Linguaggio.

Dalla biologia all'IA

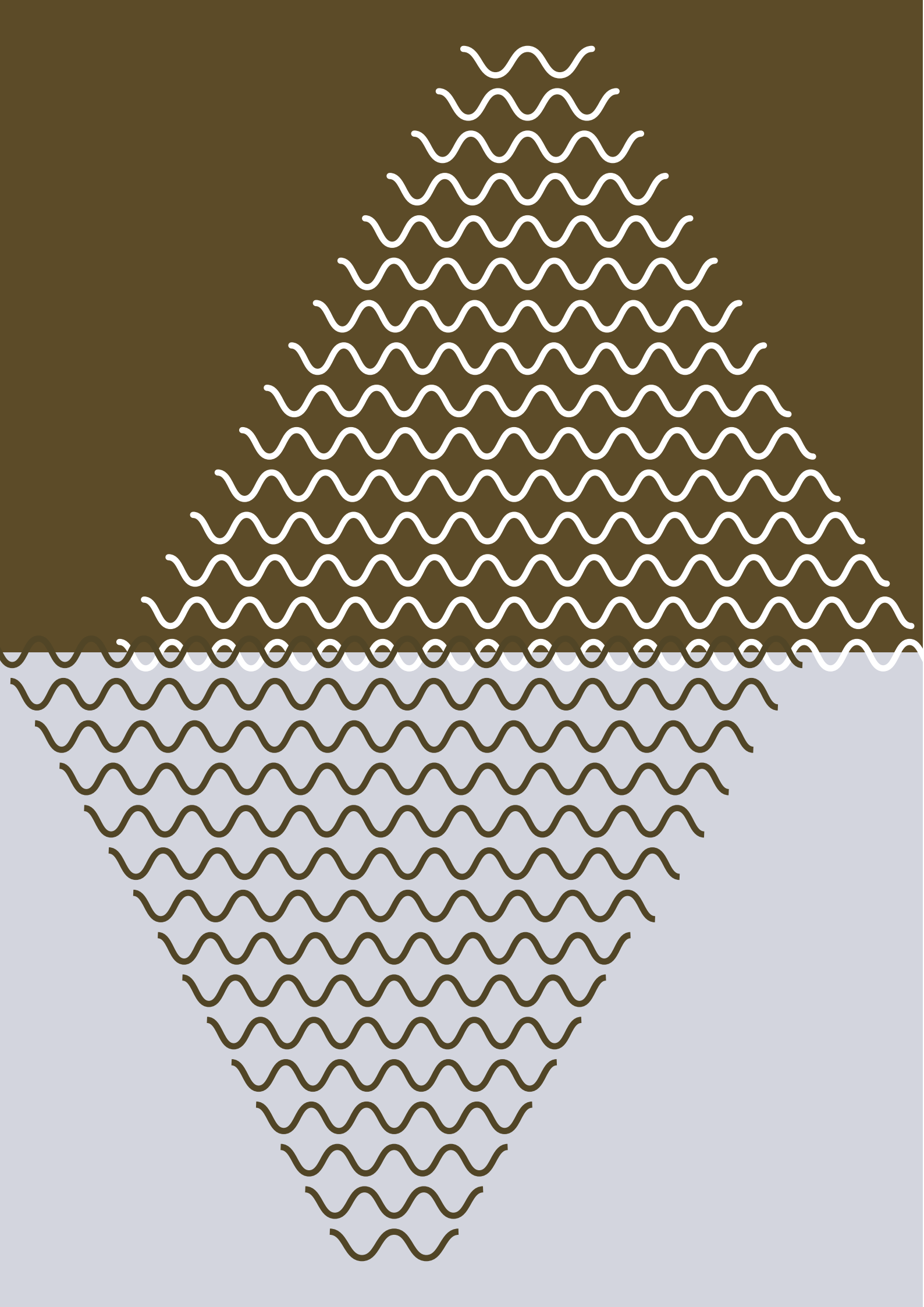


a cura di Damiano Cantone

II/2026
ISSN: 2385-1945

Philosophy Kitchen #25

P
K



Linguaggio.

Dalla biologia all'IA

a cura di Damiano Cantone

7 STATEMENT

EDITORIALE
9 Damiano Cantone

PARTE I

17 **L'evoluzione del linguaggio: un'indagine paleoneurologica**
Fabio Macciardi
Viola Castellani
Fabio Martini

33 **La natura simbolica e combinatoria dell'informazione biologica. A livello genetico, neuropsicologico e linguistico**
Franco Fabbro

47 **Natura e artificio. Il Linguaggio come Meta-Istituzione e Prassi sociale trasformativa**
Nicola Garau

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026

59 **Artificial Language and Human Cognition. Towards a New Computational Functionalism?**
Francesco Maria Cianciaruso

73 **San Francesco e l'IA. Dalla zoosemiotica all'intelligenza artificiale relazionale**
Daria Arkhipova
Nicola Zengiaro

PARTE II

93 **Why Stories Enchant the Brain. Splendors and Miseries of an Extraordinary Human Faculty**
Francesco Ferretti

107 **Alice in Worderland. The Language(s) of Architecture in the Age of AI**
Carlo Deregibus

125 **Sulla tecnologia del linguaggio e il suo futuro**
Damiano Cantone

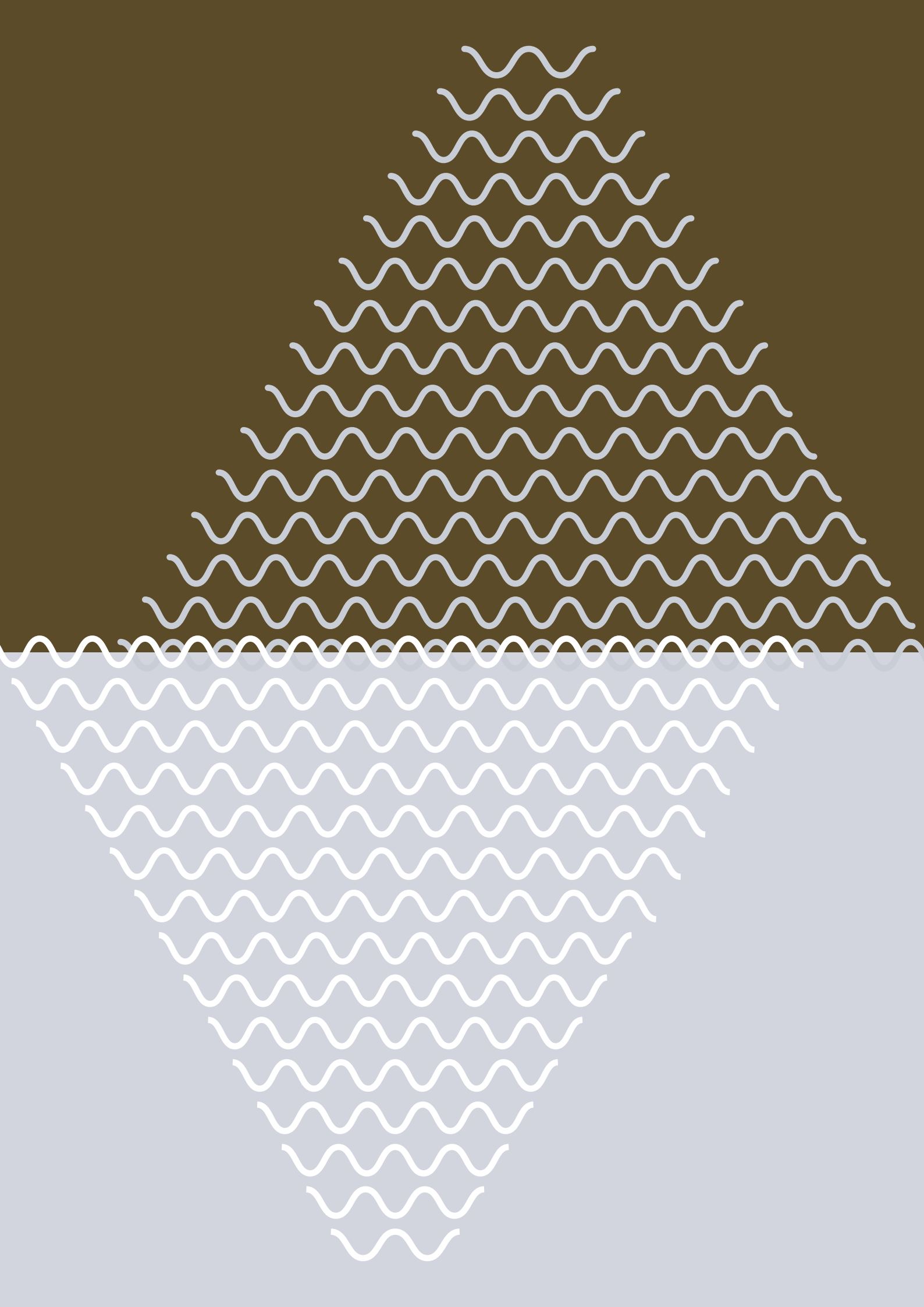
PARTE III

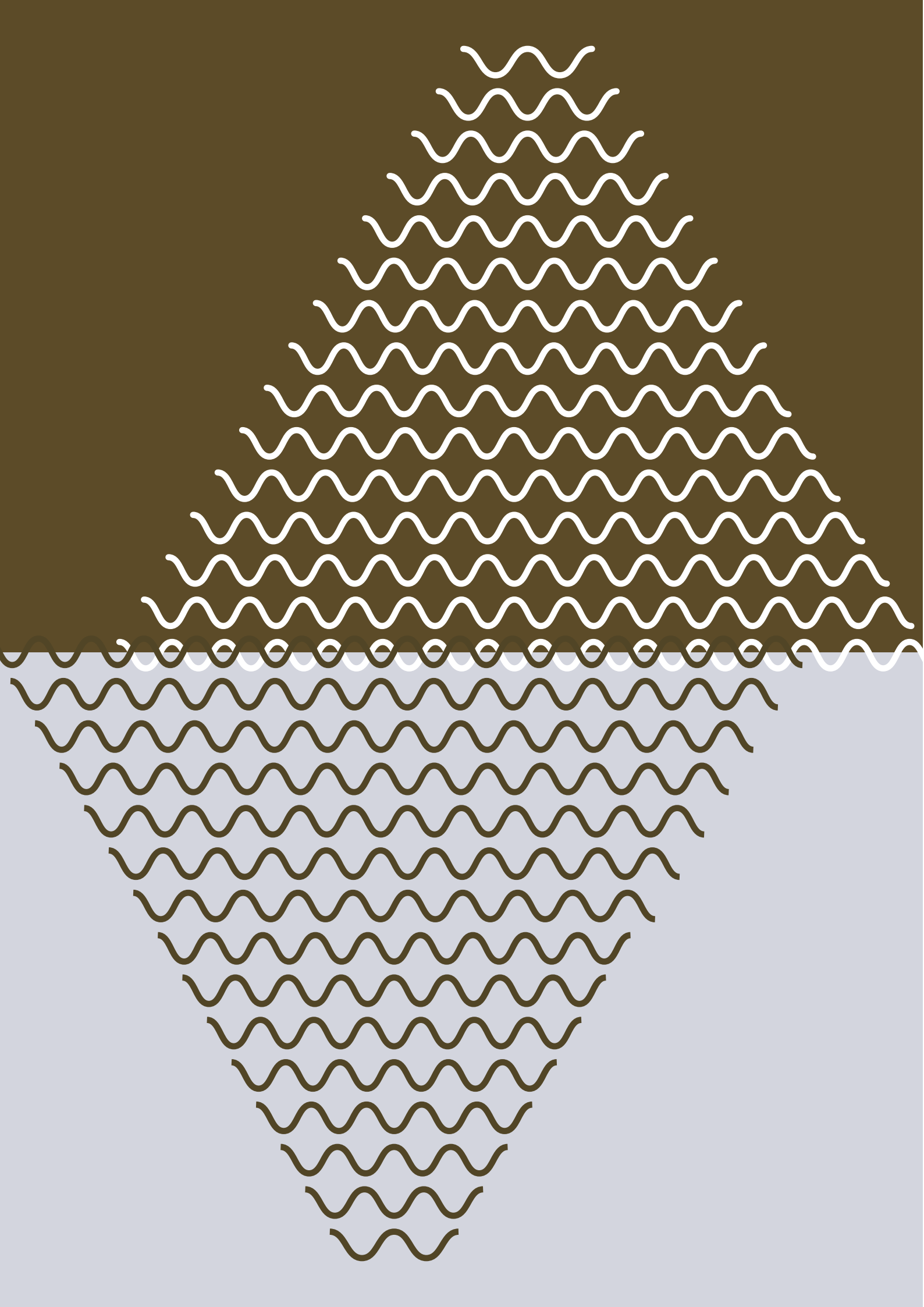
141 **Continuity under Variation. Radical Absence and the Form of Language**
Filippo Scafi

157 **Words and Vectors: The Distributed View of Language**
Niccolò Monti

171 **LLMs and Normativity. A View From Languaging**
Filippo Batisti

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026,





Nell’ambito del senso comune si fa ancora fatica a liberarsi da quel pensiero binario che poneva da una parte ciò che è naturale e dall’altra ciò che è artificiale. Tuttavia, se escludiamo quei filosofi che di solito non amano dialogare con le altre discipline presenti nell’enciclopedia, nell’ambito della riflessione scientifica e filosofica da tempo ormai si è preso atto del fatto che tale separazione non tiene. Proprio lo studio del linguaggio lo dimostra in modo esemplare. Per lungo tempo posto quale elemento centrale dell’universo discorsivo che ospitava non solo la differenza tra il naturale e l’artificiale, ma anche quella, conseguente, tra l’animale umano e l’animale di altre specie, il linguaggio ora si pone come punto di incrocio tra sguardi disciplinari diversi, che in comune hanno la vocazione a interrogare la matrice sistemico-evolutiva da cui scaturisce la facoltà di servirsi del linguaggio per costruire la nicchia abitata da Homo (e non solo sapiens). Animale protesico, l’animale parlante costruisce se stesso nel mentre forgia gli artefatti di cui si munisce, finendo poi con l’istituire legami di dipendenza dagli artefatti stessi. In tale prospettiva, il presente fascicolo di PK ha voluto non solo interrogare l’origine del linguaggio, ma anche il modo in cui i linguaggi artificiali contemporanei si inseriscano entro la lunga storia delle modificazioni che l’animale parlante apporta ai propri mondi.

Giovanni Leghissa per Philosophy Kitchen

Damiano Cantone
damiano.cantone@uniud.it

Che cos'è il linguaggio? Una facoltà biologica, un sistema simbolico, una tecnologia, una pratica sociale? E, soprattutto, fino a che punto queste diverse definizioni possono essere considerate separatamente? La riflessione contemporanea sul linguaggio sembra rendere sempre meno praticabile l'idea di poter isolare un unico livello di analisi capace di abbracciarle tutte. Le acquisizioni delle neuroscienze e della biologia evolutiva, lo sviluppo delle prospettive *embodied* e situate, il rinnovamento della biolinguistica e, più recentemente, la diffusione dei sistemi di intelligenza artificiale capaci di produrre e manipolare linguaggio naturale hanno contribuito a riaprire questioni che sembravano, almeno in parte, stabilizzate. È da questa constatazione che ha preso avvio il presente numero di *Philosophy Kitchen*. La call for papers da cui esso nasce proponeva di spostare l'attenzione dal funzionamento interno del linguaggio alla complessa relazione tra il livello biologico, psichico e sociale che il linguaggio contribuisce a costruire e dal quale, al tempo stesso, dipende. Le prospettive dell'*Embodied Cognition* e gli sviluppi delle neuroscienze hanno infatti progressivamente messo in discussione rappresentazioni eccessivamente localizzazioniste o modulari delle funzioni linguistiche; parallelamente, gli studi sull'evoluzione hanno riaperto il problema dell'origine del linguaggio e della sua relazione con l'evoluzione biologica e culturale della specie umana.

A questo quadro già complesso si è aggiunta negli ultimi anni una trasformazione tecnologica difficilmente eludibile. Il rapido sviluppo dei *Large Language Models* ha riportato al centro del dibattito interrogativi fondamentali che riguardano cosa distingue il linguaggio umano da quello

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 9 – 12

prodotto dalle macchine, quale rapporto esiste tra competenza linguistica e comprensione e se è possibile separare il linguaggio dal corpo, dall'esperienza e dall'interazione sociale. Non è più chiaro, insomma, se ciò che chiamiamo "linguaggio artificiale" rappresenta davvero una forma radicalmente altra rispetto al linguaggio umano oppure può aiutarci, proprio per differenza, a comprendere meglio la natura di quest'ultimo. In questa prospettiva, l'intelligenza artificiale costituisce un nuovo oggetto di indagine e al contempo una sorta di banco di prova teorico: rende nuovamente visibili problemi relativi al rapporto tra sintassi e semantica, informazione e significato, simbolo e referente, linguaggio e corpo, competenza e interazione. Il titolo scelto per questo numero, *Il linguaggio dalla biologia all'intelligenza artificiale*, non indica pertanto una traiettoria lineare nella quale il biologico venga progressivamente sostituito dall'artificiale. Il "dalla… all'…" va inteso piuttosto come l'indicazione di un campo di tensioni e di passaggi, Tra i cui estremi si collocano infatti il cervello e la cognizione, l'evoluzione e la cultura, la dimensione simbolica, la narrazione, le istituzioni e le pratiche sociali, le altre specie animali, le tecnologie e, infine, i sistemi computazionali. È precisamente la possibilità di attraversare questi livelli senza ridurre l'uno all'altro a costituire il presupposto metodologico che accomuna i contributi raccolti nel volume.

