fronte a tali limiti neuroanatomici, ed in cui le relazioni simboliche non solo venivano fissate, ma addirittura trasmesse. Naturalmente ciò dovette avvenire molto gradualmente e con una serie di prove ed errori evolutivi, che nondimeno riuscirono a raggiungere una certa soglia di stabilità referenziale. Vista dall'interno della prospettiva teorica che stiamo illustrando, la tesi di Deacon, che identifica il *primum movens* del linguaggio in una promessa fra due amanti socialmente riconosciuta, rinvia ad una considerazione che vale la pena mettere in luce. Infatti, nell'incentivare la stabilità di una coppia, la promessa conferma allo stesso tempo la fiducia nella struttura sociale su cui poggia, che a sua volta ne guadagna in saldezza; e ciò contribuisce, secondo una dinamica circolare,

---

<sup>294</sup> *Ivi*, p. 387.

alla validità delle promesse pubbliche che occorrono al suo interno. Una dinamica, dunque, che nell'alimentare la forza del gruppo sociale - aspetto come abbiamo visto cruciale ai fini della sopravvivenza – conferisce maggiore solidità anche ai legami emozionali che vi occorrono, favorendo il clima di armonia intra-gruppale. Ma quale può essere stata la valvola di alimentazione di un simile processo di evoluzione simbolica?

Al pari degli scimpanzé, gli ominidi primitivi dovettero imparare un complesso di associazioni tra segni e oggetti, a ripeterli in continuazione, e col tempo, a disimparare l'associazione concreta a favore di una più astratta. Il processo ha dovuto essere mantenuto finché non si scoprì il sistema completo di relazioni combinatorie tra simboli. Che cosa di comparabile avrebbe potuto alimentare queste esigenze nelle prime culture dove sono nati i simboli? La risposta è, in una parola, il rituale.<sup>295</sup>

Nel rituale, dunque, è possibile vedere una pratica che, continuamente attuata con l'accompagnamento di musiche, danze, gesti o proferimenti specifici – ha consentito di spostare l'attenzione dai riferimenti concreti ad oggetti, eventi, o comportamenti isolati, a riferimenti più astratti, fondati dall'insieme delle relazioni tra essi

---

<sup>295</sup> *Ivi*, p. 388. Vedi *Ivi*, pp. 393-94: “In ogni caso non dovremmo dimenticare che i simboli sono ancora strettamente legati a pratiche culturali e strumentali di tipo rituale. Anche se la parola trasmette molte forme di informazione libere da qualsivoglia supporto materiale, in pratica ampi supporti contestuali fisici e sociali influiscono spesso sul contenuto della comunicazione. L'acquisizione del linguaggio si affida ancora a una significativa ritualizzazione e irreggimentazione di tipo ludico del contesto di acquisizione dei simboli [...]. Allora, venendo al dunque, la teoria delle origini simboliche or ora descritta [...] non è una teoria delle origini del comportamento sociale, ma lo è della traduzione del comportamento sociale in forma simbolica.”

intercorrenti, ovvero ai riferimenti non più indicali ma simbolici, quindi pubblicamente manifesti e riconosciuti.

Così, proprio le pratiche sociali, alimentate dal contributo simultaneo e sinestesico di più canali funzionali anzitutto dal punto di vista emozionale (gestuale, mimico, vocale-) e guidate dall'esigenza di solidificare i legami intragruppali, quindi di raffinare le modalità relazionali, avrebbero favorito l'evoluzione della comunicazione simbolica. Tra queste pratiche, come abbiamo visto, un ruolo cruciale era verosimilmente giocato da quelle concernenti i gossip, o i rituali di matrimonio. Vale la pena, a tal proposito, illustrare le idee di Camilla Power, il cui intreccio abbraccia le tesi esposte in questo paragrafo in un unico pensiero organicamente strutturato. La studiosa in primo luogo si chiede da cosa potesse desumere la sua capacità di alimentare la coesione sociale un mezzo - come quello vocale - che a differenza del *grooming* solleva gli individui da ogni impegno in termini di tempo o fatica fisica, e che per di più non provoca alla stessa stregua il rilascio di sostanze oppiacee. E a tal proposito osserva:

We then need to ask what is the reinforcing mechanism that makes vocal grooming intrinsically pleasureable? [...] It is possible to imagine similar opiate stimulation arising from emotive vocalizations of 'sing-song' or 'motherese' types involving collective or synchronized singing that imparted a sense of security and belonging [...].<sup>296</sup>

---

<sup>296</sup> Power C., *Old wives' tales: the gossip hypothesis and the reliability of cheap signals*, in C. Knight C., J. R. Hurford, M. Studdert-Kennedy, "Approaches to the Evolution of Language", Cambridge University Press, MA, 1998, p. 114.

In questo passo emerge un punto chiave per il nostro studio, ovvero la caratterizzazione emotiva delle vocalizzazioni, intese non più come mere “esibizioni comunicative”, bensì come segnali con un incipiente potenziale di significazione, cui contribuisce in qualche misura la dimensione emozionale, e nello specifico le emozioni legate al senso di sicurezza e di appartenenza suscitate da canti collettivi sincronizzati, effettuati quindi da individui avvolti in una rete sociale armoniosa. Proprio al fine di garantire una sì piacevole armonia sociale, urgeva trovare un espediente con un forte potere deterrente relativamente all’inganno e alle sovversioni. In linea con le idee di Dunbar, l’autrice identifica tale espediente nel gossip, la cui pratica presuppone uno sfondo sociale di fiducia reciproca. Inoltre, la studiosa identifica nei rituali un contesto indispensabile per l’evoluzione del linguaggio, in quanto pratica la cui attuazione attraverso musiche, danze, e altri generi di sforzi cognitivi e fisici di gruppo implicava un grande impiego di energia che, nel chiamare i soggetti a fornire una sorta di prova fisica in segno della loro onestà, ne mostrava la disponibilità ad integrarsi nel gruppo. Tale precipua caratterizzazione avrebbe spinto i primi ominidi a credere alle informazioni di cui non potevano essere i diretti testimoni, e che, per di più, essendo vocali costavano anche meno in termini di fatica. Così, attraverso le faticose prestazioni richieste dalla stessa realizzazione del rito, ogni individuo, avrebbe sancito la propria

appartenenza al gruppo, guadagnandone al contempo la fiducia. Di conseguenza le sue informazioni (ovvero, stando alle idee dell'autrice, i suoi pettegolezzi) avrebbero desunto il merito di essere ritenute vere.

Il rituale, dunque, viene inteso dai maggiori studiosi intorno all'evoluzione del linguaggio come la culla della cognizione umana, vuoi perché fornisce un momento di forte pubblicità dei legami all'interno della rete sociale cui si appartiene, o in virtù del valore (in termini di fatica fisica) necessario a guadagnare la stima e la fiducia dei compagni, e a solidificare quei legami. Ma la storia non finisce qui. Come osserva Chris Knight, il rituale è efficace anche nel creare quel livello simbolico di rappresentazione del mondo, in grado di differenziare la cognitività umana da quella dei primati. Nello specifico, l'autore trova nel rituale la matrice della natura distintiva del linguaggio, che concepisce in termini di gioco di finzione condiviso, che presuppone uno sfondo cooperativo e di fiducia:

Pre-modern humans in this model are heavily involved in communal pretend-play or 'mimesis' – fantasy-sharing representational activity such as mime, song and dance [...]; this drives selection pressures for subtle volitional control over emotionally expressive vocalizations and linked gestural representations.<sup>297</sup>

---

<sup>297</sup> Knight C., *Ritual/speech co-evolution: a 'selfish gene' solution to the problem of deception*, in C. Knight C., J. R. Hurford, M. Studdert-Kennedy, "Approaches to the Evolution of Language", cit., p. 87.

Così, mentre i richiami vocali delle scimmie non possono non rinviare alla specifica situazione emozionale, o di pericolo cui sono associate, lo stesso non vale di certo per il linguaggio umano, in cui l'autore vede anzitutto un mezzo per ingannare (e in seguito, mentire). A tal proposito si chiede: «Imagine a population in which volitional signals are becoming commonplace, thanks initially to skills in deception. How can a new honest strategy invade the deceptive one and become evolutionary stable?»<sup>298</sup>. Come può, dunque, un dispositivo cognitivo che, non può prescindere da un contesto di mutua collaborazione e fiducia prender piede in una fase inizialmente votata all'inganno per fini egoistici? L'autore trova la soluzione a tale quesito concependo il linguaggio come un mezzo per creare un livello di finzione comunemente condiviso, tale da non minare alla base un contesto di mutua cooperazione e, utile nel creare quell'identità di gruppo necessaria ad alimentare un senso di co-appartenenza al suo interno:

Myths, dramatic performance, art and indeed all expressions of human symbolic culture may in this light be understood as 'collusion in deception' [...] – collaboration in the maintenance of fictions which have social support.<sup>299</sup>

---

<sup>298</sup> *Ivi*, p. 74.

<sup>299</sup> *Ivi*, p. 76.

Tanto le ricerche sperimentali comparative quanto lo studio teorico sull'evoluzione del linguaggio verbale chiamano a riconoscere una componente chiave di tale processo proprio nell'intima natura sociale degli uomini, ovvero nell'esigenza costitutivamente umana di "far gruppo", di condividere emozioni e pensieri con i propri simili, o di esternare le proprie passioni. Troviamo che si possa identificare la dimensione sociale come la madre e, allo stesso tempo, l'anima del linguaggio verbale. Madre, anzitutto perchè coevolvendo con la natura prosodica delle modulazioni protolinguistiche, ha creato le condizioni essenziali all'evoluzione della necessità di comunicare intenzionalmente con l'altro; tale inclinazione si è tradotta *ipso facto* nella possibilità cognitiva di dar luogo a complessità strutturali facilmente acquisibili, in grado di promuovere la l'evoluzione del linguaggio stesso, e la sua trasmissione a livello culturale.

Per capire in che senso ne costituisce l'anima, propongo il seguente passo del *De Anima* di Aristotele:

Se l'occhio fosse un animale, la sua anima sarebbe la vista: questa è la sostanza dell'occhio rispetto alla sua definizione. L'occhio è la materia della vista: quando la vista viene meno non c'è più l'occhio tranne che per omonimia come nel caso dell'occhio di pietra o disegnato. [...] L'anima è attività orientata a un fine come lo sono la vista e la potenzialità dell'organo-strumento: il corpo è ciò



che è capace di fare. Come l'occhio è pupilla e vista così l'animale è anima e corpo<sup>300</sup>.

In termini aristotelici, riteniamo che “l'anima” dell'uomo sia il suo intrinseco essere sociale, in quanto la sua naturale, ancestrale predisposizione ad integrarsi in un gruppo ha trainato l'evoluzione del linguaggio proposizionale stesso. Tale abilità si traduce retroattivamente nella possibilità di articolare pensieri estremamente complessi, teorie scientifiche, nella pianificazione di progetti, nella capacità di mentire, o banalmente, di organizzare una festa. In altri termini, il linguaggio proposizionale fonda la stessa forma di vita specie-tipicamente umana: un paradigma di riconoscimento e comprensione intersoggettiva in cui si l'abilità di condividere le esperienze vissute avvolge le espressioni stesse del linguaggio, colorandole di quella familiarità tipica delle parole, rese così simili a “volti ben noti”<sup>301</sup>.