La struttura del numero riflette questo percorso. La prima parte affronta le condizioni biologiche, evolutive, cognitive e sociali che rendono possibile il linguaggio. L'apertura è affidata a Fabio Macciardi, Viola Castellani e Fabio Martini con *L'evoluzione del linguaggio: un'indagine paleoneurologica*, che ricostruisce il problema dell'emergere del linguaggio attraverso il confronto tra dati paleoneurologici, genetici e archeologici. La prospettiva evolutiva consente di sottrarsi tanto all'idea di un'origine improvvisa quanto alla ricerca di un singolo fattore esplicativo, restituendo invece la complessità dell'interazione tra trasformazioni biologiche, organizzazione cerebrale, capacità cognitive e cultura materiale. Franco Fabbro, in *La natura simbolica e combinatoria dell'informazione biologica. A livello genetico, neuropsicologico e linguistico*, amplia ulteriormente la scala temporale e concettuale dell'indagine. Il linguaggio viene inserito nella storia dell'informazione biologica, dalle strutture genetiche ai sistemi neuronali e alla dimensione psichica. Ciò che emerge è la presenza, a livelli profondamente differenti, di proprietà simboliche e combinatorie: il linguaggio umano diviene così comprensibile soltanto all'interno di una storia biologica e cognitiva più ampia, pur senza essere riducibile a essa. È proprio contro ogni possibile riduzionismo che si colloca il contributo di Nicola Garau, *Natura e artificio: il linguaggio come meta-istituzione e prassi sociale trasformativa*. Il passaggio dal biologico al sociale non viene concepito come una cesura, ma come l'emergere di una dimensione nella quale il linguaggio non si limita a rappresentare una realtà già costituita: attraverso le pratiche linguistiche gli esseri umani producono, stabilizzano e trasformano significati e istituzioni. Il linguaggio appare pertanto come una tecnologia e una pratica costitutivamente sociale. Chiude questa prima sezione Francesco Maria Cianciaruso con *Artificial Language and Human Cognition. Towards a New Computational Functionalism?*. Il confronto con le reti neurali artificiali introduce già il secondo polo tematico del numero. L'artificiale occasione per ripensare i modelli della cognizione e

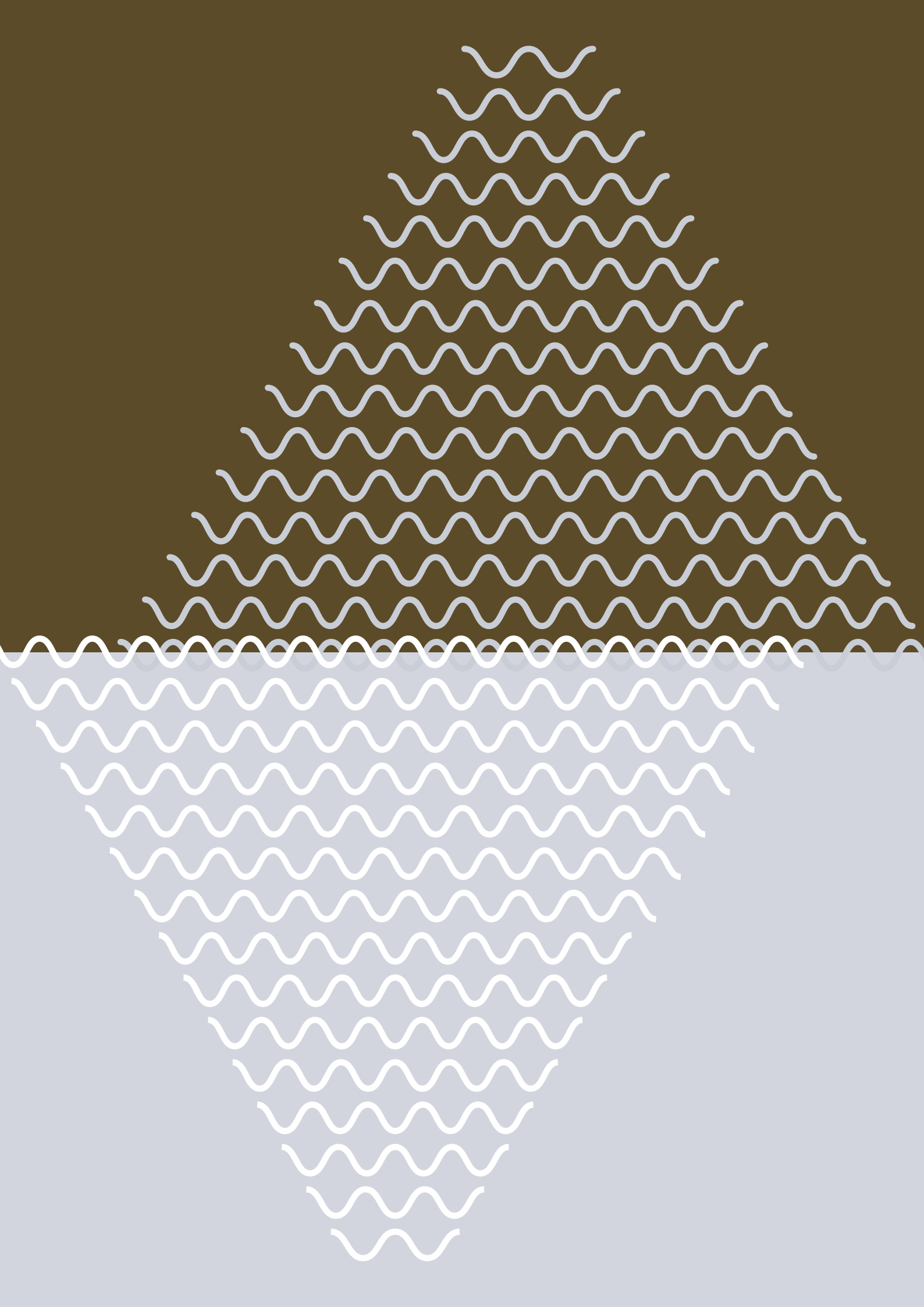
le relazioni tra rappresentazione simbolica, apprendimento distribuito e dimensione incarnata.

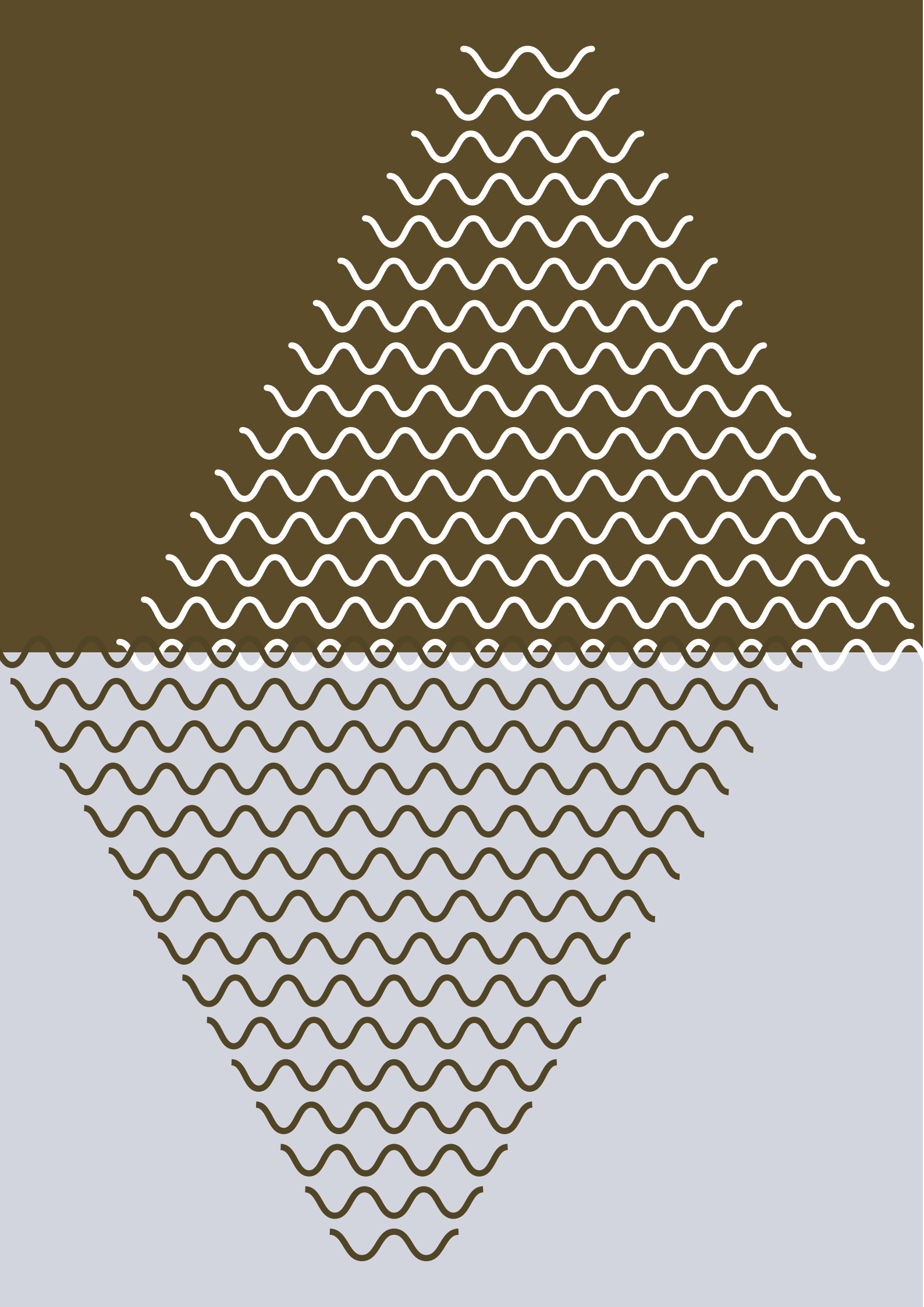
La seconda parte amplia e, in un certo senso, mette alla prova i confini del linguaggio. Se nella prima sezione la domanda riguardava soprattutto le condizioni che rendono possibile il linguaggio umano, qui il problema diventa che cosa accada quando il linguaggio viene osservato nei suoi punti di contatto con altre specie, con la narrazione, con le tecnologie e con differenti sistemi di produzione simbolica. Daria Arkhipova e Nicola Zengiaro, in *San Francesco e l'IA: dalla zoosemiotica all'intelligenza artificiale relazionale*, interrogano criticamente l'applicazione dell'intelligenza artificiale alla comunicazione animale. Il tentativo di "tradurre" le vocalizzazioni di altre specie rende evidente il rischio di assumere il linguaggio umano come modello universale della comunicazione. La prospettiva biosemiotica suggerisce invece una concezione relazionale, contestuale e multisensoriale della semiosi, nella quale l'IA può essere ripensata non come dispositivo universale di decodifica, ma come tecnologia di mediazione. Con *Why Stories Enchant the Brain. Splendors and Miseries of an Extraordinary Human Faculty*, Francesco Ferretti affronta un'altra dimensione fondamentale della specificità umana: la capacità narrativa. Il rapporto tra linguaggio e narrazione viene ricostruito in chiave evolutiva e cognitiva, mostrando come il pensiero narrativo coinvolga capacità di proiezione nel tempo, nello spazio e nella mente altrui. Ma il potere delle storie introduce anche la dimensione pragmatica del linguaggio: la capacità di persuadere, di produrre cooperazione e identità condivise può trasformarsi in strumento di manipolazione. Il linguaggio come dispositivo operativo emerge con particolare evidenza nel contributo di Carlo Deregibus, *The Language(s) of Architecture in the Age of AI*. Il caso dell'architettura mostra come parola e rappresentazione visiva possano entrare, attraverso l'intelligenza artificiale generativa, in una relazione nuova. Il linguaggio non serve più soltanto a descrivere un progetto: prompt, parole chiave e processi iterativi diventano strumenti capaci di concorrere alla produzione della forma stessa. Il confine tra linguaggio, rappresentazione e azione si fa così sempre più poroso. La sezione si conclude con *Sulla tecnologia del linguaggio e il suo futuro* di Damiano Cantone. Considerare il linguaggio stesso come una tecnologia consente di inserire le innovazioni contemporanee nella storia molto più lunga delle tecnologie cognitive umane. Dagli LLM alle interfacce cervello-macchina e cervello-cervello, il problema diventa allora quello delle possibili trasformazioni della mediazione simbolica e, in ultima analisi, del futuro stesso del linguaggio.

La terza parte concentra infine l'attenzione sul confronto tra linguaggio umano, modelli computazionali e Large Language Models. È qui che alcuni interrogativi disseminati nelle sezioni precedenti diventano espliciti: che cosa significa comprendere? Da dove deriva la coerenza di un sistema linguistico? Quale rapporto esiste tra distribuzione statistica e significato? E che cosa accade alla normatività propria delle pratiche linguistiche quando la produzione di enunciati viene affidata a sistemi artificiali? Filippo Scafi, in *Continuity under Variation: Radical Absence and the Form of Language*, affronta il problema distinguendo la coerenza strutturale dalla comprensione. La capacità dei modelli linguistici di preservare determinate relazioni anche in condizioni di variazione

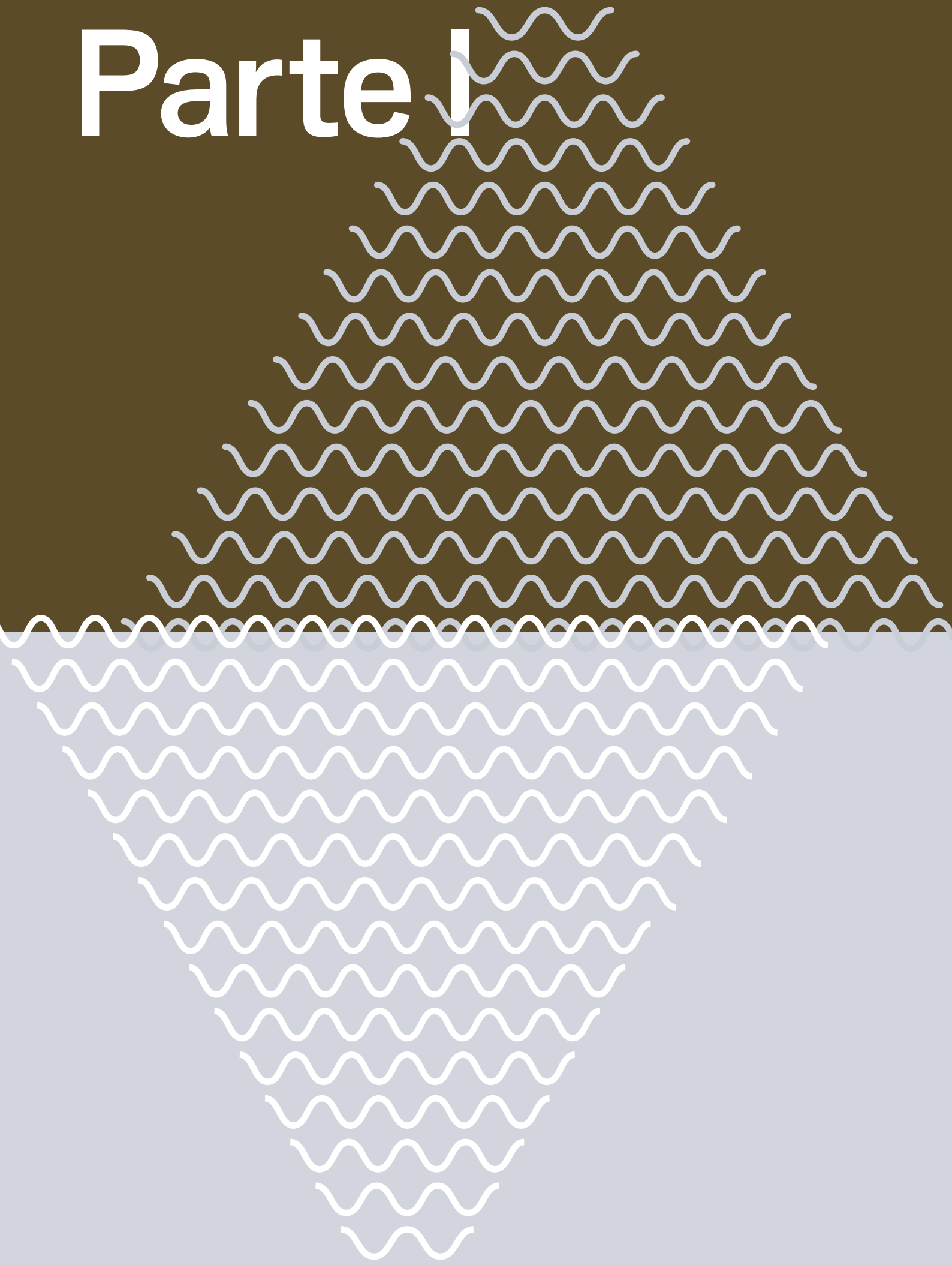
o assenza del riferimento permette di interrogare ciò che permane nella forma linguistica indipendentemente dal suo supporto e dalla sua interpretazione. Niccolò Monti, con *Words and Vectors: The Distributed View of Language*, ricostruisce invece il percorso che dalla linguistica distribuzionale conduce alle rappresentazioni vettoriali e ai moderni sistemi di *Natural Language Processing*. Guardare al linguaggio attraverso gli LLM significa, in questa prospettiva, interrogare anche la materialità dei dati e dei dataset, i processi attraverso cui il linguaggio viene reso computabile e l'opacità dei sistemi che ne risultano. In *LLMs and Normativity: A View from Languaging*, Filippo Batisti riporta la questione sul terreno della normatività. La sostituzione della nozione astratta di *language* con quella processuale di *languaging* consente di mettere al centro cooperazione, interazione e contesto. Il problema non è più soltanto se un modello artificiale sappia produrre una frase corretta, ma quale tipo di pratica renda possibile attribuire a quella produzione un significato e una validità condivisi. Chiude il percorso Tommaso Guariento con *The Dawns of Meaning. Monism, Dualism and Memory Machines as Different Models of Language*. Attraverso il confronto tra differenti modelli teorici dell'origine del significato, il contributo ricolloca gli LLM all'interno di una storia più ampia delle tecnologie, delle forme di memoria e delle concezioni del linguaggio. Definire questi sistemi *memory machines* piuttosto che semplicemente *language machines* significa anche resistere alla tentazione di naturalizzare il modello computazionale, ricordando la rete di processi biologici, storici, culturali e materiali dalla quale esso emerge.