---

<sup>300</sup> Aristotele, *De Anima*, 412b 18-413a 3, in Lo Piparo F., *Aristotele e il linguaggio. cosa fa di una lingua una lingua*, Laterza, Roma-Bari, 2003, p. 9.

<sup>301</sup> Vedi. Wittgenstein L., *Philosophische Untersuchungen*, cit., § 167: “[...] Infatti la già semplice vista di una riga stampata è straordinariamente caratteristica, è, cioè, un'immagine del tutto speciale: [...] le parole, che in gran parte si ripetono continuamente e ci sono estremamente familiari, proprio come volti ben noti. [...] Poni mente al fatto che l'aspetto delle parole ci è familiare in grado analogo a quello in cui ci è familiare il suo suono.”

## Conclusioni

Passando per l'analisi critica degli aspetti chiave dell'emergenza del linguaggio (sia a livello filogenetico che ontogenetico), il nostro lavoro di ricerca ha tratto linfa, nel corso del suo sviluppo, dalla volontà di metterne in chiaro le componenti cognitive che animano l'attività di linguaggio presso gli uomini. Si è cercato, dunque, di congiungere i numerosi e complessi aspetti di questo affascinante fenomeno, con l'obiettivo di sortire un risultato coerente, unitario, e il più possibile aderente alla realtà evolutiva della cognizione umana. Il metodo teorico che abbiamo adottato per la risoluzione di tale questione è quello il cui approccio fa capo alla concezione del linguaggio umano non come il prodotto finale dell'evoluzione lineare e graduale di una singola cellula, ma di una dinamica che molto plausibilmente ha interessato simultaneamente diversi ambiti che, come le fila di un'unica maglia, sovrapponendosi ed intrecciandosi entro una medesima cornice sociale, ha dato luogo a ciò che oggi conosciamo come il *nostro* linguaggio: una dimensione cognitiva in grado di dotare di un senso pragmaticamente condivisibile persino enunciati o situazioni contestuali logicamente difettivi e incompiuti. Una dimensione che, proprio da una dinamica di continua costruzione di senso

entro contesti normativi di interazione sociale, desume il suo ordine perfetto:

Da un lato è chiaro che ogni proposizione “è in ordine così com’è”. [...] D’altra parte sembra chiaro questo: che, dove c’è senso, là deve esserci ordine perfetto. – L’ordine perfetto deve dunque essere presente anche nella proposizione più vaga.<sup>302</sup>

---

<sup>302</sup> *Ivi*, § 98.

## Bibliografia

Abe, K., Watanabe, D., *Songbirds possess the spontaneous ability to discriminate syntactic rules*, "Nature Neuroscience", 14, 2011, pp. 1067–1074.

Arbib, M. A., "The Evolving Mirror System: A Neural Basis for Language Readiness", M. H. Christiansen, S. Kirby, (ed.), "Language evolution", cit., pp. 182-200.

Arcadi, A., Robert, D., Buttress drumming by wild chimpanzees: *Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness*, "Primates", 39, 1998, pp. 505-518.

Arnold, K., Zuberbühler, K., *Semantic combinations in primate calls*, "Nature", 441, 2006, p. 303.

Arnold, K., Zuberbühler, K., *Meaningful call combinations in a non-human primate*, "Current Biology", 18, 2008, pp. R202–R203.

Berwick, R. C., Okanoya, K., Beckers, G. J., L., Bolhuis, J., *Songs to syntax: the linguistics of birdsong*, "Trends in Cognitive Science", 15, 2011, pp. 113-121.

Bickerton, D., *Foraging versus Social Intelligence in the Evolution of protolanguage*, in A. Wray (ed.), "The Transition to Language", Oxford University Press, 2002, pp. 207-225.

Bickerton, D., *Symbol and Structure: A Comprehensive Framework for Language Evolution*, in M. H. Christiansen, S. Kirby, (ed.), "Language evolution", Oxford University Press, 2004, pp. 92-93.

Brady, T., Oliva, A., *Statistical learning using real-world scenes*, "Psychological science", 19, 2008, pp. 678- 685.

Brown, S., *The "Musilanguage" Model of Music Evolution*, in N. L. Wallin, B. Merker, S. Brown, (ed.), "The Origins of Music", Mit Press, Cambridge, 2001, pp. 271-300.

Burnham, D., Kitamura, C. *What's new, pussycat? On talking to babies and animals*, "Science", 296, 2002, p. 1435.

Carapezza, M., *Uexkiill, la nozione di Umwelt e il parlare di animali*, in Annali del dipartimento di Filosofia, Storia e Critica dei Saperi (Università di Palermo), 2006.

Carapezza, M., *Segno e Simbolo in Wittgenstein*, Bonanno, Catania, 2006.

Chomsky, N., "Three models for the description of language", *Information Theory*, 1956.

Chomsky, N., *Rules and Representations*, Columbia University Press, 1980; trad. it. di G. Gallo, *Regole e rappresentazioni*, Baldini Castoldi Dalai, Milano, 2008.

Chomsky, N., *Language and problems of knowledge. The managua lectures*, The MIT press, Cambridge, 1988, trad. it. di C. Donati e A. Moro, *Linguaggio e problemi della conoscenza*, Il Mulino, Bologna, 1998

Christiansen, M. H., Curtin, S., *Transfer of learning: Rule acquisition or statistical learning?*, "Trends in Cognitive Science", 3, 1999, pp. 289-290.

Christophe, A., Guasti, T., Nespore, M., Dupoux, E., Van Ooyen, B., *Reflections on phonological bootstrapping: Its role in lexical and syntactic acquisition*, "Language and Cognitive Processes", 12, 1997, pp. 585-612

Christophe, A., Millotte, S., Bernal, S., Lidz, J., *Bootstrapping Lexical and Syntactic Acquisition*, "Language and Speech", 51, 2008, pp. 61-75.

Clarke, E., Reichard, U. H., & Zuberbühler, K., *The syntax and meaning of wild gibbon songs* “PLoS ONE”, 1, 2006, pp. 1-10.

Clay, Z., Zuberbühler, K., *Bonobos Extract Meaning from Call Sequences*, “PLoS ONE”, 6, 2011, pp.1-10

Conway, C., *Sequential learning in non-human primates*, “Trends in Cognitive Sciences”, 5, 2001, pp. 539-546.

Cooper, R. P., Abraham, J., Berman S., Stask, M., *The development of infants' preference for motherese*, “Infants Behavior and development”, 20, 1997, pp. 477-488.

Corballis, M. C., *Did Language Evolve from Manual Gestures?*, in A. Wray, A. (ed.), “The Transition to Language”, Oxford University Press, 2002, pp. 161-79.

Corballis, M. C., *From Hand to Mouth: The Gestural Origins of Language*, in M. H. Christiansen, S. Kirby, (ed.), “Language evolution”, Oxford University Press, 2004, pp. 201-218.

Corballis, M. C., *Recursion, language, and starlings*, “Cognitive Science”, 31, 2007, pp. 697-704.

Crockford, C., Herbinger, I., Vigilant, L. (2004). *Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning*, “Ethology”, 110, 2004, pp. 221-243

Crockford, C., Boesch, C., *Call combinations in wild chimpanzees*, “Behaviour” 142, 2005, pp. 397–421.

Darwin, C., *The Descent of Man, and Selection in Relation to Sex*, John Murray, London, 1871.

De Mauro, T., *Prima lezione sul linguaggio*, Laterza, Roma-Bari, 2002.

De Mauro, T., *Minisemantica*, Laterza, Roma-Bari, 2004.

De Waal, F., *Primates and Philosophers*, Princeton University Press, 2006; trad. it. di F. Conte, *Primati e filosofi. Evoluzione e moralità*, Garzanti, Milano, 2008.

Deacon, T. W., *The Symbolic Specie. The coevolution of Language and the Brain*, Norton & Company, New York, 1997; trad. it. di S. Ferraresi, *La specie simbolica. Coevoluzione di linguaggio e cervello*, Giovanni Fioriti, Roma, 2001.

Dennett, D. C., *The Intentional Stance*, MIT press, Cambridge, 1987; trad. it. Di E. Bassato, *L'atteggiamento Intenzionale*, Il Mulino, Bologna, 1993, p. 328.

Dessalles, J., *Altruism, status, and the origin of relevance*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy, (ed.), "Approaches to the Evolution of Language", Cambridge University Press, 1998, pp. 130-147.

Doolittle, E. L., *Other Species' Counterpoint. An investigation of the relationship between Human Music and Animal Song*, tesi di dottorato presentata alla University of Princeton, 2007.

Doupe, A. J., Kuhl, P. K., *Birdsong and human speech: common themes and mechanisms*, "Annual Review of Neuroscience", 22, 1999, pp. 567-631.

Estes, K. G., Evans, J. L., Alibali, M. W., Saffran, J. R., *Can Infants Map Meaning to Newly Segmented Words? Statistical Segmentation and Word Learning?*, "Psychological science", 18, 2007, pp. 254-260.

Dunbar, R. I. M., *The Origin and Subsequent Evolution of Language*, in M. H. Christiansen, S. Kirby (ed.), "Language evolution", Oxford University Press, 2004, pp. 219-234.

Dunbar, R. I. M., *The Human Story. A New History of Mankind's Evolution*, Faber and Faber Limited, London, 2004; trad. it. di D. Giusti, *La scimmia pensante. Storia dell'evoluzione umana*, Il Mulino, Bologna, 2009.

Dunbar, R. I. M., *Theory of mind and the evolution of language*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy, "Approaches to the Evolution of Language", Cambridge University Press, MA, 1998, pp. 92-110.

Fabre-Thorpe, M., *Visual categorization: accessing abstraction in non-human primates*, "Philosophical Transactions of the Royal Society B: Biological Sciences", 358, 2003, pp. 1215-1223

Falk, D., *Prelinguistic evolution in early hominins: Whence motherese?*, "Behavioral and Brain Sciences", 27, 2004, pp. 491-541.

Fernald, A., *Four-month-old infants prefer to listen to motherese*, "Infant Behavior and Development", 8, 1985, pp. 181-195.

Fernald, A., *Acoustic determinants of infant preference for motherese speech*, "Infant Behavior and Development", 10, 1987, pp. 279-293.

Fernald, A., Taeschner, T., & Dunn, J. *A cross-language study of prosodic modifications in mothers and fathers' speech to preverbal infants*, "Journal of child language", 16, 1989, pp. 477-501.

Fernald, A. *Intonation and communicative intent in mothers' speech to infants: Is the melody the message?*, "Child development", 60, 1989, pp. 1497-1510.

Fernald, A., *Prosody and focus in speech to infants and adults*, "Developmental Psychology", 27, 1991, pp. 209-221.

Fernald, A., *Meaningful melodies in mothers' speech to infants*, in H. Papousek, U. Jurgens, M. Papousek (ed.), "Nonverbal vocal communication: comparative and developmental approaches", Cambridge University Press, Cambridge, 1992, pp. 262-282.

Fernald, A., McRoberts, G., *Prosodic bootstrapping: A critical analysis of the argument and the evidence*, in J.L. Morgan, K. Demuth, "Signal to Syntax: from speech to grammar in early Acquisition", Lawrence Erlbaum Associates, NJ, 1996, pp. 365-387.

Ferretti, F., *Perché non siamo speciali*, Laterza, Roma-Bari, 2007.