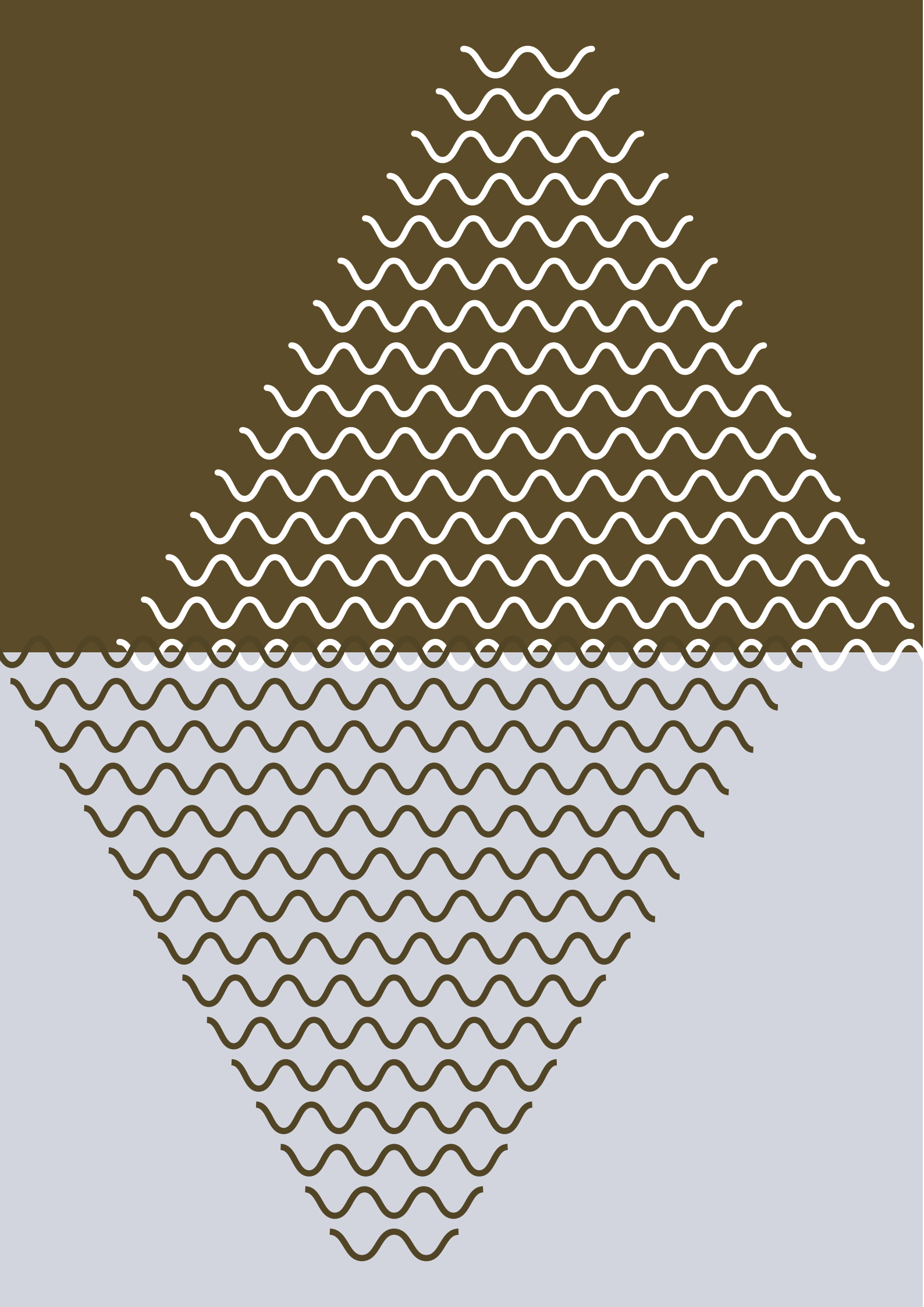
Il linguaggio, in ultima istanza, appare come un fenomeno costitutivamente stratificato: biologico senza essere soltanto biologico, cognitivo ma inseparabile dal corpo, sociale senza essere riducibile a convenzione, simbolico e al contempo operativo, tecnologico e, proprio per questo, capace di trasformare gli esseri umani che lo utilizzano. Questo lo rendono una perfetta cartina di tornasole per misurare gli effetti che le trasformazioni che le tecnologie della cosiddetta intelligenza artificiale operano sull'esperienza umana nella sua complessità. Cercare di avere una prospettiva filosofica su un simile quadro teorico significa confrontarsi con diverse discipline, rispettandone i limiti e i risultati, per costruire una concettualità relativa al linguaggio che esca dalle strette maglie degli approcci affidati alle singole discipline.





Parte I





L’evoluzione del linguaggio: un’indagine paleoneurologica

Fabio Macciardi

Professore emerito presso l’Università della California, Irvine (UCI) e responsabile del laboratorio di Antropologia Cognitiva Evoluzionistica del MIFiP (Museo e Istituto Fiorentino di Preistoria). La sua ricerca si concentra sull’analisi dell’evoluzione dei geni associati a tratti neurocognitivi complessi. In particolare, ha sviluppato metodologie di indagine che integrano il sequenziamento del DNA antico (paleogenetica) con la ricostruzione virtuale del cervello di paleo-ominini tramite tecniche di paleoneurologia.

fmacciar@uci.edu

Viola Castellani

Archeologa affiliata al Museo Fiorentino di Preistoria. È responsabile di progetti di archeologia preventiva nel territorio toscano, con la supervisione di scavi correlati a progetti infrastrutturali. I suoi contributi di ricerca si concentrano sull’archeologia cognitiva e sull’arte preistorica, indagando in particolare l’evoluzione cognitiva e le origini neurali del pensiero numerico negli esseri umani preistorici.

viola.castellani@unifi.it

Fabio Martini

Professore ordinario di Paletnologia presso l’Università di Firenze e attualmente direttore del Museo e Istituto Fiorentino di Preistoria. Conduce scavi in Italia, soprattutto nel centro-sud, e all’estero (Eritrea, Sudan). Settori privilegiati della sua ricerca sono le culture del Paleolitico e del Mesolitico, le produzioni litiche, l’arte dei popoli cacciatori-raccoglitori.

fabio.martini@unifi.it

This article examines the origin and evolution of human language through a paleoneurological and paleogenetic perspective, integrating evidence from cognitive archaeology, comparative neuroanatomy, and evolutionary genetics. Building upon the primary theoretical models of language, it addresses the limitations of accounts based on a sudden emergence or single genetic mechanisms, with a particular focus on the concepts of Merge and recursivity. The analysis considers the evolution of brain regions involved in language, the continuities and divergences among *Homo sapiens*, Neanderthals, and other hominins, as well as the role of genes, regulatory regions, and epigenetic elements in the progressive complexification of neural networks. The available evidence suggests that language cannot be reduced to a single evolutionary event; rather, it should be understood as the outcome of a complex, multicausal process shaped by the dynamic interplay among genetic transformations, brain reorganization, cognitive capacities, material culture, and social practices.

— EVOLUZIONE DEL LINGUAGGIO
— EVOLUZIONE DEL CERVELLO UMANO
— EVOLUTIONARY COGNITIVE ARCHEOLOGY
— NEUROSCIENZE COGNITIVE
— PALEOGENETICA
— PALEONEUROLOGI

Nel 1866, la Società Linguistica di Parigi proibì la presentazione di studi sull'origine del linguaggio, considerandoli speculazioni prive di basi empiriche solide, un divieto reiterato anche dalla Philological Society of London attorno al 1870. Per quanto oggi possa sembrare paradossale, questi divieti, emanati da quelle che all'epoca erano tra le principali società linguistiche, di fatto bloccarono la ricerca sull'evoluzione del linguaggio fino alla seconda metà del XX secolo. A più di un secolo di distanza dall'infelice decisione della Società Linguistica di Parigi, la ripresa delle indagini sull'evoluzione del linguaggio ha permesso di sviluppare metodologie di studio e di ottenere risultati che possono rivelarne i meccanismi biologici e culturali. Oggi, infatti, l'indagine sull'origine del linguaggio, considerato un tratto evolutivo caratteristico del genere *Homo* e, in particolare, di *Homo sapiens*, è estesa fino a includere molteplici discipline, dall'antropologia alle neuroscienze, dall'archeologia alla paleogenetica e alla riflessione filosofica (*sensu lato*). Si configura allora un campo di ricerca che integra neurobiologia ed evoluzione culturale in una prospettiva darwiniana aggiornata che ha dato origine a ciò che oggi è definito Evolutionary Cognitive Archaeology [ECA] o Evolutionary Neuropsychology (Coolidge, 2020). Con lo sviluppo dell'ECA, la paleoneurologia diviene in questi anni una metodologia di indagine sempre più raffinata. Le strategie e le tecniche paleoneurologiche spaziano dall'analisi morfologica dei cervelli antichi, ricostruiti in modo virtuale grazie allo sviluppo dei metodi di endocast, all'analisi genetica comparativa e all'analisi del DNA antico estratto dai resti ossei di reperti paleoantropologici (Paleogenetica) e delle proteine antiche (Paleoproteomica). Oltre ai dati biologici, le strategie di studio della Evolutionary Cognitive Archaeology includono i metodi di indagine della psicologia evoluzionistica (Tooby & Cosmides 2015) in costante sviluppo con la proposta di nuove (e vecchie) teorie interpretative che adattano schemi analitici derivati dall'indagine dei comportamenti umani del mondo presente agli ominini del tempo profondo. Tra i temi della psicologia evoluzionistica, sono rilevanti quelli derivati dalle neuroscienze cognitive, con i lavori pionieristici di T. W. Deacon applicati all'evoluzione (Deacon 1997), fino a un tema oggi considerato di attualità: la Material Engagement Theory, anche conosciuta con l'acronimo MET (Malafouris 2013). La MET ha trovato un'applicazione utile in una vasta gamma di discipline, tra cui l'archeologia, l'antropologia, la semiotica, le scienze cognitive e la filosofia della mente. Secondo Malafouris, la MET ruota attorno a tre ipotesi centrali: "l'ipotesi della mente estesa, che esplora l'intreccio costitutivo della cognizione con la cultura materiale; l'ipotesi della significazione enattiva, che esplora la natura del segno materiale in un approccio cognitivo che lega apprendimento e conoscenza all'azione e all'esperienza corporea [...]; [e] l'ipotesi dell'agentività materiale (nella formulazione originale in inglese: *material agency*), che esplora l'agentività non come una proprietà umana ma come il prodotto emergente dell'attività situata". La MET, quindi, si differenzia da altre teorie cognitive di impostazione materialistica attraverso la sua duplice attenzione alla cognizione incarnata (*embodied cognition*) e alla cultura materiale, e più specificamente: attraverso la sua esplorazione di come il divario tra questi campi epistemologici – che storicamente sono stati trattati in modo isolato – possa essere colmato. Molto ci sarebbe da dire sui modi con i quali le varie teorie interpretative psicologiche sono state applicate ai dati della paleoantropologia,

ma è impossibile non rilevare almeno come troppo spesso si sia cercato di interpretare i comportamenti di menti di specie diverse o comunque di individui vissuti in tempi molto antichi, come nel caso dei primi *H. sapiens*, adattando a questi gli schemi culturali dell'uomo contemporaneo (Wynn 2002; Wynn & Coolidge 2016; Wynn *et al.* 2020).

Insieme con l'emergere di nuovi dati relativi ai meccanismi – neurali, molecolari e culturali – che concorrono a illuminare l'origine e l'evoluzione del linguaggio, restano tuttavia evidenti vari problemi irrisolti che rendono difficile non solo la sintesi di quanto finora compreso, ma rendono anche complessa l'identificazione di future linee di ricerca. Un'esposizione dettagliata di tutti i vari problemi richiederebbe uno spazio maggiore di quello consentito a un articolo; pertanto, analizzeremo in modo più approfondito solo alcuni dei punti ritenuti più rilevanti in relazione all'origine paleoneurologica del linguaggio, rimandando il lettore incuriosito alla bibliografia che accompagna le sezioni successive.

Dove stanno andando le neuroscienze evoluzionistiche del linguaggio?

Un lavoro recente pubblicato su *Science* (Arnon *et al.* 2025) propone tre domini di indagine sull'evoluzione del linguaggio secondo una prospettiva definita "bioculturale", lasciando tuttavia nell'ambiguità la spiegazione di che cosa gli autori intendano per "bioculturale". Il primo dominio suggerito per uno sviluppo di ricerca sull'origine del linguaggio si riferisce all'apprendimento della produzione vocale (VPL: *Vocal Production Learning*), inteso come la capacità di un organismo di espandere e modificare il proprio repertorio di vocalizzazioni in base alle proprie esperienze uditive. La seconda linea di ricerca suggerita è quella della struttura linguistica (LS: *Linguistic Structure*), cioè come l'emergere di strutture "linguistiche", a partire dal linguaggio dei segni, implichi una combinazione di condizioni biologiche, cognitive e culturali. Infine, la terza linea di indagine si focalizza sull'interazione sociale come determinante essenziale per l'apprendimento del linguaggio nell'uomo e per l'apprendimento di comportamenti acquisiti in altre specie. Oltre alla prevedibile inutilità di proporre linee guida di indagine per altri ricercatori, dato che non si capisce perché questi ultimi dovrebbero seguire queste (o altre) indicazioni sulla base di non si sa quale autorità, l'ambiguità della prospettiva bioculturale, non meglio definita se non come un evento complesso multideterminato, dimostra come l'identificazione delle tre linee di ricerca sottolinei piuttosto la preoccupazione di aderire ad uno schema evolutivo darwiniano classico e tradizionalista – evoluzione per selezione naturale ad adattamenti lenti e progressivi – che era in buona parte già superato dagli sviluppi della Sintesi Moderna (*Modern Synthesis*) a metà del XX secolo. Grazie al lavoro teorico e concettuale di T. Dobzhansky, S. Wright, S.E. Fisher, E. Mayr e altri (Smocovitis 1992), l'unificazione di concetti apparentemente inconciliabili – la processulità dell'evoluzione per selezione naturale insieme con la strutturalità, almeno apparente, della genetica mendeliana – ha permesso la formazione di discipline come la Paleoneurologia e la Paleogenetica che sono oggi in rapido sviluppo (Bruner 2015; Cavalli Sforza 1994; Higuchi *et al.* 1984; Paabo *et al.* 2004; Pääbo 1985; Pauling L. 1963). Occorre solo qui ricordare che sia l'originale proposta darwiniana di evoluzione per selezione

naturale che l'unificazione di questa con la teoria genetica mendeliana operata dalla sintesi moderna, non sono costruzioni epistemologiche statiche, ma dinamiche e richiedono un costante aggiornamento ai metodi e alle scoperte delle altre discipline che fanno dell'evoluzione il loro campo di indagine. Non si può peraltro dimenticare che Darwin stesso aveva considerato la possibilità che l'evoluzione potesse procedere non solo per "adattamenti" progressivi, ma anche per modificazioni di tratti "pre-adattati", indirizzate poi verso percorsi evolutivi alternativi, sotto la spinta della selezione naturale e non, come troppo spesso interpretato, secondo finalismi teleologici. Darwin stesso, nel suo libro *The Descent of Man and Selection in Relation to Sex* del 1871, scriveva: "Non posso dubitare che il linguaggio debba la sua origine all'imitazione e alla modifica, aiutate da segni e gesti, di vari suoni naturali, le voci di altri animali e le grida istintive dell'uomo."

Per comprendere come la paleoneurologia e la paleogenetica possano contribuire allo studio dell'evoluzione del linguaggio, è opportuno allora rivisitare almeno alcuni dei problemi irrisolti del linguaggio in una prospettiva che integri (paleo)antropologia, (paleo)genetica e (paleo)neuroanatomia funzionale.

Il linguaggio umano: una possibile definizione e i suoi (molti) limiti

Un primo problema fondamentale è come definire il linguaggio, non tanto come struttura linguistica, quanto piuttosto secondo un approccio che ne consideri l'evoluzione, ossia come un processo in divenire. Senza una prospettiva che consideri il tempo e non solo i modi dell'evoluzione, qualunque definizione del linguaggio che non tenga conto del tempo del suo sviluppo evolutivo rimane soltanto una riflessione tecnica interna al sistema linguistico.

Secondo molti ricercatori, il linguaggio non è altro che una forma di comunicazione avanzata, con origini nelle modalità comunicative già presenti nel mondo animale, queste ultime a loro volta intese come trasmissione di informazioni. E quindi, come forma di comunicazione, il linguaggio segue un tipico sviluppo adattativo darwiniano, come altri tratti fisici o psicologici di *H. sapiens*, e può essere studiato con i metodi tradizionali dell'antropologia evoluzionistica. Secondo altri, il linguaggio umano nasce invece come una modalità formale di organizzazione del pensiero, strutturato in forme complesse a struttura gerarchica come la ricorsività, e la sua funzione comunicativa è soltanto derivata. In questa prospettiva, il linguaggio umano è un tratto innovativo che non trova un immediato riscontro negli esempi del mondo animale, perché la sua origine è legata a fenomeni di exaptation. Diviene allora essenziale definire che cosa sia il linguaggio: solo in seguito, a seconda della definizione del tratto cognitivo linguistico, è possibile analizzarne i probabili meccanismi evolutivi.

La teoria del linguaggio ancor oggi più ampiamente accettata, e quella su cui si fondano le considerazioni che seguono, si basa sul lavoro fondamentale di Chomsky, che ha introdotto e perfezionato nel tempo un modello linguistico che prevede l'interazione tra un sistema concettuale-intenzionale (C-I) e un sistema sensomotorio (S-M) (Hauser *et al.* 2002). L'interazione tra una componente percettivo-motoria e una componente

idetica costituisce un modello cognitivo generalista che si può applicare a molti tratti cognitivi condivisi *da Homo* con il mondo animale, dagli uccelli ai primati. Per quanto riguarda il linguaggio, il modello chomskiano si estende da questo schema generalista ad uno più complesso, articolato in due componenti. La prima componente, che comprende l'interazione tra il sistema C-I e quello S-M, mostra come un'infinita varietà di espressioni possa essere generata da un insieme limitato di elementi e viene indicata come Facoltà del linguaggio in senso esteso (FLB: Faculty of Language in the Broad sense). Questa capacità di generare un numero potenzialmente infinito di espressioni a partire dall'insieme di due elementi cui si può poi aggiungere un terzo elemento complementare, e così via ("discrete infinity"), è possibile grazie a una caratteristica unicamente umana (seconda componente): la capacità di strutturare le espressioni ricorsivamente in modo gerarchico. Questa gerarchizzazione sintattica, dimostrata da frasi come "Giovanni ha detto che ha incontrato Giorgio che gli ha detto che Maria si incontra con Alberto", è nota come Facoltà del Linguaggio in senso stretto (FLN: Faculty of Language in the Narrow sense). La chiave di questo programma ricorsivo, definito nel tempo come programma minimalistico o Grammatica Universale (UG: Universal Grammar), è fondata sull'ipotesi che la struttura gerarchica della sintassi del linguaggio umano consista in una singola operazione, *Merge* (Hauser *et al.* 2002). Merge consiste nell'unione di due elementi semplici (merge, primo livello), come, per esempio, due parole o una parola e un verbo, a cui poi si può aggiungere un terzo elemento (merge, secondo livello), e così via, potenzialmente all'infinito. Il linguaggio può essere definito come un sistema simbolico combinatorio, strutturalmente organizzato in fonologia, morfologia, e sintassi, grazie alla funzione merge-ricorsività, indipendentemente da come la funzione *merge* si espliciti in schemi alternativi grammatico-sintattici storicamente indicativi della formazione di una "lingua". Alcune lingue, come quelle romanze, germaniche, alcune lingue asiatiche e molte d'origine africana, prevedono uno schema sintattico Soggetto-Verbo-Oggetto (SVO), mentre altre lingue storicamente determinate seguono un altro ordine, per esempio Soggetto-Oggetto-Verbo (SOV), come molte lingue asiatiche e orientali. In ogni caso, e qualunque sia la formulazione sintattica, il concetto e il senso di *merge* non cambiano e una funzione merge con ricorsività permette comunque la formazione di linguaggio secondo quanto predetto da FLN. Sebbene Chomsky e altri suggeriscano che la FLN si sia evoluta attraverso la selezione naturale e riconoscano che il linguaggio abbia una solida base genetica (Berwick & Chomsky, 2015), testare sperimentalmente l'ipotesi "Merge" e la conseguente ricorsività come tratto genetico a base evolutiva resta una sfida significativa. La proposta originale di Chomsky per giustificare l'origine genetica (ergo, innata) della capacità del linguaggio umano era infatti fondata sull'ipotesi ingenua che il linguaggio fosse apparso in *H. sapiens* in seguito ad una singola mutazione a un singolo gene X e che tale mutazione si fosse poi rapidamente estesa a tutti gli individui dell'epoca, più o meno circa 50,000 anni fa (50 Kya). Inoltre, la comparsa rapida e improvvisa del linguaggio avrebbe generato anche la produzione di cultura e di un mondo simbolico, come prodotti del linguaggio-pensiero.

Pur riconoscendo la novità di un'impostazione genetico-evoluzionistica sull'origine del linguaggio, questa assumerebbe caratteri implausibili:

un evento improvviso "tutto o niente" che si sarebbe esteso a tutta l'umanità dell'epoca in modo immediato, l'impossibilità logica di derivare una complessità emergente da una causa priva di fondamento biologico e, non ultimo, l'assunto implicito di un solo gene responsabile di un sistema cognitivo fondato su dimensioni estese e complesse come la FLB e la FLN.

Una diversa interpretazione del patrimonio archeologico, derivata da una rianalisi dei reperti disponibili, suggerisce ipotesi alternative. Per prima cosa, l'ipotesi della singola mutazione in un solo gene è oggi smentita da analisi genetico-formali (Benítez-Burraco & Boeckx 2014; Martins & Boeckx 2019) e, come vedremo in seguito, dall'identificazione di diversi geni correlati allo sviluppo e alla formazione del linguaggio. Inoltre, la ricorsività non è più considerata appannaggio esclusivo del linguaggio (Di Vincenzo & Manzi 2013). Le modalità di incisione dei più antichi reperti di *H. sapiens*, per esempio le incisioni presenti sul blocco di ocra ritrovato a Blombos (Sud Africa) e datato attorno a 80 Kya e quelle sui gusci di uova di struzzo datate alla stessa epoca, suggeriscono che il processo di incisione segua uno schema che mostra gli stessi meccanismi ricorsivi del linguaggio, indipendentemente da un ipotetico e del tutto speculativo significato linguistico né tantomeno simbolico delle incisioni analizzate (Decembrini *et al.* 2026; Longa 2013). Allo stesso modo, alcune incisioni sicuramente attribuibili a Neanderthal – ma certamente non tutte – sono oggi considerate un esempio di ricorsività numerica (gruppi di linee equidistanti in moduli ripetitivi dello stesso numero di incisioni) e interpretate cognitivamente come una generazione di segni possibile grazie a un procedimento di "merge" ricorsivo (Macciardi *et al.* 2025). La procedura, poi, della produzione di utensili litici a partire dal periodo Acheuleano fino al Neolitico e alle età dei metalli sembra seguire del pari una procedura ricorsiva, anche in questo caso simile a quella del linguaggio. Secondo questa lettura della produzione tecno-litica, la scheggiatura di un ciottolo di pietra grezza produce (a0) una prima scheggia litica e (b0) un nucleo di pietra (*core*) modificato, sul quale il lavoro di progressiva scheggiatura ricorsiva produce una seconda scheggia (a1) e un core ulteriormente modificato (b1), e così via, fino ad ottenere la forma desiderata dell'utensile di pietra (bX) insieme con schegge che possono a loro volta essere utilizzate (aX). Se queste analisi cognitive del patrimonio archeologico fossero poi confermate da altri studi, ne conseguirebbe che l'origine di merge e ricorsività potrebbe essere studiata ed analizzata come un tratto cognitivo non soltanto attribuibile al linguaggio. Una prima conseguenza in questo caso, è che l'origine di merge/ricorsività segue un'evoluzione (adattativa o ex-attiva) di origine decisamente più antica di quella proposta da Chomsky, retrodatando la sua origine ad 250-400 Kya se non più, con le prime produzioni tecno-litiche di scheggiatura per ritocchi minimi (Harris & Miniussi 2003; Li G. *et al.* 2022; Macciardi *et al.* 2025; Moore 2020; 2024). Una seconda conseguenza, forse ancora più rilevante, è che, se il procedimento ricorsivo è presente almeno da 400 Kya, è ipotizzabile che il linguaggio – o almeno i suoi presupposti cognitivo-formali – non sia patrimonio esclusivo di *H. sapiens*, ma possa essere condiviso anche con altre specie ominine (Dediu & Levinson 2018). Infine, il progressivo sviluppo evolutivo di merge e di ricorsività supera l'ipotesi genetica di un solo gene e di una sola mutazione causale per spiegare l'origine di un tratto di tale complessità, a favore di un'origine multigenica più verosimile, che si struttura nel tempo secondo meccanismi

potenzialmente verificabili sperimentalmente (Macciardi & Martini 2022; Martins & Boeckx 2019).

Un'ulteriore complessità deriva dall'ipotesi che il linguaggio, definito dalla regola ricorsiva di FLN, possa essere inteso principalmente come un processo di organizzazione del pensiero, mentre il suo ruolo di strumento di comunicazione resta secondario e derivato. L'ipotesi che il linguaggio sia uno strumento di regolazione del pensiero, inteso come "linguaggio interno", risale a Vigotskij (2008) ed è stata poi ripresa ed elaborata, anche se spesso implicitamente, nei lavori di Chomsky. Recenti studi di neuroimaging, tuttavia, mettono in discussione questa ipotesi, dimostrando che le reti neurali che supportano il linguaggio e il pensiero sembrano essere distinte, in condizioni sperimentali controllate, separando così i processi neurali alla base del linguaggio da quelli correlati al pensiero (Fedorenko *et al.* 2024; Friederici & Becker, 2025). Da un punto di vista *strutturale*, i sistemi percettivi e motori del linguaggio sono definiti come sottosistemi separati, situati rispettivamente nella corteccia uditiva e nelle aree di percezione del linguaggio nella circonvoluzione temporale superiore (CTS), per la percezione, mentre la produzione del linguaggio è preminente nella corteccia motoria e nelle aree di pianificazione motoria, tra cui l'area di Broca nel lobo frontale inferiore. È importante notare che queste aree cerebrali, funzionalmente definite come percettive e motorie, sono strutturalmente connesse tramite tratti di fibre di sostanza bianca ("fascicoli"), localizzati dorsalmente e ventralmente nella corteccia cerebrale, che garantiscono il flusso di informazioni tra le aree. Negli esseri umani, però, sono state identificate due vie dorsali distinte che forniscono connessioni strutturali a due reti indipendenti. Analisi comparative interspecie, cioè secondo una logica evoluzionistica *processuale*, hanno rivelato che queste due reti strutturali dorsali presentano traiettorie evolutive diverse, sia nelle regioni bersaglio dei tratti di fibre sia nella loro lateralità. Nel corso dell'evoluzione, la regione bersaglio, definita come Brodmann Area 44 (BA44), situata nella circonvoluzione frontale inferiore sinistra (area di Broca, *pars opercularis*) e coinvolta nella produzione fonologica e motoria del linguaggio secondo la neurologia classica, si è espansa negli esseri umani (rispetto agli scimpanzé) e ha in parte cambiato funzione, passando da regione motoria e comunicativa nei primati non umani a regione correlata al linguaggio negli umani. L'area di Broca nell'uomo controlla infatti anche altre funzioni motorie a componente ritmica, come la danza. Pur non spiegando quasi nulla sull'origine di merge, la danza è un mezzo espressivo dotato di ricorsività – coreografia, strutture del movimento – e capace di comunicare emozioni e sentimenti in modo non verbale, con un'intensità e una profondità della comunicazione che non sono inferiori a quelle di altre espressioni gestuali, per esempio la lingua dei segni, considerate equivalenti alla produzione fonologica del linguaggio.

Morfologia, funzione ed evoluzione di regioni cerebrali coinvolte nel linguaggio

Un'analisi comparativa dell'evoluzione dell'area di Broca (aree 44 e 45 di Brodmann, secondo la nomenclatura corrente della ripartizione citologica della corteccia cerebrale umana) e delle sue connessioni con le regioni cerebrali deputate alla percezione del linguaggio – area di Wernicke, situata

nel lobo temporale superiore sinistro, e connessa con l'area di Broca tramite il Fascicolo Arcuato – ci introduce a una dimensione applicativa della paleoneurologia, come ambito preferenziale per lo studio dell'evoluzione del cervello nella linea ominina. (Hecht *et al.* 2025; Rilling *et al.* 2008; Sierpowska *et al.* 2022).

Se il dato più evidente dell'evoluzione del cervello negli ominini da *H. habilis* a *H. sapiens* dipende da un incremento volumetrico delle dimensioni cerebrali, risultato di un lungo processo scientifico basato sui dati dell'antropologia fisica e della paleontologia tra il XIX e il XX secolo e consolidato nella formalizzazione del concetto di quoziente di encefalizzazione (EQ), gli sviluppi recenti della paleoneurologia mostrano che il processo di evoluzione cerebrale è dovuto principalmente a una progressiva complessizzazione della struttura neuronale e dell'architettura cerebrale (Kaas & Herculano-Houzel 2026). L'incremento volumetrico è pertanto una conseguenza macromorfologica di processi di riorganizzazione molecolare e strutturale del tessuto neurale.

In particolare, per l'area di Broca, recenti studi funzionali e di architettura recettoriale supportano la conclusione che l'area 44 di Brodmann (BA44) si sia evoluta a partire da una regione correlata all'azione, fino a diventare un sistema bipartito, con una porzione posteriore a supporto dell'azione e una porzione anteriore a supporto dei processi sintattici (Gallardo *et al.* 2023). Questa doppia funzione dell'area di Broca, la pianificazione motoria complessa, inclusa l'elaborazione dei gesti, e la produzione e la comprensione di strutture grammaticali, rende anche ragione dell'equivalenza tra il linguaggio gestuale e quello in forma fonologica. Non è pertanto necessario invocare meccanismi specifici e diversi per le due modalità linguistiche, gestuali o fonologiche. Sembra inoltre quasi ovvio che il linguaggio dei segni abbia sviluppato e segua regole sintattiche simili, se non identiche, a quelle del linguaggio parlato, incluse le differenze grammaticali delle lingue storicamente determinate: un individuo che utilizzi il linguaggio dei segni si esprimerà con modalità sintattiche differenti a seconda della cultura cui appartiene, per esempio europea, asiatica o amerindia.

Dalle differenze volumetriche a una più complessa architettura neuronale, le variazioni evolutive della struttura cerebrale sono numerose. L'analisi di tali varianti è un procedimento a sua volta complesso, per cui in questa sede proviamo a sintetizzarle secondo alcuni principi guida. Per prima cosa, l'evoluzione del cervello umano attuale non è ascrivibile a un singolo evento, sia esso la progressiva encefalizzazione – tale per cui la dimensione del cervello umano è cinque volte superiore al valore che ci si aspetterebbe in un confronto comparativo con primati e mammiferi – o la più complessa architettura del tessuto neurale, che comprende sia le cellule (neuroni) che le loro connessioni. Molti paleoneurologi oggi tendono invece a privilegiare la spiegazione dell'unicorno anatomico a giustificazione dell'unicità del cervello di *H. sapiens*. Un esempio è l'interesse attuale per una particolare regione del lobo parietale, chiamata "precuneo". Situato nella faccia mediale del lobo parietale, da alcuni paleontropologi il precuneo viene considerato il maggior responsabile dell'attuale conformazione globulare del cervello di *H. sapiens*, attribuendogli inoltre funzioni di integrazione di stimoli sensoriali e funzioni mnemoniche peculiari, che sono però perlopiù speculative e rappresentate in via indiretta (Bruner &

Iriki 2016; Bruner *et al.* 2017b; Bruner *et al.* 2017a; Li, Y. *et al.* 2022; Lyu *et al.* 2021; Takashima *et al.* 2020). Integrando invece dati di paleoneurologia con analisi della conformazione cerebrale secondo un indice di globularità (dal cervello meno globulare a quello più globulare) nelle popolazioni attuali, e correlando le analisi morfologiche con dati genetici, è possibile individuare i meccanismi di un processo di globularizzazione che coinvolge diverse strutture cerebrali (il lobo frontale, i nuclei sottocorticali, il cervelletto e i fasci anatomici che corrono attraverso queste regioni) e un pattern complesso di reti geniche e neurali (Gunz *et al.* 2019; Neubauer *et al.* 2018). I risultati di queste indagini indicano una divergenza evolutiva inconciliabile tra Neanderthal (caratterizzati da un cervello "allungato" in senso antero-posteriore) e *H. sapiens*, nonostante una probabile origine da un antenato comune, oggi identificato in *H. heidelbergensis* o in *H. antecessor*. La forma dei cervelli di Neanderthal e *H. sapiens* dipende infatti da un elevato numero di varianti (morfologiche, anatomiche e genetiche) che probabilmente risalgono a circa 450-600 Kya, prima che la speciazione divergente tra le due specie diventasse evidente nel patrimonio paleoantropologico (Boeckx & Benitez-Burraco 2014; Gomez-Robles 2019; Gomez-Robles *et al.* 2025; Melchionna *et al.* 2020; Sehlebusch *et al.* 2017; Sherwood & Bradley 2019). In secondo luogo, non sono solo le strutture morfo-anatomiche e cellulari a essere variate secondo uno schema a mosaico, ma un ruolo essenziale è svolto anche dal tempo impiegato dai processi evolutivi (Tilot *et al.* 2021). Le espressioni morfologiche del processo di globularizzazione, evidenti dai reperti fossili, e considerato l'evento di speciazione di *H. sapiens*, risalgono a circa 250 Kya in Africa (McDougall *et al.* 2005; Vidal *et al.* 2022), quindi molto dopo l'inizio del processo di divergenza genetico-molecolare. In seguito, questa variazione morfologica accelera in modo esponenziale nel corso degli ultimi 100.000 anni (DeSilva *et al.* 2019; DeSilva *et al.* 2021; Gunz *et al.* 2012; Neubauer *et al.* 2018).