Fiser, J., Aslin, R.N., *Statistical learning of new visual feature combinations by infants*, "Proceedings of the National Academy of Sciences", 99, 2002, pp. 15822-15826.

Fitch, W. T., *The evolution of speech: a comparative review*, "Trends in Cognitive Sciences", 4, 2000, pp. 258-267.

Fitch, W. T., Hauser, M. D., *Computational constraints on syntactic processing in a nonhuman primate*, "Science", 303, 2004, pp. 377-380.



Fitch, T. W., *Kin selection and “Mother Tongues”: A Neglected Component in Language Evolution*, in D. K. Oller, U. Griebel (ed.), “Evolution of Communication Systems: A Comparative Approach”, MIT Press, Cambridge, Massachusetts, 2004, pp. 276-96.

Fitch, W. T., Hauser, M. D., Chomsky, N., *The evolution of the language faculty: clarifications and implications*, “Cognition”, 97, 2005, pp. 179-210.

Fitch, W. T., *The evolution of language: a comparative review*, “Biology and Philosophy”, 20, 2005, pp. 193-230.

Fitch, W. T., *On the Biology and Evolution of Music*, “Music Perception”, 24, 2006, pp. 85-88.

Fitch W. T., *The biology and evolution of music: A comparative perspective*, “Cognition”, 100, 2006, pp. 173-215.

Fitch, W. T., *Evolving Meaning: The roles of Kin Selection, Allomothering, and Paternal Care in Language Evolution*, in C. Lyon, C. L. Nehaniv, A. Cangelosi (ed), “Emergence of Communication and Language”, Springer-Verlag, 2007, pp. 29-52

Fitch, W. T., *The evolution of language*, Cambridge University Press, 2010.

Fitch, W. T., Bugnyar, T., Huber, L., *Cognition and the Evolution of Language: Constructing Cognitive Phylogenies*, “Neuron”, 65, 2010, pp. 795-814.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., Koelsch, S., *Universal Recognition of Three Basic Emotions in Music*, “Current Biology”, 19, 2009, pp. 1–4.

Gargani, A. G., *Wittgenstein. Musica, parola, gesto*, Raffaello Cortina, Milano, 2008.

Gentner, T., *Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris**, “Animal Behaviour”, 56, 1998, pp. 579-594.

Gentner, T. Q., Fenn, K. M., Margoliash, D., Nusbaum, H. C., *Recursive syntactic pattern learning by songbirds*, "Nature", 440, 2006, pp. 1204–1207.

Geissmann, T., *Gibbon songs and human music from an evolutionary perspective*, in N. L. Wallin, B. Merker, S. Brown, (ed.), "The Origins of Music", Mit Press, Cambridge, 2001, pp. 103-123.

Geissmann, T., *The relationship between duet songs and pair bonds in siamangs, Hylobates syndactylus*, "Animal Behaviour", 2000, pp. 805-809.

Goldstein, M., King, A., *Social interaction shapes babbling: Testing parallels between birdsong and speech*, "Proceedings of the National Academy of Science", 100, 2003, pp. 8030-8035.

Gómez, R., *Infant artificial language learning and language acquisition*, "Trends in Cognitive Sciences", 4, 2000, pp. 178-186.

Hagen, E., *Music and Dance as a Coalition System*, "Human Nature", 14, 2003, pp. 21-50.

Hall, M., *A Review of Vocal Duetting in Birds*, "Advances in the Study of Behavior", 2009, pp. 67-120.

Hauser, M. D., Newport, E. L., Aslin, R. N., *Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins*, "Cognition", 78, 2001, pp. B53 - B64.

Hauser, M. D., Weiss, D., & Marcus, G., *Rule learning by cotton-top tamarins*, Cognition, 86, 2002, pp. B15–22.

Hauser, M. D., Chomsky, N., Fitch, W. T., *The faculty of language: what is it, who has it, and how did it evolve?*, "Science", 298, 2002, pp 1569-79.

Heidegger, M., *Die Grundbegriffe der Metaphysic. Welt – Endlichkeit – Einsamkeit*, Klostermann Verlag, Frankfurt ma Main, 1983; trad. it di P. Coriando, *Concetti fondamentali della metafisica. Mondo – Finitezza – Solitudine*, Il Melangolo, Genova, 1992.

Heijningen, van, C., Visser, de, J., Zuidema, W., Cate, de, C., *Simple rules can explain discrimination of putative recursive syntactic structures by a*

*songbird species*, “Proceedings of the National Academy of Sciences”, 106., 2009, pp. 20538-20543.

Heimbauer, L., Beran, M., *A Chimpanzee Recognizes Synthetic Speech with Significantly Reduced Acoustic Cues to Phonetic Content*, “Current Biology”, 21, 2011, pp. 1210-1214.

Heinz, J., Idsardi, W., *Sentence and Word Complexity*, “Science”, 333, 2011, pp. 295–297.

Heugten, van, M., Johnson, E. K., *Linking infants' distributional learning abilities to natural language acquisition*, “Journal of Memory and Language”, 63, 2010, pp. 197-209.

Hewes, G. W., *Primate Communication and the gestural origin of language*, “Current Anthropology”, 14, 1973, pp. 5-24

Hulse, S. H., Bernard D. J., Braten, R. F., 1995, *Auditory discrimination of chord-based spectral structures by European starlings (Sturnus vulgaris)*, “Journal of Experimental Psychology: General”, 124, 1995, pp. 409-423.

Hunyadi, L. *Cognitive grouping and recursion in prosody*, in H. van der Hulst, J. Koster, H. van Riemsdij (ed.), “Recursion and Human Language. Studies in Generative Grammar”, Berlin/New York: De Gruyter Mouton, 2010, pp. 343-370

Ilie, G., Thompson, F., *A comparison of acoustic cues in music and speech for three dimensions of affect*, “Music Perception”, 23, 2006, pp. 319-329.