Altri eventi paleoneurologici considerati essenziali per l'origine del linguaggio riguardano una diversa morfologia della circonvoluzione angolare e della circonvoluzione marginale del lobo parietale inferiore, una regione cerebrale che avevamo già incontrato nel caso del Fascicolo Arcuato, la struttura anatomica che collega l'area di Broca con l'area di Wernicke. Dai rilievi endocranici di Neanderthal e *H. sapiens* "arcaico" (prima di 80-100 Kya) effettuati tramite analisi dei rispettivi endocast in vari reperti archeologici, si può apprezzare, un'espansione del lobo frontale e un'espansione laterale del lobo parietale, a indicare che entrambe le circonvoluzioni (angolare e marginale) – come pure il Fascicolo Arcuato per assunto indiretto – sono evidenti in entrambe le specie. Se a questi rilievi cerebrali, pur parziali, si aggiunge che anche una formazione anatomica periferica, l'osso ioide, che si colloca alla base della lingua tra mandibola e laringe e garantisce una mobilità fondamentale per deglutire, respirare e parlare, è già presente anche in Neanderthal (nei rari casi in cui è stato possibile reperirlo (Abady *et al.* 2020; Haeusler *et al.* 2019), è evidente che in quest'ultimo sono già presenti le strutture anatomiche e funzionali che permettono un linguaggio simile a quello di *H. sapiens*, pur se con probabili tratti di arcaicità o più che verosimili differenze legate ad un diverso sviluppo del cervello/mente tra *H. neanderthalensis* e *sapiens*. Non sappiamo e non potremo mai sapere se Neanderthal avesse già sviluppato il tratto distintivo del linguaggio moderno (merge con ricorsività) o

se invece avesse un'altra forma di cognitività linguistica, ma oggi appare indubbio che in ogni caso, anche *H. neanderthalensis* possedesse "un" linguaggio. Se poi, come hanno dimostrato T. Fitch e i suoi collaboratori, anche le scimmie possiedono una conformazione del tratto vocale che permetterebbe loro di parlare, ma non lo fanno (Fitch *et al.* 2016), allora l'origine e la possibilità del linguaggio devono essere ulteriormente ricercate in altri meccanismi neurobiologici che possano spiegare la complessità dell'organizzazione molecolare dei neuroni e della loro architettura cerebrale, cioè in meccanismi di origine genetica.

Meccanismi genetici all'origine delle strutture molecolari e cellulari che rendono possibile il linguaggio

Abbiamo visto nei paragrafi precedenti che un processo di speciazione a livello genetico-molecolare precede di molto, anche di centinaia di migliaia di anni, la presenza di reperti paleoantropologici che segnalano la differenza morfologica di specie: la divergenza molecolare tra Neanderthal e H. sapiens può essere datata a 450//550 Kya, mentre i primi esempi di H. sapiens datati con metodi paleoantropologici risalgono a circa 250 Kya (Di Vincenzo & Manzi 2023; Fu *et al.* 2016; Harris *et al.* 2023; Mounier & Mirazon Lahr 2019). Il meccanismo di evoluzione a livello molecolare, secondo una lettura della selezione naturale per piccoli e lenti adattamenti, tipicamente prevede che una singola mutazione favorevole, un evento stocastico, venga poi selezionata nella popolazione, aumentandone di fatto le probabilità di riproduzione. Esempi di questa modalità evolutiva si trovano, ad esempio, nell'espressione del carattere "capacità di digestione del latte in età adulta" (tecnicamente definito come "persistenza della lattasi") o nell'adattamento di alcune popolazioni a uno stile di vita ad alte altitudini, come nelle popolazioni tibetane o in quelle andine. Nel caso della persistenza della lattasi, è coinvolto un solo gene, LCT, ma le varianti che rendono tolleranti al lattosio in età adulta sono differenti a seconda delle macropopolazioni attualmente esistenti (europea, nordafricana, est africana, dell'Asia centrale, arabe), mentre per l'adattamento ad alte altitudini, geni diversi sono stati selezionati in differenti popolazioni, tibetani, andini o etiopi. Oggi poi, possiamo anche datare l'origine delle mutazioni, da cui possiamo derivare che le mutazioni dei geni che permettono l'adattamento ad alte altitudini nelle popolazioni tibetane datano al Paleolitico Medio, e sono probabilmente avvenute tramite inbreeding dei primi *H. sapiens* che colonizzarono la regione con individui della specie Denisovan (Zhang *et al.* 2021), mentre le mutazioni al gene LCT datano al periodo di transizione Mesolitico-Neolitico (Marnetto *et al.* 2022). Ancora più articolato è il quadro evolutivo di altri caratteri, come la resistenza alla malaria: data la persistenza della patologia da tempi sicuramente lunghi, in questo caso sono coinvolti vari geni, con mutazioni che variano non solo a seconda delle popolazioni, ma anche nei tempi della loro apparizione, con varianti geniche apparse più volte nel quadro generale dell'evoluzione.

I modi e i tempi dell'evoluzione molecolare che sottendono le origini del cervello delle specie umane, a partire da circa un milione di anni fa, sono ancora più complessi. Come abbiamo già visto a proposito dell'ipotesi originaria di Chomsky relativa all'origine genetica di *Merge*, una (o, se per questo, anche più di una) mutazione di un singolo gene non è compatibile

con l'espressione di un fenotipo complesso. Per spiegare l'eziologia della componente genetica di un tratto complesso sono state avanzate diverse ipotesi, tra loro complementari più che alternative: tra le più rilevanti, la formazione di nuovi geni, l'ex-adattamento (exaptation) di elementi molecolari regolatori della funzione o dell'espressione di geni già esistenti, la duplicazione (o la delezione) di segmenti cromosomici di dimensione variabile. In funzione dell'evoluzione del linguaggio, possiamo ricordare, per esempio, la duplicazione segmentale, cioè all'interno dello stesso cromosoma, del gene SRGAP2, che nel corso dell'evoluzione della linea ominina è duplicato ben quattro volte, dando origine ai geni SRGAP2A, B, C e D circa un milione di anni fa. Le varianti più recenti SRGAPC e D determinano una diversa conformazione delle sinapsi neuronali, in grado di trasmettere informazioni più efficientemente (Charrier *et al.* 2012; Dennis *et al.* 2012). A questo si aggiungono altri geni, sempre coinvolti nella formazione, sviluppo e genesi del linguaggio, come il gene ARHGAP11B, presente nella sua conformazione solo in *H. sapiens*, che promuove l'espansione delle strutture corticali del cervello (Florio *et al.* 2015), o il gene NOVA1 che in *H. sapiens* rallenta la formazione del tessuto cerebrale mentre ne rende più complessa la struttura interna, a differenza della variante presente in Neanderthal (Trujillo *et al.* 2021). Paradigmatico è l'esempio del gene FOXP2, considerato all'inizio come "il" gene del linguaggio (MacDermot *et al.* 2005), ma in seguito derubricato a un'altra funzione, pur rilevante, cioè quella di un gene regolatore che a cascata controlla una moltitudine di altri geni. FOXP2 è correlato a una patologia, l'aprassia del linguaggio dell'adolescenza (CAS: Childhood Apraxia of Language), un disturbo motorio che si manifesta come difficoltà a pronunciare sillabe o suoni dovuta a un deficit di *pianificazione* dei movimenti necessari alla produzione fonologica, anche se i muscoli coinvolti non presentano un difetto strutturale; si tratta quindi di un difetto strutturale dei meccanismi motori cerebrali e non di un difetto muscolare periferico (Fisher 2019). Infine, secondo una prospettiva evolutiva, abbiamo solo due nucleotidi di differenza nella sequenza del gene tra lo scimpanzè e *H.sapiens* e nessuna differenza di sequenza tra *H.sapiens* e Neanderthal. Tuttavia, solo la nostra specie possiede un gene regolatore a RNA, posto all'interno della sequenza di FOXP2, che controlla l'espressione di quest'ultimo (il fattore di trascrizione POU3F2) e che non è condiviso con altri ominini, Neanderthal e Denisovan (Maricic *et al.* 2013).

Dagli studi genetici più recenti, il fenomeno più importante per l'evoluzione del cervello è legato non tanto ai geni, le cui sequenze nucleotidiche rappresentano solo il 2% dell'intero genoma umano, quanto del genoma regolatore non-codificante che ne rappresenta il restante 98%. Più della metà di quest'ultimo è rappresentato da trasposoni e retrotrasposoni, elementi di antica derivazione virale: se solo una piccola frazione dei retrotrasposoni ha mantenuto la capacità di muoversi (traspondere, in termini tecnici) all'interno del genoma duplicandosi da una posizione ad un'altra, tutti hanno acquisito per meccanismi ex-adattativi una funzione regolatrice sulla funzione e sull'espressione dei geni che codificano per proteine (Hoyt *et al.* 2021; McClintock 1956). Trasposoni e retrotrasposoni comprendono la maggior parte degli elementi epigenetici che regolano le dinamiche evolutive, di sviluppo e funzionali del genoma umano. Nel solo lobo frontale sinistro, sono stati identificati circa 650,000 retrotrasposoni che

regolano i circa 14,000 geni che sono espressi nei neuroni di quest'area; di questi, il 60% circa lo condividiamo con gli altri primati e sono elementi deputati al controllo di geni "antichi" del sistema nervoso centrale, per esempio geni che codificano per le sinapsi ei neurotrasmettitori; il 30/35% è invece condiviso solo con altri ominini (Neanderthal e Denisovan) e serve per le regolazione dei geni coinvolti nello sviluppo della memoria, dei sistemi cognitivi e per il controllo di molti dei geni correlati al linguaggio, mentre non condividiamo con nessuna altra specie circa il 10/15%. Questi ultimi controllano la funzione e l'espressione di geni attivi nel sistema nervoso centrale e deputati alla migrazione e alla complessizzazione delle cellule neurali (Guffanti *et al.* 2018; Guffanti *et al.* 2019): lo studio della funzione e dell'evoluzione di questi geni e dei loro regolatori trasposonici è la frontiera delle neuroscienze cognitive di oggi.

Gli elementi genetici a funzione regolatrice comprendono non solo trasposoni e retrotrasposoni, ma anche altre classi di oggetti molecolari: una gran parte di questi è ulteriormente caratterizzata da un fenomeno di "accelerazione evolutiva", tale per cui anche elementi che sono conservati nel genoma di altre specie, primati non umani o roditori, accelerano modificando la loro sequenza quando si trovano all'interno del genoma di *H. sapiens* e sono definiti come HARS (Human Accelerated Regions) (Franchini & Pollard 2015; Franchini 2021; Ryzhkova *et al.* 2022) o "Human-specific regulatory regions" (Marnetto *et al.*, 2014), quando non sono condivisi con nessuna altra specie. Una evoluzione accelerata delle sequenze codificanti e regolatrici caratterizza anche geni che sono correlati al controllo dei sistemi cognitivi umani, con una velocità che è pari o superiore a dieci volte il tasso di mutazione evolutiva del genoma umano nel caso di geni non a funzione cerebrale, una variabilità che riflette una selezione positiva a funzione adattativa o una selezione negativa che tende ad eliminare sequenze interferenti con i meccanismi di selezione naturale (Bustamante *et al.* 2005; Macciardi *et al.* 2020)

Appare evidente anche da questa sintesi estrema che l'evoluzione del cervello umano richieda l'azione congiunta di un grande numero di elementi genici, codificanti e regolatori: la maggior parte di questi è coinvolta in funzioni che controllano tratti cognitivi, il linguaggio, ma anche la memoria, l'apprendimento e altri, tramite la costruzione di reti neurali complesse, tuttora in rapida evoluzione e specifiche della nostra specie.

Con gli sviluppi della paleoneurologia e della paleogenetica, è stato introdotto un nuovo metodo che combina evidenze provenienti da archeologia, genetica e neuroscienze e che permette di valutare sperimentalmente se i reperti provenienti dagli scavi, i dati genetici – come l'evoluzione molecolare del gene FOXP2 – e i cambiamenti nella struttura cerebrale in tutta la linea evolutiva degli ominini possano rivelare la complessa interazione di fattori biologici e culturali necessaria per spiegare l'origine del linguaggio e del pensiero astratto precoce.

Questo approccio supera lo schema comportamentalista della "scatola nera" del cervello umano, in cui i meccanismi sono sconosciuti e solo gli effetti sono osservabili, e invece esamina e studia le dinamiche cerebrali interne e le strutture neurobiologiche nel contesto delle evidenze archeologiche. I nuovi metodi e le strategie analitiche sviluppatesi nel corso della ancor breve vita della paleoneurologia ci permettono così di rintracciare i primi segni di cognizione nelle specie di ominini che ci hanno preceduto,

tra cui *Homo erectus*, *Homo neanderthalensis* e i primi *Homo sapiens*. Analogamente ai paradigmi delle neuroscienze cognitive molecolari, il metodo della paleo-neuro-genetica si concentra sulla scoperta di elementi che contribuiscono alla formazione dei meccanismi neurali alla base della cognizione, considerato un fenomeno emergente e culturalmente influenzato. L'obiettivo di questo paradigma di ricerca non è affermare l'esistenza di "geni del linguaggio" o di regioni cerebrali responsabili di tratti cognitivi specifici, come l'ipotesi ormai superata di un'area cerebrale deputata al linguaggio, bensì sottolineare che i geni e le strutture cerebrali forniscono gli elementi fondamentali per lo sviluppo delle funzioni cognitive, che perlopiù non sono localizzate in regioni anatomiche, come la vecchia scuola frenologica proponeva in passato. Indagando la morfologia e la funzione del cervello attraverso la paleoneurologia, analizziamo anche le transizioni di sequenza molecolare e genomica tra le specie di ominidi tramite la paleogenetica con l'obiettivo finale di comprendere l'interazione in corso tra geni, cervello e cultura come meccanismo fondamentale dell'evoluzione cognitiva (Macciardi *et al.* 2025).

Cambiamenti fondamentali nella cognizione sono già evidenti nella documentazione archeologica durante la transizione dall'industria tecnolitica dell'Olduvaiano a quella dell'Acheuleano, avvenuta all'incirca tra 1,75 milioni e 800.000 anni fa (Macciardi *et al.* 2025). Studi sperimentali di scheggiatura indicano che la creazione di bifacciali acheuleani attiva regioni cerebrali simili a quelle coinvolte nel ragionamento simbolico moderno, in particolare la corteccia prefrontale (Hecht *et al.* 2015). Un altro periodo importante nell'evoluzione della cognizione umana si verificò circa 400.000 anni fa, caratterizzato da strutture sociali più organizzate tra gli *H. neanderthalensis* e, un po' più tardi, dall'emergere dei primi *H. sapiens* con sistemi sociali ancora più complessi. È quasi impossibile credere che questi progressi evolutivi siano avvenuti senza alcuna forma di linguaggio. Durante lo stesso periodo, nel Paleolitico medio e superiore, osserviamo non solo un aumento delle dimensioni del cervello, ma anche lo sviluppo di strutture cerebrali specifiche. Tra circa 700.000 e 500.000 anni fa, nuovi geni sono comparsi nel nostro genoma e la nostra specie ha sperimentato una profonda riorganizzazione dei meccanismi regolatori che controllano il nostro genoma funzionale (Hoyt *et al.* 2022; Vollger *et al.* 2022). Sintetizzando tutti i dati attuali, è altamente verosimile ipotizzare che il linguaggio, la memoria, l'apprendimento e altri tratti cognitivi, fino a comprendere l'origine del pensiero simbolico, si siano evoluti attraverso l'azione congiunta di sistemi neurali, nuovi geni, riorganizzazioni genomiche e pratiche socio-culturali. Molto ovviamente resta da fare.

Bibliografia

Abady, I., Bar-Yosef, O. & Belfer-Cohen, A. (2020). Kebara V — A Contribution for the Study of the Middle-Upper Paleolithic Transition in the Levant. *Paleoanthropology*, 2020, 1–28.

Arnon, I. et al. (2025). What enables human language? A biocultural framework. *Science*, 390, eadq8303.

Benítez-Burraco, A. & Boeckx, C. (2014). Universal Grammar and Biological Variation: An EvoDevo Agenda for Comparative Biolinguistics. *Biological Theory*, 9, 122–134.

Berwick, R. C. & Chomsky, N. (2015). *Why Only Us: Language and Evolution*. The MIT Press.

Boeckx, C. & Benítez-Burraco, A. (2014). Globularity and language-readiness: generating new predictions by expanding the set of genes of interest. *Frontiers in Psychology*, 5, 1324.

Bruner, E. (2015). *Human paleoneurology*. Springer.

Bruner, E. & Iriki, A. (2016). Extending mind, visuospatial integration, and the evolution of the parietal lobes in the human genus. *Quaternary International*, 405, 98–110.

Bruner, E. et al. (2017a). Precuneus proportions and cortical folding: A morphometric evaluation on a racially diverse human sample. *Annals of Anatomy*, 211, 120–128.

Bruner, E. et al. (2017b). Evidence for expansion of the precuneus in human evolution. *Brain Structure and Function*, 222, 1053–1060.

Bustamante, C. D. et al. (2005). Natural selection on protein-coding genes in the human genome. *Nature*, 437, 1153–1157.

Cavalli Sforza, L. L., Menozzi, P. & Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press.

Charrier, C. et al. (2012). Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell*, 149, 923–935.

Coolidge, F. L. (2020). *Evolutionary Neuropsychology: an introduction to the evolution of the structures and functions of the human brain*. Oxford University Press.

Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. W. W. Norton & Co. Inc.

Decembrini, V. et al. (2026). Earliest geometries: A cognitive investigation of Howiesons Poort engraved ostrich eggshells. *PLoS ONE*, 21, e0338509.

Dediu, D. & Levinson, S. C. (2018). Neanderthal language revisited: not only us. *Current Opinion in Behavioral Sciences*, 21, 49–55.

Dennis, M. Y. et al. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell*, 149, 912–922.

DeSilva, J. M. et al. (2019). One small step: A review of Plio-Pleistocene hominin foot evolution. *American Journal of Physical Anthropology*, 168, 63–140.

DeSilva, J. M. et al. (2021). When and Why Did Human Brains Decrease in Size? A New Change-Point Analysis and Insights From Brain Evolution in Ants. *Frontiers in Ecology and Evolution*, 9.

Di Vincenzo, F. & Manzi, G. (2013). Social learning and origin of the language faculty by means of natural selection. *Journal of Anthropological Sciences*, 91, 261–267.