Izumi, A., *Japanese monkeys perceive sensory consonance of chords*, “The Journal of the Acoustical Society of America”, 108, 2000, pp. 3073-3078.

Jackendoff, R., *Parallels and Nonparallels between Language and Music*, “*Music Perception*”, 26, 2009, pp. 195-204.

Jespersen, O., *Language: Its nature, development and origin*, W.W. Norton & Co., New York, 1922.

Johnson, E. K., Jusczyk, P.W., *Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics*, “Journal Of Memory And Language”, 44, 2001, pp. 548-567.

Johnson, E. K., *Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech*, “The Journal of the Acoustical Society of America”, 123, 2008, pp. EL144–EL148

Johnson, E. K., Seidl A. H., *At 11 month, prosody still outranks statistics*, “Developmental Science”, 12, 2009, pp. 131-141.

Juslin, P. N., Laukka, P., *Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?*, “Psychological Bulletin”, 129, 2003, pp. 770–814.

Kedar, Y., Casasola, M., Lust, B., *Getting there faster: 18- and 24-month-old infants' use of function word to determine reference*, “Child Development”, 77, 2006, pp. 325-338.

Kirschner, S., Tomasello, M., *Joint music making promotes prosocial behavior in 4-year-old children*, “Evolution and Human Behavior”, 31, 2010, pp. 354–364.

Klammer, T., Schulz, M: R., Della Volpe, A., *Analyzing English Grammar*, Longman, 2009, in [http://en.wikipedia.org/wiki/Function\\_word#cite\\_ref-0](http://en.wikipedia.org/wiki/Function_word#cite_ref-0)

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D., *Music, language and meaning: brain signatures of semantic processing*, “Nature Neuroscience”, 7, 2004, pp. 302–307.

Kuhl, P. K., *Human speech and birdsong: Communication and the social brain*, “Proceedings of the National Academy of Science”, 100, 2003, pp. 9645-9646.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I, Ulla Sundberg, U., Lacerda, F., *Cross-Language Analysis of Phonetic Units in Language Addressed to Infants*, “Science”, 277, 1997, pp. 684–686

Langmore, N., *Functions of duet and solo songs of female birds*, “Trends in Ecology & Evolution”, 13, 1998, pp. 136-140.

Livingstone, R. S., Thompson, W. F., *The emergence of music from the Theory of Mind*, “Musicae Scientiae”, 2009, pp. 83-115.

Lo Piparo, F., *Il mondo, le specie animali e il linguaggio. la teoria zocognitiva di Wittgenstein*, in M. Carenini e M. Matteuzzi (a cura di), *Percezione, Linguaggio, Coscienza. Saggi di filosofia della mente*, Quodlibet, Macerata, 1998.

Lo Piparo F., *Aristotele e il linguaggio. cosa fa di una lingua una lingua*, Laterza, Roma-Bari, 2003.

Maraini, F., *Gnosi delle fanfole*, Baldini e Castoldi, 1994.

Marcus, G., Vijayan, S., Rao, S. B., Vishton, P. M., *Rule Learning by Seven-Month-Old Infants*, “Science”, 283, 1999, pp. 77-79.

Marcus, G. F., *Reply to Christiansen and Curtin*, “Trends in Cognitive Sciences”, 3, 1999, pp. 290–291.

Marler, P., *Origins of Music and Speech: Insights from Animals*, in N. L. Wallin, B. Merker, S. Brown, (ed.), “The Origins of Music”, Mit Press, Cambridge, 2001, pp. 31-48.

Maslow, A. H., *The Psychology of Science*, Henry Regnery Co, 1970, pp. 15-16.

McDermott, J., Hauser, M. D., *The origins of music: Innateness, uniqueness, and evolution*, “Music Perception: An Interdisciplinary Journal”, 23, 2005, pp. 29-59.

Medinaa, T. N., Snedekerc, J., Trueswella, J. C., & Gleitmana, L. R., *How words can and cannot be learned by observation*, “Proceedings of the National Academy of Sciences”, 108, 2011, pp. 9015–9019.

Millotte, S., Wales, R., Christophe, A., *Phrasal prosody disambiguates syntax*, “Language and Cognitive Processes”, 22, 2007, pp. 898–909.

Mithen, S., *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*, Harvard University Press, 2007.

Moro, A., *I confini di Babele*, Longanesi, Milano, 2006.

Nakata, T., *Infants' responsiveness to maternal speech and singing*, "Infant Behavior and Development", 27, 2004, pp. 455-464.

Nuebauer, R., *Super-normal length song preferences of female zebra finches (Taeniopygia guttata) and a theory of the evolution of bird song*, "Evolutionary Ecology", 13, 1999, pp. 365-380.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N., *Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate*, "Cognitive Psychology", 49, 2004, pp. 85-117.

O'Donnell, T. J., Hauser, M. D., & Fitch, W. T., *Using mathematical models of language experimentally*, "Trends in Cognitive Sciences", 9, 2005, pp. 284-289.

Okanoya, K., *Sexual Display as a Syntactical Vehicle: The Evolution of Syntax in Birdsong and Human Language through Sexual Selection*, in A. Wray, "The Transition to Language", Oxford University Press, 2002, pp. 46-63.

Okanoya, K., *The Bengalese Finch: A Window on the Behavioral Neurobiology of Birdsong Syntax*, "Annals of the New York Academy of Sciences", 1016, 2004a, pp. 724-735.

Okanoya, K., *Song Syntax in Bengalese Finches: Proximate and Ultimate Analyses*, "Advances in the Study of Behavior", 14, 2004b, pp. 297-346.

Ouattara, K., Lemasson, A., Zuberbühler, K., *Campbell's Monkeys Use Affixation to Alter Call Meaning*, "PLoS ONE", 4, 2009, pp. 1-7.