Di Vincenzo, F. & Manzi, G. (2023). Homo heidelbergensis as the Middle Pleistocene common ancestor of Denisovans, Neanderthals and modern humans. *Journal of Mediterranean Earth Sciences*, 15, 161–175.

Fedorenko, E., Piantadosi, S. T. & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature*, 630, 575–586.

Fisher, S. E. (2019). Human Genetics: The Evolving Story of FOXP2. *Current Biology*, 29, R65–R67.

Fitch, W. T. (2018). What animals can teach us about human language: the phonological continuity hypothesis. *Current Opinion in Behavioral Sciences*, 21, 68–75.

Fitch, W. T. et al. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2, e1600723.

Florio, M. et al. (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science*, 347, 1465–1470.

Franchini, L. F. (2021). Genetic Mechanisms Underlying Cortical Evolution in Mammals. *Frontiers in Cell and Developmental Biology*, 9, 591017.

Franchini, L. F. & Pollard, K. S. (2015). Can a few non-coding mutations make a human brain? *Bioessays*, 37, 1054–1061.

Friederici, A. D. & Becker, Y. (2025). The core language network separated from other networks during primate evolution. *Nature Reviews Neuroscience*, 26, 131–132.

Fu, Q. et al. (2016). The genetic history of Ice Age Europe. *Nature*, 534, 200–205.

Gallardo, G. et al. (2023). Morphological evolution of language-relevant brain areas. *PLoS Biology*, 21, e3002266.

Gomez-Robles, A. (2019). Dental evolutionary rates and its implications for the Neanderthal-modern human divergence. *Science Advances*, 5, eaaw1268.

Gomez-Robles, A. et al. (2025). Accelerated evolution increased craniofacial divergence between humans and great apes. *Proceedings of the Royal Society B: Biological Sciences*, 292, 20251507.

Gould, S. J. (1997). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 10750–10755.

Gould, S. J. & Lewontin, R. C. (1979). Spandrels of San-Marco and the Panglossian Paradigm – a Critique of the Adaptationist Program. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 205, 581–598.

Gould, S. J. & Vrba, E. S. (1982). Exaptation – a missing term in the science of form. *Paleobiology*, 8, 4–15.

Guffanti, G. et al. (2018). Novel Bioinformatics Approach Identifies Transcriptional Profiles of Lineage-Specific Transposable Elements at Distinct Loci in the Human Dorsolateral Prefrontal Cortex. *Molecular Biology and Evolution*, 35, 2435–2453.

Guffanti, G. et al. (2019). Transposable Elements. *Current Topics in Behavioral Neurosciences*, 42, 221–246.

Gunz, P. et al. (2012). A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. *Journal of Human Evolution*, 62, 300–313.

Gunz, P. et al. (2019). Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity. *Current Biology*, 29, 120–127.

Haeusler, M. et al. (2019). Morphology, pathology, and the vertebral posture of the La Chapelle-aux-Saints Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, 3, 201820745.

Harris, D. N. et al. (2023). Diverse African genomes reveal selection on ancient modern human introgressions in Neanderthals. *Current Biology*, 10.1016/j.cub.2023.09.066.

Harris, I. M. & Miniussi, C. (2003). Parietal lobe contribution to mental rotation demonstrated with rTMS. *Journal of Cognitive Neuroscience*, 15, 315–323.

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.

Hecht, E. E. et al. (2025). Individual variation in the chimpanzee arcuate fasciculus predicts vocal and gestural communication. *Nature Communications*, 16, 3681.

Higuchi, R. et al. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312, 282–284.

Hoyt, S. J. et al. (2021). From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *bioRxiv*, 10.1101/2021.07.12.451456.

Hoyt, S. J. et al. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science*, 376, eabk3112.

Jerison, H. J. (1973). *Evolution of the Brain and Intelligence*. Academic Press.

Kaas, J. & Herculano-Houzel, S. (2026). *Evolution of Nervous Systems*. Academic Press.

Li, G. et al. (2022). Translating words into actions in working memory: The role of spatial-motoric coding. *Quarterly Journal of Experimental Psychology*, 75, 1959–1975.

Li, Y. et al. (2022). Enhanced effective connectivity from the middle frontal gyrus to the parietal lobe is associated with impaired mental rotation after total sleep deprivation: An electroencephalogram study. *Frontiers in Neuroscience*, 16, 910618.

Longa, V. M. (2013). The evolution of the Faculty of Language from a Chomskyan perspective: bridging linguistics and biology. *Journal of Anthropological Sciences*, 91, 15–62.

Lyu, D. et al. (2021). A Precuneal Causal Loop Mediates External and Internal Information Integration in the Human Brain. *Journal of Neuroscience*, 41, 9944–9956.

Macciardi, F. et al. (2020). Accelerated mutation rate of genes controlling for cognitive brain functions in humans. *The 89th Annual Meeting of the American Association of Physical Anthropologists*. Los Angeles.

Macciardi, F. & Martini, F. (2022). The Neanderthal Brain: biological and cognitive evolution. In F. Romagnoli, F. Rivals & S. Benazzi (eds), *Updating Neanderthals* (89–108). Elsevier.

Macciardi, F. et al. (2025). Dealing with measures: Archaeology, cognitive neuroscience and the evolutionary origin of numerical thinking. *L'Anthropologie*, 129.

MacDermot, K. D. et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American Journal of Human Genetics*, 76, 1074–1080.

Malafouris, L. (2013). *How Things Shape the Mind. A theory of material engagement*. MIT Press.

Maricic, T. et al. (2013). A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Molecular Biology and Evolution*, 30, 844–852.

Marnetto, D. et al. (2014). Genome-wide identification and characterization of fixed human-specific regulatory regions. *American Journal of Human Genetics*, 95, 39–48.

Marnetto, D. et al. (2022). Ancestral genomic contributions to complex traits in contemporary Europeans. *Current Biology*, 32, 1412–1419.

Martins, P. T. & Boeckx, C. (2019). Language evolution and complexity considerations: The no half-Merge fallacy. *PLoS Biology*, 17, e3000389.

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology*, 21, 197–216.

McDougall, I., Brown, F. H. & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, 433, 733–736.

Melchionna, M. et al. (2020). From Smart Apes to Human Brain Boxes. A Uniquely Derived Brain Shape in Late Hominins Clade. *Frontiers in Earth Science*, 8.

Moore, M. W. (2020). Hominin Stone Flaking and the Emergence of 'Top-down' Design in Human Evolution. *Cambridge Archaeological Journal*, 30, 647–664.

Moore, M. W. (2024). Hominin Stone Tool Manufacture: Design Space Constraints and Implications for Interpreting Cognitive Evolution. In T. Wynn, K. A. Overmann & F. L. Coolidge (eds), *The Oxford Handbook of Cognitive Archaeology* (C20S1–C20S13). Oxford University Press.

Mounier, A. & Mirazón Lahr, M. (2019). Deciphering African late middle Pleistocene hominin diversity and the origin of our species. *Nature Communications*, 10, 3406.

Neubauer, S., Hublin, J. J. & Gunz, P. (2018). The evolution of modern human brain shape. *Science Advances*, 4, eaao5961.

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, 314, 644–645.

Pääbo, S. et al. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, 38, 645–679.

Pauling, L. et al. (1963). Chemical Paleogenetics: Molecular "Restoration Studies" of Extinct Forms of Life. *Acta Chemica Scandinavica*, 17, 9–16.

Rilling, J. K. et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, 11, 426–428.

Ryzhkova, A. S. et al. (2022). HARs: History, Functions, and Role in the Evolution and Pathogenesis of Human Diseases. *Cell and Tissue Biology*, 16, 499–512.

Schlebusch, C. M. et al. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 358, 652–655.

Sherwood, C. C. & Bradley, B. J. (2019). Brain Evolution: Mapping the Inner Neandertal. *Current Biology*, 29, R95–R97.

Sierpowska, J. et al. (2022). Comparing human and chimpanzee temporal lobe neuroanatomy reveals modifications to human language hubs beyond the frontotemporal arcuate fasciculus. *Proceedings of the National Academy of Sciences of the United States of America*, 119, e2118295119.

Smocovitis, V. B. (1992). Unifying biology: the evolutionary synthesis and evolutionary biology. *Journal of the History of Biology*, 25, 1–65.

Takashima, A. et al. (2020). Speaking in the Brain: The Interaction between Words and Syntax in Sentence Production. *Journal of Cognitive Neuroscience*, 10.1162/jocn_a_01563, 1–18.

Tilot, A. K. et al. (2021). The Evolutionary History of Common Genetic Variants Influencing Human Cortical Surface Area. *Cerebral Cortex*, 31, 1873–1887.

Tooby, J. & Cosmides, L. (2015). Conceptual Foundations of Evolutionary Psychology. In *The Handbook of Evolutionary Psychology* (5–67). John Wiley & Sons.

Trujillo, C. A. et al. (2021). Reintroduction of the archaic variant of NOVA1 in cortical organoids alters neurodevelopment. *Science*, 371.

Vidal, C. M. et al. (2022). Age of the oldest known Homo sapiens from eastern Africa. *Nature*, 10.1038/s41586-021-04275-8.

Vollger, M. R. et al. (2022). Segmental duplications and their variation in a complete human genome. *Science*, 376, eabj6965.

Vygotskij, L. S. (2008). *Pensiero e linguaggio. Ricerche psicologiche* [ed. orig. 1934]. Laterza.

Wynn, T. (2002). Archaeology and cognitive evolution. *Behavioral and Brain Sciences*, 25, 389–402.

Wynn, T. & Coolidge, F. L. (2016). Archeological insights into hominin cognitive evolution. *Evolutionary Anthropology*, 25, 200–213.

Wynn, T., Overmann, K. A. & Malafouris, L. (2020). 4E cognition in the Lower Palaeolithic. *Adaptive Behavior*, 29, 99–106.

Zhang, X. et al. (2021). The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proceedings of the National Academy of Sciences of the United States of America*, 118.

La natura simbolica e combinatoria dell'informazione biologica. A livello genetico, neuropsicologico e linguistico

Franco Fabbro

Laureato in Medicina e Chirurgia (Padova 1982) e specializzato in Neurologia (Verona 1986). Ha svolto attività di ricercatore e medico neurologo presso l'Ospedale infantile Burlo Garofalo di Trieste. Per più di vent'anni ha svolto il ruolo di professore ordinario di Fisiologia, Neuropsichiatria infantile e Psicologia clinica presso l'Università di Udine.

franco.fabbro@hotmail.com

The origin of biological information is linked to the appearance of the first RNA and DNA polymers over four billion years ago. In these polymers, information is linked to the sequence of four different nucleotides. Beyond the genetic level, information plays a crucial role in the regulation of cells and between cells. The presence of a nervous system in multicellular animals has given rise to analog and digital modes of organization of biological information. In some animal groups (chordates, arthropods, mollusks), the complexity of the nervous system has favored the emergence of psychic activity, a predominantly predictive informational dimension. The biological and mental characteristics present in humans have made possible the invention of articulated language, a tool for promoting the social sharing of the mind's imaginative products. Symbolic and combinatorial characteristics are present at each of the levels at which biological information manifests itself (genetic, cellular, neuronal, psychic, and linguistic), particularly at the levels of DNA, the psyche, and language.

— ORDINE/DISORDINE
— ORGANISMI
— COMUNICAZIONE
— CERVELLO
— MENTE

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 33 — 45

Il concetto di informazione è stato elaborato nella prima metà del Novecento nell'ambito delle ricerche volte a mettere a punto dei sistemi automatici di artiglieria contraerea. Durante la Seconda Guerra mondiale gli Stati Uniti d'America cercarono di definire dei sistemi di contraerea automatizzati per abbattere i bombardieri tedeschi che volavano a 10.000 metri d'altezza, alla velocità di 500 chilometri orari. In queste condizioni un proiettile di artiglieria impiegava più di 10 secondi per raggiungere il bersaglio, in questo lasso di tempo il bombardiere si era spostato di più un chilometro. Per riuscire a colpire il bersaglio era necessario sviluppare delle previsioni ed informarsi sul risultato degli spari eseguiti.

Responsabile del team di ricerca sulla messa a punto di questi sistemi automatizzati di contraerea fu nominato Norbert Wiener, un docente di matematica del Massachusetts Institute of Technology (MIT) di Boston. Egli sviluppò dei sistemi di matematica statistica per predire il punto in cui si sarebbe trovato l'aeroplano in un determinato momento. Inoltre, introdusse nel sistema la possibilità di essere informato sul risultato della predizione, cioè la valutazione della "differenza" tra il punto dove era stato ipotizzato trovarsi il bombardiere e il punto effettivo in cui si trovava (misurato attraverso dei sistemi Radar). Wiener introdusse in questi sistemi dei processi di retroazione negativa (feedback negativo) che permettevano di valutare l'efficacia delle predizioni e di migliorarle monitorando costantemente il risultato dei tiri.

Nell'elaborazione di questi dispositivi automatizzati Wiener si accorse che nei sistemi elettromeccanici di contraerea non circolavano soltanto deboli correnti elettriche, ma anche un'entità del tutto nuova ed enigmatica, cioè l' "informazione". Si trattava di qualcosa che aveva sì necessità di supporti materiali, ma che era indipendente da tali supporti. Infatti, in un celebre passo di un suo libro Wiener sostenne che l'informazione non era «né materia né energia» (Wiener 1948, 199; cfr. Bynum 2016).

1. L'informazione in ambito tecnologico

Qualche decennio prima, Ralph Hartley, un esperto di comunicazione che lavorava presso la Bell Telephone Company di Boston, aveva sviluppato un modello matematico per misurare la quantità di informazione trasmessa lungo una linea telegrafica (Hartley 1928). Secondo Hartley durante una comunicazione erano sempre coinvolti dei simboli (lettere dell'alfabeto, sequenze di punti e linee), che per comune accordo, veicolano dei significati. Egli propose di eseguire dei calcoli matematici sui simboli senza curarsi del loro significato. Ciascun simbolo rappresenta una scelta tra un insieme di simboli possibili. Quindi, la «quantità di informazione» trasmessa (n) era in relazione logaritmica con il numero di simboli possibili (s), ossia con le dimensioni dell'alfabeto.

Le riflessioni di Hartley e di Wiener vennero successivamente raccolte e rielaborate da Claude Shannon, un giovane matematico e ingegnere che lavorava presso il MIT (Shannon e Weaver 1949). Egli propose una teoria matematica della comunicazione organizzata secondo un modello lineare simile alla trasmissione dei messaggi telegrafici. Il modello elaborato da Shannon era costituito da: i) una *sorgente*, in grado di generare un *messaggio* (un essere umano o una macchina); ii) un *sistema codificatore*

(che trasformava – mediante il codice Morse – le lettere dell'alfabeto in sequenze di punti e linee); iii) un *emettitore* che produceva i *segnali*; iv) un *canale* che permetteva di trasmettere i segnali; v) un *ricevitore* che invertiva l'operazione eseguita dall'emettitore; vi) un *decodificatore* che trasformava le sequenze di punti e linee in lettere dell'alfabeto; vii) un *destinatario* che riceveva il messaggio trasmesso.

Il modello sviluppato da Shannon costituiva una forma semplificata della comunicazione poiché non erano coinvolti gli aspetti semantici, che egli considerava, da un punto di vista ingegneristico, irrilevanti. Nel modello di Shannon l'informazione contenuta in un messaggio veniva associata all'incertezza e alla probabilità. Meno è probabile un messaggio, più è informativo. Inoltre, essa veniva collegata all'"entropia". «Trasmettere informazione significa trasmettere ordine. [...] un messaggio (cioè una sequenza di segni) trasmette tanta più informazione (cioè tanto più ordine) quanto maggiore è la sua entropia, cioè il suo potenziale disordine» (Silvestrini 2011, 69).

Sviluppando i concetti di incertezza, probabilità ed entropia, Shannon fu in grado di elaborare un'equazione che permetteva di misurare la quantità di informazione rappresentata in un messaggio (indipendentemente dal suo significato) (Shannon & Weaver 1949, 53-57). Poiché l'informazione non è un'entità materiale, cioè non ha a che vedere né con la massa né con l'energia, non era possibile utilizzare alcuna unità di misura del mondo fisico, cioè alcuna unità del Sistema Internazionale delle Grandezze, per questo motivo Shannon ebbe l'idea di considerare l'informazione come qualcosa che ti permette di rispondere a una domanda dove sono possibili soltanto due risposte equiprobabili: si (1) o no (0). In questo modo ogni domanda corrisponde a una cifra binaria (*Binary Digit*) ossia a un *bit*.

Qualsiasi domanda che possa prevedere una risposta binaria equiprobabile (si o no) può essere misurata in bit. Dato che il linguaggio è costituito da una sequenza di simboli (fonemi o lettere dell'alfabeto) è possibile misurarne l'informazione. Si è potuto stabilire che ogni lettera dell'alfabeto della lingua italiana può essere identificata mediante cinque domande (5 bit), quindi un articolo di circa 40.000 battute contiene mediamente 200.000 bit d'informazione.

2. La natura dell'informazione

L'"informazione" è un'entità immateriale che è presente nei processi di comunicazione, che si svolgono all'interno degli organismi viventi (regolazione genetica, metabolismo cellulare, interazione tra organi e apparati), nella comunicazione tra esseri viventi, oppure nei dispositivi tecnologici realizzati da esseri senzienti. Come è stato accennato, gli elementi fondamentali di ogni sistema di comunicazione sono: la "sorgente", il "destinatario" e il "codice" (un alfabeto, il codice Morse, il codice genetico, eccetera). Ogni messaggio prima di essere inviato viene codificato in una stringa di segni o simboli. Con la parola "segno" intendiamo "qualcosa che rinvia a qualcos'altro", in accordo con l'espressione latina coniata dai filosofi medievali: *aliquid stat pro aliquo* (Mazzone 2005, 16); invece, con la parola "simbolo" intendiamo una tipologia di segni nei quali il rapporto tra

la rappresentazione e la cosa rappresentata è completamente arbitrario (Peirce 1931-1935, §228; Callaghan 1978, 29).

Diversi fisici e filosofi si sono interrogati sulla natura dell'informazione. Una parte rilevante di questi autori aderisce a una visione fisicalista sostenendo che l'informazione deve essere considerata come una entità del mondo fisico, in grado di fluire nei sistemi, di essere accumulata, immagazzinata e convertita da una forma a un'altra (Lombardi & Lopez 2021). Ad esempio per Nathan Harshman l'informazione è una proprietà del sistema, indipendente dall'osservatore (Harshman 2016). Per Rolf Landauer (1991) l'informazione è fisica, e per questo non deve essere considerata come un'entità astratta o un'entità disincarnata. Di parere opposto era Norbert Wiener, che oltre a essere un matematico aveva avuto anche una formazione filosofica; egli sosteneva che l'informazione non aveva uno status ontologico collegato né alla massa, né all'energia (Wiener 1948, 199; cfr. von Baeyer 2003, 25).

Un contributo fondamentale alla precisazione della natura dell'informazione è stato avanzato da Gregory Bateson (1972). Secondo questo celebre pensatore britannico il concetto di informazione poteva essere compreso soltanto all'interno della dimensione biologica, che egli definiva come "mondo della Creatura". A suo parere, soltanto con l'origine dei primi organismi biologici si era realizzata una "separazione" tra l'organismo (la prima cellula) e il resto del mondo. Ed è all'interno di queste separazioni (che debbono essere considerate secondo una prospettiva funzionale, perché tutti gli organismi viventi sono in costante interazione con il loro ambiente) che è possibile generare strutture molto complesse (ordine) e riconoscere *differenze*.

Bateson ha sostenuto che le differenze sono qualcosa di astratto. Non pertinenti al mondo della Fisica, un mondo (che egli definisce Pleroma) dove sono presenti soltanto «forze, urti e scambi di energia» (Bateson 1972, 469). La "differenza" è un concetto che appartiene al mondo della comunicazione. Ma la comunicazione è possibile soltanto se si ammette la possibilità di separare, qualcosa da qualcos'altro, se cioè si ammette l'esistenza di entità separate come gli organismi (o le macchine create da esseri senzienti). Per sopravvivere in un ambiente complesso, gli organismi viventi debbono essere dotati di sistemi di valutazione (organi di senso), in grado di generare dei modelli (impliciti o espliciti) del loro mondo. Attraverso gli organi di senso alcuni esseri viventi sono in grado di riconoscere in un mucchio di palline, quelle nere da quelle bianche, e se lo desiderano di separarle (mettendo le palline bianche in un contenitore e quelle nere in un altro). In questa maniera gli organismi viventi sono in grado di generare ordine, cioè creare qualcosa di improbabile.

La presenza di organismi viventi, dotati di sistemi di valutazione (recettori sensoriali) permette di generare una dicotomia fondamentale tra il *territorio* e la *mappa*. Il territorio è costituito dall'universo materiale (Pleroma) nel quale sono situati gli organismi viventi (Creatura). La mappa è costituita dalla rappresentazione del territorio generata da un organismo vivente (o da una macchina). Bateson si è chiesto: quali parti del territorio vengono rappresentate sulla mappa? Se il territorio fosse completamente uniforme, cioè senza valli, fiumi, colline, montagne, strade, edifici, eccetera, sulla mappa non verrebbe riportato nulla. Ciò che viene

trasferito su di una mappa sono le *differenze* di quota, di vegetazione, di superficie. «Le differenze sono le cose che vengono riportate su di una mappa» (Bateson 1972, 468).

Come è stato detto, le differenze sono entità astratte che appartengono al mondo della comunicazione. Un oggetto del mondo newtoniano, come un pezzo di gesso, un sasso o una sedia, presenta per ogni loro atomo e per ogni molecola un numero infinito di differenze (che riguardano la posizione in cui si possono trovare le particelle elementari, gli atomi, le molecole), tra queste infinite possibilità un organismo ne sceglie un numero limitato; queste differenze diventano informazioni. «Ciò che intendiamo per informazione (per unità elementare di informazione) è una differenza che produce una differenza» (Bateson 1972, 470; Bynum 2016, 206). Qualche decennio dopo, Giuseppe O. Longo, uno dei più noti teorici dell'informazione, ha proposto di considerare come informazione: «una differenza che genera una differenza *per qualcuno*» (Longo 1998, 33; cfr. Longo & Vaccaro 2013; von Foerster 1980, 1984).

3. Il ruolo dell'informazione a livello del DNA

La vita è comparsa sulla Terra circa 4,2 miliardi di anni fa, quando è comparso il primo organismo vivente, chiamato *Last Universal Common Ancestor* (LUCA). Da questo primo organismo derivano tutti gli esseri viventi presenti sul nostro pianeta. Sono state proposte diverse ipotesi sull'origine della vita sulla Terra, ma in realtà non abbiamo idee chiare su questo avvenimento. Alcune strutture presenti nelle cellule, ad esempio i ribosomi, sono così complesse che è molto difficile immaginare la loro evoluzione e il loro assemblaggio (de Duve 2005; Nurse 2020). Un dato significativo è che gli atomi che compongono le diverse strutture degli organismi viventi sono gli stessi presenti nella materia ordinaria.

La vita è costituita da una serie di processi che si svolgono in un organismo. I più semplici organismi noti sono i batteri, composti da un'unica cellula (molti dei quali non possiedono un nucleo). Gli organismi viventi sono delle entità complesse, delimitate (funzionalmente) nello spazio, lontane dall'equilibrio termodinamico, in continua interazione con l'esterno, capaci di sopravvivere (cioè di mantenere una condizione definita come omeostasi) e di riprodursi. Possono essere paragonati a degli "automi" in grado di aggiustare e riprodurre se stessi. Per questa ragione – come aveva correttamente ipotizzato il matematico di origine ungherese John von Neumann (1951; 1966) – essi debbono possedere le informazioni necessarie per sopravvivere e riprodursi. Inoltre, queste informazioni debbono essere registrate su dei supporti di natura (il DNA) differenti rispetto alle strutture dell'organismo stesso (le proteine). Alla stessa stregua il progetto di una casa (le informazioni necessarie per costruire un edificio) sono rappresentate in un disegno su carta, mentre l'edificio è costruito con mattoni, cemento e tegole di terracotta.

Una delle più grandi scoperte del Novecento, compiuta da James Watson e Francis Crick, ha permesso di descrivere la struttura e le funzioni del DNA (acido desossiribonucleico), un polimero organico composto da una catena di acidi nucleici in grado di memorizzare le informazioni necessarie per permettere la sopravvivenza e la replicazione di un organismo (Watson & Crick 1953; Watson 1968; Crick 1981). Il DNA è composto da due

catene di acidi nucleici avvolte in una doppia elica e costituite da una sequenza di 4 basi azotate: Adenina, Timina, Citosina, Guanina (nell'RNA, acido ribonucleico, l'Uracile prende il posto della Timina). L'informazione genetica (DNA o RNA) è presente in tutti gli organismi viventi; negli esseri umani il filamento di DNA è lungo circa 1 metro ed è formato da 3,5 miliardi di nucleotidi.

Nel DNA umano sono contenute le informazioni relative a 20.000 geni responsabili della costruzione e della regolazione delle proteine che permettono il funzionamento e la duplicazione delle cellule del corpo umano. L'informazione codificata nel DNA è rappresentata dalla sequenza dei quattro nucleotidi all'interno di un gene (un pezzo di DNA che codifica per una proteina). Si è potuto constatare che una sequenza di tre nucleotidi (detta tripletta o codone) codifica (sceglie) uno specifico aminoacido. È stato così identificato il "codice genetico" composto da 64 triplette, ciascuna delle quali codifica (seleziona) uno (o più) dei 20 aminoacidi presenti negli organismi viventi. Cioè ciascuna tripletta di nucleotidi (codone) rappresenta in maniera simbolica un determinato aminoacido, ad esempio CAA è collegato all'aminoacido Glutammina; AAA all'aminoacido Lisina e così via. Alcune triplette danno il segnale di partenza, altre il segnale di stop. Si è scoperto che il codice genetico è lo stesso per tutti gli organismi, ciò significa che tutti gli esseri viventi presenti sulla Terra discendono da un unico organismo ancestrale (LUCA).

Nelle cellule eucariote (dotate di nucleo) il DNA viene trasformato in RNA messaggero (mRNA), questo RNA esce dal nucleo e si porta nel citoplasma. L'incontro tra l'RNA messaggero, i ribosomi e gli RNA transfer permette di costruire la catena degli aminoacidi che costituisce la struttura primaria delle proteine. Nel citoplasma le proteine assumono velocemente una struttura terziaria o quaternaria (formata dall'assembramento di più proteine), queste strutture sono responsabili della funzione delle proteine (come enzimi, pompe ioniche o proteine strutturali).

A questo punto è importante sottolineare che la struttura materiale del DNA pur essendo fondamentale non è direttamente responsabile dell'informazione genetica. Per chiarire questo punto immaginiamo di raccogliere le quattro «lettere» che compongono il DNA in quattro raccoglitori differenti. Nel primo contenitore sono contenute circa 900.000 Adenine (A), nel secondo altrettante Timine (T), nel terzo le Citosine (C) e nel quarto le Guanine (G). I quattro contenitori contengono la stessa massa ed energia del DNA umano, ma non ne contengono l'informazione. L'informazione presente nel DNA dipende infatti dalla sequenza che le quattro lettere (A, T, C, G) assumono nella doppia elica. Questa sequenza può essere rappresentata su supporti differenti, nei nucleotidi a livello delle cellule, oppure come sequenze di lettere su dei supporti cartacei, infatti, l'informazione è indipendente dai supporti nei quali viene rappresentata.

Una parte fondamentale di quello che è stato definito come «segreto della vita» (Schrödinger 1944) è rappresentato dalle modalità di registrazione e lettura delle informazioni contenute nel DNA, nel codice genetico e nei processi di regolazione cellulare che permettono agli organismi viventi di sopravvivere e riprodursi (Fabbro 2021b; 2023a, 2024, 2026). Un aspetto rilevante della vita è la capacità di organizzare strutture complesse (lipidi, carboidrati, proteine, acidi nucleici) grazie a una serie continua di processi che necessitano un consumo di energia ottenuta dall'ambiente

(che proviene dal Sole o da altri organismi). Secondo una prospettiva gli organismi costruiscono e mantengono delle strutture ordinate (neg-entropia) a scapito di un aumento del disordine (entropia) nell'ambiente in cui si trovano a interagire.

4. L'informazione a livello del sistema nervoso

Lo sviluppo dei primi organismi viventi multicellulari, avvenuta circa 1,2 miliardi di anni fa, costituisce una tappa fondamentale dell'evoluzione dei viventi. In alcuni gruppi di organismi multicellulari (vertebrati, molluschi, artropodi, nematodi) si sono sviluppate le cellule del sistema nervoso, esse sono responsabili del coordinamento dei diversi organi e apparati che costituiscono l'organismo e dell'organizzazione del comportamento (in particolare del movimento attivo). Le cellule del sistema nervoso si differenziano da tutte le altre cellule dell'organismo per la capacità di generare impulsi elettrici (*spike*), costituiti da brusche depolarizzazioni in grado di trasmettersi attraverso delle strutture filiformi (assoni) che collegano i neuroni tra di loro (Fabbro 2016).

Negli esseri umani il sistema nervoso è formato da circa 80 miliardi di cellule nervose. Esse sono organizzate in diverse componenti: un sistema nervoso periferico (nervi sensoriali e motori); un sistema nervoso enterico (che circonda il lume dei visceri dell'apparato digerente); un sistema nervoso vegetativo (sistema simpatico e parasimpatico) e un sistema nervoso centrale (midollo spinale ed encefalo). Quasi tutti i neuroni del sistema nervoso umano sono localizzati nell'encefalo (costituito da tronco dell'encefalo, cervelletto, gangli della base, talamo, dalla sostanza bianca e dalla corteccia dei vari lobi cerebrali) (Fabbro 2021a).

In quasi tutti gli organismi multicellulari dotati di sistema nervoso si possono riconoscere tre tipologie di neuroni: i) i neuroni sensoriali (neuroni afferenti), ii) i neuroni motori (neuroni efferenti), e iii) gli interneuroni. Negli esseri umani le prime due categorie (neuroni sensoriali e motori) contano soltanto alcuni milioni di cellule; mentre più del 99% dei neuroni del sistema nervoso umano sono interneuroni. Ciò significa che la maggior parte del lavoro eseguito dal sistema nervoso consiste nella generazione di quadri di attività nervosa prevalentemente interiorizzata (di natura emozionale, cognitiva, relativa ai diversi stati di coscienza, ecc.) e nell'integrazione di compiti vegetativi e sensori-motori.

Per comprendere i principi generali del funzionamento del sistema nervoso credo sia utile analizzare i processi di trasduzione sensoriale. Nei polpastrelli delle dita è presente una serie molto complessa di recettori sensoriali. Alcuni di questi recettori sono responsabili della sensibilità tattile, per esempio i corpuscoli del Pacini, essi registrano le differenze di pressione applicate sulla cute. Quando viene applicata una forza sulla cute, la capsula del corpuscolo di Pacini si deforma e genera nella terminazione nervosa una depolarizzazione elettrica che a un certo punto può dare origine a un potenziale d'azione, un fenomeno tutto-o-nulla, che corrisponde a un segnale digitale, simile a quelli presenti nei computer.

In questi recettori sensoriali, come i corpuscoli del Pacini, il mondo fisico viene trasformato in una serie di segnali digitali (0 = assenza di segnale; 1 = presenza di un potenziale d'azione), cioè il mondo fisico assume una dimensione informazionale (Fabbro 2016, capp. 6 e 7). Il potenziale

d'azione, pur utilizzando dei processi elettrochimici (mondo del Pleroma), non è più soltanto un fenomeno fisico, ma è qualcosa in grado di generare una «differenza» nell'attività (elettrofisiologica) di un neurone e di una rete neuronale, cioè di generare una *informazione* (mondo della Creatura). Nei recettori sensoriali (tattili, termici, dolorifici, propriocettivi, visivi, uditivi olfattivi, eccetera) gli stimoli analogici (massa ed energia) vengono trasformati (codificati) in segnali digitali (fenomeni tutto-o-nulla), che sono collegati ai processi di elaborazione delle informazioni sviluppati dal sistema nervoso (von Neumann 1958, 127).

Secondo Gerald Edelman, l'integrazione delle informazioni sensoriali e la loro elaborazione, nei centri nervosi multisensoriali (ad esempio nei collicoli superiori, nel talamo e nella corteccia associativa), determina la generazione di una "scena" che comprende uno sfondo, dove sono presenti degli oggetti posti di fronte a un soggetto (Edelman 1992, 2004; Fabbro 2021a, 2023b, 2026).

5. La dimensione simbolica e informazionale della psiche

In ambito neuroscientifico, la psiche (mente) e la coscienza (primaria) sono considerati come sinonimi. La coscienza primaria, che deve essere distinta dall'autocoscienza (cioè alla consapevolezza di se stessi e del mondo), corrisponde alla scena del mondo che appare di fronte ad un soggetto la mattina al risveglio (Revonsuo 2010). Secondo Thomas Metzinger (2009), la consistenza ontologica della psiche è del tutto simile a quella della realtà virtuale, e più precisamente alla "realtà" generata da un simulatore di volo nel quale sono simulati non soltanto la cabina di pilotaggio, l'aereo, le condizioni di volo, ma anche il pilota stesso. Dunque la psiche è un insieme di modelli del mondo (e di se stesso), che gli individui proiettano all'esterno e che ritengono essere del tutto reali (fenomeno della trasparenza).

Uno dei primi autori a proporre un modello della mente come sistema simbolico è stato lo psicologo scozzese Kenneth Craik. Secondo questo autore tre processi generali caratterizzano la psiche (Craik 1943, 32). Il primo consiste nella "traduzione" degli stimoli del mondo esterno in "simboli" (costituita negli esseri umani da immagini, parole e numeri). Il secondo processo è costituito dall'elaborazione mentale dei simboli (mediante i processi di simulazione, ragionamento e calcolo). Il terzo processo è costituito dalla ritraduzione dei simboli in processi esterni, ossia in azioni. Craik ha sottolineato come una delle modalità di simbolizzazione più generali della psiche umana sia costituita dalla misurazione. Si tratta di una forma di simbolizzazione molto potente poiché quasi ogni cosa del mondo fisico può essere misurata e quindi rappresentata mediante un numero (Craik 1943, 56).

La capacità di costruire dei "modelli del mondo" rappresenta una strategia evolutiva presente in numerosi gruppi animali (Fabbro 2021a; 2026). Questi "modelli mentali" (di cui l'esempio più eclatante sono i sogni; cfr. Thompson 2014; Fabbro 2023b) permettono agli animali di rappresentare l'ambiente, nel quale sono presenti oggetti e organismi (predatori e prede), e se stessi. Uno dei vantaggi evolutivi collegati alla capacità di generare dei modelli mentali è collegato alla possibilità di sviluppare simulazioni (progetti, previsioni, ipotesi), che rendono possibile la capacità di scegliere una delle diverse opzioni possibili. L'adeguatezza di queste

rappresentazioni-simulazioni (e delle scelte ad esse collegate) è dimostrata dal fatto che gli organismi viventi sono in grado di sopravvivere e a riprodursi in ambienti fisici e sociali molto complessi.