Ouattara, K., Lemasson, A., Zuberbühler, K., *Campbell's monkeys concatenate vocalizations into context-specific call sequences*, "Proceedings of the National Academy of Sciences", 106, 2009, pp. 1-6.

Patel, A. D., "Music, Language, and the Brain", Oxford, Oxford University Press, 2008.

Patel, A., *Language, music, syntax and the brain*, "Nature Neuroscience", 6, 2003, pp. 674-681.

Patel, A., Iversen, J., Bregman, M., *Experimental Evidence for*

*Synchronization to a Musical Beat in a Nonhuman Animal*, "Current Biology", 19, 2009, pp. 827-830.

Patel, A., *Music, biological evolution, and the brain*, in M. Bailar (Ed.), "Emerging Disciplines", Rice University Press, Houston, 2010, pp. 91-144.

Pena, M., Bonatti, L. L., Nespor, M., Mehler, J. *Signal-Driven Computations in Speech Processing*, "Science", 298, 2002, pp. 604-607.

Perruchet, P., Rey, A., *Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates*, "Psychonomic Bulletin & Review", 12, 2005, pp. 307-313.

Pinker, S., *The Language Instinct*, 1994, trad. it di G. Origgi, *L'istinto del linguaggio*, Arnoldo Mondadori, Milano, 1997.

Pinker, S., *How the Mind works*, Penguin Books, New York, 1997.

Pinker S., *Language as an Adaptation to the Cognitive Niche*, in M. H. Christiansen M. H., S. Kirby S. (ed.), "Language evolution", Oxford University Press, 2004, pp. 16-37.

Power C., *Old wives' tales: the gossip hypothesis and the reliability of cheap signals*, in C. Knight C., J. R. Hurford, M. Studdert-Kennedy, "Approaches to the Evolution of Language", Cambridge University Press, MA, 1998, pp. 111-129.

Premack, D., *Does the chimpanzee have a theory of mind?*, "Behavioral and Brain Sciences", 4, 1978, pp. 515-526.

Quine, W.V.O., *Word and Object*, Mit Press, Cambridge, 1960, pp. 26-45.  
Ramus, F., Hauser, M. D., Miller, C., Morris, D., Mehler, J., *Language discrimination by human newborns and by cotton-top tamarin monkeys*, "Science", 288, 2000, pp. 349-351.

Rogers, J., & Hauser, M. D., *The use of formal language theory in studies of artificial language learning: A proposal for distinguishing the differences between human and nonhuman animal learners*, in H. van der Hulst (ed.), "Recursion and Human Language", Gruyter, Walter De Gmbh, Berlin/New York, 2010, pp. 213-232.

Rousseau, J. J., *Essai sur l'origine des langues*, Gallimard, Paris; trad. it. di G. Gentile, *Saggio sull'origine delle lingue*, Guida, Napoli, 1984.

Russell, B., *An outline of Philosophy*, Allen & Unwin, London 1927, in Egner R. E., Dennon L. E., (ed.), *The Basic Writings of Bertrand Russell*, Routledge Classics, London and New York, 2010.

Saffran, J. R., Newport, E., Aslin, R. N., *Word segmentation: The role of distributional cues*, "Journal of Memory and Language", 35, 1996, pp. 606-621.

Saffran, J. R., Aslin, R. N., & Newport, E. L., *Statistical learning by 8-month-old infants*, "Science", 274, 1996, pp. 1926–1928.

Saffran, J. R., Johnson, E., Aslin, R. N., *Statistical learning of tone sequences by human infants and adults*, "Cognition", 70, 1999, pp. 27-52

Searcy, W. A., Andersson, M., *Sexual selection and the evolution of song*, "Annual Review of Ecology and Systematics", 17, 1986, pp. 507-533.

Seidenberg, M. S., *Do infants learn grammar with algebra or statistics?*, "Science", 284, 1999, p. 433.

Seidenberg, M. S., MacDonald, M., *Does grammar start where statistics stop?*, "Science", 298, 2002, p. 554.

Seyfarth, R., Cheney, D., Marler, P., *Vervet monkey alarm calls: Semantic communication in a free-ranging primate*, "Animal Behaviour", 28, 1980, pp. 1070–1094.

Shannon, R., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M., *Speech Recognition with Primarily Temporal Cues*, "Science", 270, 1995, 270, pp. 303-304.

Scherer, K. L., *Expression of emotion in voice and music*, "Journal of Voice", 9, 1995, pp. 235-248.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R., *Songs as an aid for language acquisition*, "Cognition", 106, 2008, pp. 975–983.



Shi, R., Lepage, M., *The effect of functional morphemes on word segmentation in preverbal infants*, “Developmental Science”, 11, 2008, pp. 407-413.

Shukla, M., Nespor, M., *An interaction between prosody and statistics in the segmentation of fluent speech*, “Cognitive psychology”, 54, 2007, pp. 1-32.

Shukla, M., White, K. S., Aslin, R. N., *Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants*, “Proceedings of the National Academy of Sciences”, 108, 2011, pp. 6038–6043.

Siegel, D. J., *The Developing Mind*, The Guilford Press Inc., 1999, trad. it. di L. Madeddu, *La mente relazionale. Neurobiologia dell'esperienza interpersonale*, Raffaello Cortina, Milano, 2001.

Smith, L., Chen, Y., *Infants rapidly learn word-referent mappings via cross-situational statistics*, “Cognition”, 106, 2008, pp. 1558-1568.

Soderstrom, M., Seidl, A., Nelson, D. G. K., Jusczyk, D. W., *The prosodic bootstrapping of phrases: Evidence from prelinguistic infants*, “Journal of Memory and Language”, 49, 2003, pp. 249–267.

Soma, M., Hiraiwa-Hasegawa, M., Okanoya, K., *Song-learning strategies in the Bengalese finch: do chicks choose tutors based on song complexity*, “Animal Behaviour”, 78, 2009, pp. 1107-1113.

Strawson, P., *On Referring*, “Mind”, 1950, in P. Strawson, *Logico-Linguistic Papers*, Methuen, London, 1971; trad. it.: *Sul riferimento*, in A. Bonomi, *La Struttura logica del linguaggio*, Valentino Bompiani, 2001.

Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriya, Y., Takeshita, H., Nakamura, T., Hashiya, K., *Preference for consonant music over dissonant music by an infant chimpanzee*, “Primates”, 51, 2009, pp. 7–12.

Takahasi, M., Yamada, H., & Okanoya, K. *Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica)*, “Ethology”, 116, 2010, pp. 481–489.

Thiessen, E. D., Saffran, J.R., *When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants*, “Developmental Psychology”, 39, 2003, pp. 706-716.

Thiessen, E., Hill, E., Saffran, J. R., *Infant-directed speech facilitates word segmentation*, “Infancy”, 7, 2005, pp. 53-71.

Tomasello, M., *The cultural origins of human cognition*, Harvard University Press, Cambridge, Mass, 1999; trad. it. di M. Ricucci, *Le origini culturali della cognizione umana*, Il Mulino, Bologna, 2005, pp. 48-49.

Tomasello, M. *Does the chimpanzee have a theory of mind? 30 years later*, “Trends in Cognitive Sciences”, 2008, pp. 187-192.

Tomasello, M., *Origins of Human Communication*, Massachusetts Institute of Technology, 2008; trad. it. di S. Romano, *Le origini della comunicazione umana*, Raffaello Cortina, Milano, 2009.

Toro, J, Trobalon, J., Sebastian-Gallés, N., *The use of prosodic cues in language discrimination tasks by rats*, “Animal Cognition”, 6, 2003, pp. 131-136.

Toro, J. M., Nespors, M., Mehler, J., Bonatti, L. L., *Finding Words and Rules in a Speech Stream*, “Psychological science”, 19, 2008, pp. 137-144.

Trainor, L. J., Austin, C. M., Desjardins, R. N., *Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion?*, “Psychological science”, 11, 2000, pp. 188–195.

Trainor, L. J., Desjardin R. N., *Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels*, “Psychonomic Bulletin & Review”, 2002, 9, pp. 335-340

Trehub, S. E., *The developmental origins of musicality*, “Nature Neuroscience”, 6, 2003, pp. 669-673.

Uexküll, J. von e Kriszat, G., *Streifzüge durch die Umwelten von Tieren und Menschen: Ein Bilderbuch unsichtbarer Welten*, 1934; trad. it. di P. Manfredi, *I mondi invisibili*, 1936, rist. *Ambiente e comportamento*, Il Saggiatore, Milano, 1967.

Ulbaek, I., *The origin of language and cognition*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy (ed.), "Approaches to the Evolution of Language", Cambridge University Press, MA, 1998, pp. 30-43.

Wales, R., Dupoux, E., Christophe, A., *Can prosodic cues and function words guide syntactic processing and acquisition*, in R. Hofmann, H. Mixdorff (ed.), "Speech prosody: 3rd International Conference", Dresden TUD Press, 2006.

Wittgenstein, L., *Tractatus logico-philosophicus*, Routledge and Kegan Paul, Oxford, 1961, trad. it. di A. G. Conte, *Tractatus logico-philosophicus*, Einaudi, Torino, 1998.

Wittgenstein, L., *On certainty*, Basil Blackwell, Oxford, 1969; trad. it. di M. Trinchero, *Della certezza. L'analisi filosofica del senso comune*, Einaudi, Torino, 1999.

Wittgenstein L., *Philosophische Untersuchungen*, Basil Blackwell, Oxford, 1953; trad. it. di M. Trinchero, *Ricerche filosofiche*, Einaudi, Torino, 1999.

Wittgenstein, L., *The Big Typescript*, Springer-Verlag, Vienna, 2000; trad. it. di A. De Palma, *The Big Typescript*, Einaudi, Torino, 2002.

Wong, P., Skoe, E., & Russo, N., *Musical experience shapes human brainstem encoding of linguistic pitch patterns*, "Nature Neuroscience", 2007, pp. 1-2.

Wray, A., *Protolanguage as a holistic system for social interaction*, "Language & Communication", 18, 1998, pp. 47-67.

Wray, A., *Holistic utterances in protolanguage: the link from primates to humans*, in C. Knight, J. R. Hurford, M. Studdert-Kennedy, M. (ed.), "The Evolutionary emergence of language", Cambridge University Press, 2000, pp. 285-302.

Wray, A., *Dual Processing in Protolanguage: Performance without Competence*, in A. Wray, (ed.), "The Transition to Language", Oxford University Press, 2002, pp. 113-37.

Wright, A., Rivera, J., Hulse, S., Neiworth, J. J., *Music perception and octave generalization in rhesus monkeys*, “Journal of Experimental Psychology”, 129, 2000, pp. 291-307.

Yang, C., *Universal Grammar, statistics or both?*, “Trends in Cognitive Sciences”, 8, 2004, pp. 451-456.

Yu, C., Smith, L. B. *Rapid word learning under uncertainty via cross-situational statistics*, “Psychological science”, 18, 2007, pp. 414–420.

Zatorre, R. J., Belin, P., & Penhune, V. B., *Structure and function of auditory cortex: music and speech*, “Trends in Cognitive Sciences”, 6, 2002, pp. 37–46.

Zuberbühler, K., *A syntactic rule in forest monkey communication*, “Animal Behaviour”, 63, 2002, pp. 293–299