Lo studio dei diversi stati di coscienza (veglia, sonno a onde lente, sogno, sogno lucido), correlati con specifici stati dinamici interni, ha permesso di ipotizzare che la mente (umana) sia un sistema dedicato alla simulazione di scenari, che riguardano prevalentemente futuro (mente predittiva) (Clark 2024). Infatti, il futuro rappresenta una dimensione immaginativa della mente nella quale sono presenti le maggiori opportunità e i più grandi pericoli. Secondo questa prospettiva il passato (memoria episodica) deve essere considerato come una dimensione cognitiva derivata dalla capacità della mente di immaginare il futuro. Il ricordo multisensoriale del passato (memoria episodica) sarebbe un prodotto immaginativo della mente proiettato non nel futuro ma nel passato (Tulving 2002). La mente è dunque una costruzione immaginativa (simbolica), che è stata considerata – da un punto di vista esperienziale ed ontologico – simile all'attività onirica o alla visione di un film (Seth 2019, Naccache 2020). Tale attività viene «aggiornata» in maniera continuativa dalle informazioni sensoriali, che selezionano la migliore simulazione o film mentale a disposizione.

6. Il linguaggio come sistema simbolico e informazionale

Gli studi di linguistica e le riflessioni filosofiche hanno confermato la natura simbolica del linguaggio e delle lingue umane. I concetti più diffusi, come quelli che rappresentano le parti del corpo, oppure gli alimenti (acqua, pane, riso) o gli animali più comuni (cane, gatto, mucca), vengono rappresentati da parole che sono differenti in ciascuna lingua. La stringa di fonemi che mi permette di pronunciare la parola [*pane*] è completamente differente dalla parola utilizzata in inglese [*bred*]. Le due stringhe di fonemi sono delle rappresentazioni simboliche (arbitrarie) del concetto di pane.

Diverse indagini, sviluppate in ambito antropologico, linguistico e genetico, hanno permesso di capire che il linguaggio umano non si è evoluto nel corso di milioni di anni, come a lungo si è erroneamente creduto, ma è stato «inventato» (in maniera inconsapevole) meno di 100.000 anni fa (Tattersall 2012; Fabbro 2021a, 2021b). Poiché gli esseri umani (*Homo sapiens*) sono presenti sulla Terra da almeno 250.000 anni, ne deriva che per 150.000 anni non hanno avuto a disposizione il linguaggio.

Alcune ricerche, sviluppate alla fine dello scorso secolo, hanno evidenziato che gruppi di bambini (sordomuti) sufficientemente numerosi, sono in grado di inventare, nell'arco di due generazioni, un linguaggio dal nulla (Senghas *et al*. 2004; Senghas 2005). Per essi il linguaggio non rappresenta soltanto un potenziamento della capacità di condividere prodotti mentali (Dor 2015) ma anche una forma di gioco, simile ai giochi delle carte (briscola, tressette, scopa, scala quaranta, poker), che ritengo abbiano molto a che vedere con il linguaggio, con la psiche e con il DNA (in particolare per quanto riguarda gli aspetti combinatori) (Jacob 1970; Fabbro 2024, 2026).

Il linguaggio, come tecnologia in grado di permettere di condividere prodotti immaginativi della mente, ha permesso agli umani: i) di uscire

da una condizione di estrema solitudine mentale; ii) di condividere conoscenze. Alcuni individui sordomuti, che non conoscevano alcuna lingua (*languageless*), dopo aver appreso il linguaggio (attraverso l'acquisizione di una lingua dei segni) hanno potuto descrivere l'estrema solitudine della loro vita mentale prima dell'appropriazione del linguaggio (Schaller 1991). In secondo luogo il linguaggio si è rivelato un sistema molto efficace per trasmettere informazioni e conoscenze. Infatti, il linguaggio della scienza non è – come credono molti – la matematica (che rimane uno strumento formidabile) ma il linguaggio naturale (Smolin 2013, 253).

Una delle caratteristiche che separano gli esseri umani dagli altri animali è costituita dalla tecnologia, ossia dalla sistematica capacità di costruire, conservare e utilizzare strumenti, presente soltanto negli esseri umani. Come è stato detto, anche il linguaggio è una tecnologia, probabilmente una delle tecnologie più potenti che gli esseri umani hanno inventato (Leroi-Gourhan 2018; Dor 2015). E come tutte le tecnologie, in particolare quelle offensive (pistole, fucili, droni, missili, bombe nucleari), anche il linguaggio non è completamente controllato dagli esseri umani. Martin Heidegger ha sostenuto che il linguaggio ha una sorta di autonomia; nei suoi scritti ha sostenuto che "il linguaggio parla" (Heidegger 1950, 28). Non sembra sia possibile eliminare tutte le armi presenti sulla Terra (cfr. Heidegger 1954; Severino 2002, 2019), così non sembra essere mai stato un compito facile tacere (Wittgenstein 1921).

7. Conclusioni: verso una filosofia dell'informazione biologica

Gli esseri umani tendono a dimenticare che la scienza e la filosofia sono dei prodotti dell'ingegno umano. Esse dipendono da variabili storico-culturali ma anche da alcuni vincoli biologici determinanti. Entrambe presuppongono l'esistenza del linguaggio, uno strumento che favorisce la diffusione dei prodotti immaginativi della mente (Dor 2015; Fabbro 2024). Tuttavia, senza una specifica strutturazione della mente, nella quale sono presenti alcune caratteristiche cognitive peculiari, come l'attenzione congiunta, il sistema per viaggiare mentalmente nel tempo, la teoria della mente, eccetera, non è possibile acquisire e utilizzare alcun linguaggio (Tomasello 2014; Corballis 2018). Inoltre, queste specifiche caratteristiche mentali sono in relazione con alcune strutture del sistema nervoso centrale, in particolare con una specifica organizzazione neuroanatomica e funzionale del «sistema frontale» (lobo frontale, gangli della base ed emisferi cerebellari), presente unicamente negli esseri umani (Fabbro 2021a, 2021b).

Nell'affrontare il problema di che cosa sia l'informazione biologica è utile considerare la presenza di una serie di differenti livelli nella natura. Il livello fondamentale è quello microfisico, descritto dalla meccanica quantistica, una teoria che ha ricevuto numerosissime conferme sperimentali, anche se è stata considerata per molti aspetti inverosimile e incomprensibile (Feyman 1964; Mussardo 2025), poiché a livello quantistico si evidenzia una serie di fenomeni inaspettati come: la duplice natura delle particelle elementari (corpuscolare e ondulatoria) (Lederman & Hill 2011), il principio di indeterminazione, che stabilisce dei limiti nella misurazione delle grandezze fisiche (Heisenberg 1958; de Stefano 2022) e il fenomeno dell'*entanglement* (che suggerisce un carattere non locale della realtà microfisica) (Aczel 2001).

Le stranezze riscontrate a livello quantistico scompaiono nel livello successivo, quello degli "atomi", un ambito nel quale conoscenze scientifiche possono considerarsi come sostanzialmente complete (Cromer 1993, 20). Gli atomi sono composti da un nucleo (concentrato in un piccolissimo spazio), composto da protoni (con carica positiva) e neutroni, attorno al quale orbitano gli elettroni, con carica negativa e massa molto più piccola del nucleo. Ciascun protone è composto da 2 quark up e 1 quark down; mentre ciascun neutrone è composto da 1 quark up e 2 quark down. L'atomo più semplice è quello dell'Idrogeno (H), di cui l'isotopo più diffuso è formato da un protone e un elettrone. Più del 70% della massa della nostra galassia è composto da idrogeno. Il secondo elemento più diffuso è l'Elio, costituisce più del 25% della massa della nostra galassia. L'elio è un atomo composto da 2 protoni, 2 neutroni e 2 elettroni. Ciò che caratterizza un elemento chimico è rappresentato dal numero degli elettroni. L'Idrogeno possiede 1 elettrone, l'Elio ne possiede 2, il Litio 3, il sodio 11 e l'Uranio, l'elemento più pesante presente sulla Terra possiede 92 protoni e lo stesso numero di elettroni (sono stati descritti altri 26 elementi transuranici con più di 92 protoni).

Ogni cosa sulla Terra è formata dai 92 elementi chimici presenti in natura (dall'idrogeno all'uranio). A loro volta, questi elementi sono composti dalla combinazione di alcune particelle elementari (quark up, quark down, elettroni). Le proprietà dei diversi elementi chimici determinano la possibilità che gli atomi hanno di legarsi tra di loro per formare delle molecole. L'acqua, una delle molecole più importanti per la biologia, è formata da due atomi di idrogeno e uno di ossigeno (H_2O). Le più importanti strutture che compongono le cellule sono formate da "macromolecole", come i lipidi (membrane cellulari), i carboidrati (fonti di energia), le proteine (enzimi, canali e pompe ioniche), gli acidi nucleici (DNA e RNA). Come è stato accennato, le cellule costituiscono il livello biologico fondamentale, da cui si originano tutti i livelli successivi che caratterizzano il vivente, cioè gli organismi multicellulari (dotati di organi e apparati, tra cui il sistema nervoso), la psiche e, negli esseri umani, il linguaggio (Fabbro 2021b; 2024; 2026).

Gli esseri umani sono in grado di interpretare ciascuno di questi livelli come dotato di informazione. Per esempio, il fisico statunitense John Archibald Wheeler ha sostenuto la famosa frase: "*It from bit*", secondo la quale ogni grandezza fisica trae il suo significato ultimo dai bit (Wheeler 1989, 309). Tuttavia, se ci chiediamo quando sia comparsa per la prima volta l'informazione sulla Terra, la risposta più ragionevole consiste nel sostenere che l'informazione sia comparsa sul nostro pianeta circa 4,2 miliardi di anni fa, quando sono comparsi i primi polimeri di RNA e del DNA. L'origine della vita e l'origine delle informazioni coincidono. Gli elementi materiali (particelle elementari, protoni, neutroni, atomi, molecole) costituiscono le strutture fisiche attraverso cui prendono origine i fenomeni comunicativi e l'informazione; questi, a loro volta, si basano essenzialmente sugli aspetti simbolici e combinatori (tipici del DNA e del linguaggio) (Jacob 1970, Fabbro 2021b, 2024). L'informazione è quindi prima di tutto una caratteristica degli organismi viventi, essa può essere successivamente interpretata come presente nel mondo fisico e biologico, ed implementata nei dispositivi tecnologici (Fabbro 2026).

Bibliografia

Aczel, A. (2001). *Entanglement. Il più grande mistero della fisica*. Trad. it. di M. Pagani. Cortina.

Bateson, G. (1972). *Verso un'ecologia della mente*. Trad. it. di G. Longo. Adelphi.

Bynum, T.W. (2016). Informational Metaphysics. The Informational Nature of Reality (203-218). In L. Floridi (ed), *The Routledge Handbook of Philosophy of Information* . Routledge.

Callaghan, W. (1978). Introduzione a Charles S. Peirce (39-45). In C.S. Peirce (2004), *Scritti di filosofia*. Trad. it. di L. Leone. Fabbri.

Clark, A. (2024). *The Experience Machine. How Our Minds Predict and Shape Reality*. Penguin.

Corballis, M. (2018). Space, Time and Language. *Cognitive Processing*, 19 (1), 589–592.

Craik, K. (1943). Ipotesi sulla natura del pensiero (32-79). Trad. it. di G. Gabella. In K. Craik, *La fisica della mente*. Boringhieri.

Crick, F.H. (1981). *L'origine della vita*. Trad. it. di B. Vitale. Garzanti.

Cromer, A. (1993). *L'eresia della scienza. L'essenziale per capire l'impresa scientifica*. A cura di G. Giorello. Edizioni CDE.

de Duve, C. (2005). *Alle origini della vita*. Trad. it. di L. Sosio. Bollati Boringhieri, 2008.

de Stefano, F. (2022). *Dialogo sopra i massimi sistemi quantistici. Il dibattito sull'epistemologia della meccanica quantistica*. Mimesis.

Dor, D. (2015). *The Instruction of Imagination: Language as a Social Communication Technology*. Oxford University Press.

Ereditato, A. (2017). *Le particelle elementari*. Il Saggiatore.

Fabbro, F. (2016). *Le neuroscienze dalla fisiologia alla clinica*. Carocci.

Fabbro, F. (2021a). *Che cos'è la psiche. Filosofia e neuroscienze*. Astrolabio.

Fabbro, F. (2021b). *I fondamenti biologici della filosofia*. Mimesis.

Fabbro, F. (2023a). *Biological and Neuroscientific Foundations of Philosophy*. Routledge.

Fabbro, F. (2023b). *Ipnosi e cervello sociale. Neuroscienze e filosofia politica*. Mimesis.

Fabbro, F. (2024). *La svolta biopsichica. Quarantotto brevi meditazioni filosofiche controcorrente*. Astrolabio.

Fabbro, F. (2026). *Filosofia dell'informazione biologica*. Mimesis (in stampa).

Feynman, R. (1964). *The Character of Physical Law*. MIT Press.

Harshman, N.L. (2016). Physics of Information (7-14). In L. Floridi (ed), *The Routledge Handbook of Philosophy of Information*. Routledge.

Hartley, R.V. (1928). Transmission of Information. *Bell System Technical Journal*, 7 (3), 535–563.

Heidegger, M. (1950). Il linguaggio (27-44). In M. Heidegger (1973), *In cammino verso il linguaggio*. Trad. it. di A. Caracciolo e M. Caracciolo Perotti. Mursia.

Heidegger, M. (1954). La questione della tecnica (5-27). In M. Heidegger (1976), *Saggi e discorsi*. A cura di G. Vattimo. Mursia.

Heisenberg, W. (1958). *Fisica e filosofia*. Trad. it. di P. Gulmanelli. Bollati Boringhieri.

Jacob, F. (1970). *La logica del vivente. Storia dell'ereditarietà*. Trad. it. di A. Serafini e S. Serafini. Einaudi.

Landauer, R. (1991). Information Is Physical. *Physics Today*, 44, 23–29.

Lederman, L. & Hill, C. (2011). *Fisica quantistica per poeti*. Trad. it. di L. Civalleri. Bollati Boringhieri.

Leroi-Gourhan, A. (2018). *Il gesto e la parola. Tecnica e linguaggio*. Trad. it. di F. Zannino. Mimesis.

Lombardi, O. & López, C. (2021). The Many Faces of Shannon Information (324-340). In C. Meyns (ed), *Information and the History of Philosophy*. Routledge.

Longo, G.O. (1998). *Il nuovo Golem. Come il computer cambia la nostra cultura*. Laterza.

Longo, G.O. & Vaccaro, A. (2013). *Bit Bang. La nascita della filosofia digitale*. Apogeo.

Metzinger, T. (2009). *Il tunnel dell'io. Scienza della mente e mito del soggetto*. Trad. it. di M. Baccarini. Cortina.

Mussardo, G. (2025). *Dio gioca a dadi con il mondo. La storia della meccanica quantistica*. Castelvecchi.

Naccache, L. (2020). *Le cinéma intérieur. Projection privée au cœur de la conscience*. Odile Jacob.

Nurse, P. (2020). *Che cos'è la vita. I cinque principi fondamentali della biologia*. Trad. it. di L. Serra. Mondadori.

Peirce, C.S. (1998). *Collected Papers of Charles Sanders Peirce*. Voll. II. C. Hartshorne e P. Weiss (eds). Harvard University Press.

Revonsuo, A. (2010). *Consciousness, The Science of Subjectivity*. Psychology Press.

Schaller, S. (1991). *A Man Without Words*. University of California Press.

Schrödinger, E. (1944). *Che cos'è la vita? La cellula vivente dal punto di vista fisico*. Trad. it. di M. Ageno. Adelphi.

Senghas, A. (2005). Language Emergence: Clues From a New Bedouin Sign Language. *Current Biology*, 15 (12), R463–R465.

Senghas, A., Sotaro, K. & Özyürek, A. (2004). Children Creating Core Properties of Language: Evidence From an Emerging Sign Language in Nicaragua. *Science*, 305, 1779-1782.

Seth, A. (2019). Il creatore della realtà. *Le Scienze*, 615, 42-49.

Severino, E. (2002). *Técne. Le radici della violenza*. Rizzoli.

Shannon, C. & Weaver, W. (1949). *The Mathematical Theory of Communication*. The University of Illinois Press.

Silvestrini, V. (2011). *Che cos'è l'entropia. Ordine, disordine ed evoluzione dei sistemi*. Editori Riuniti University Press.

Smolin, L. (2013). *La rinascita del tempo. Dalla crisi della fisica al futuro dell'universo*. Trad. it. di S. Frediani. Einaudi.

Tattersall, I. (2012). *I signori del pianeta. La ricerca delle origini dell'uomo*. Trad. it. di A. Panini. Codice.

Thompson, E. (2014). *La veglia, il sogno, l'essere. Il sé e la coscienza nelle neuroscienze, nella meditazione e nella filosofia*. Trad. it. di M. Passavanti. Astrolabio.

Tomasello, M. (2014). *Unicamente umano. Storia naturale del pensiero*. Trad. it. di M. Riccucci. il Mulino.

Tonelli, G. (2023). *Materia. La magnifica illusione*. Feltrinelli.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, 53, 1-25.

von Baeyer, H.C. (2003). *Informazione. Il nuovo linguaggio della scienza*. Trad. it. di S. Bianchi. Dedalo.

von Foerster, H. (1980). Epistemology of Communication (18-27). In K. Woodward (ed), *The Myths of Information: Technology and Postindustrial Culture*. Routledge.

von Foerster, H. (1981). Disordine/ordine: scoperta o invenzione (191-201). In H. von Foerster, *Sistemi che osservano*. Trad. it. di B. Draghi. Astrolabio.

von Neumann, J. (1951). La logica degli automi e la loro autoriproduzione (222–245). In V. Somenzi (a cura di). *La filosofia degli automi*. Boringhieri.

von Neumann, J. (1958). *Computer e cervello*. Trad. it. di P. Bartesaghi. Il Saggiatore.

von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press.

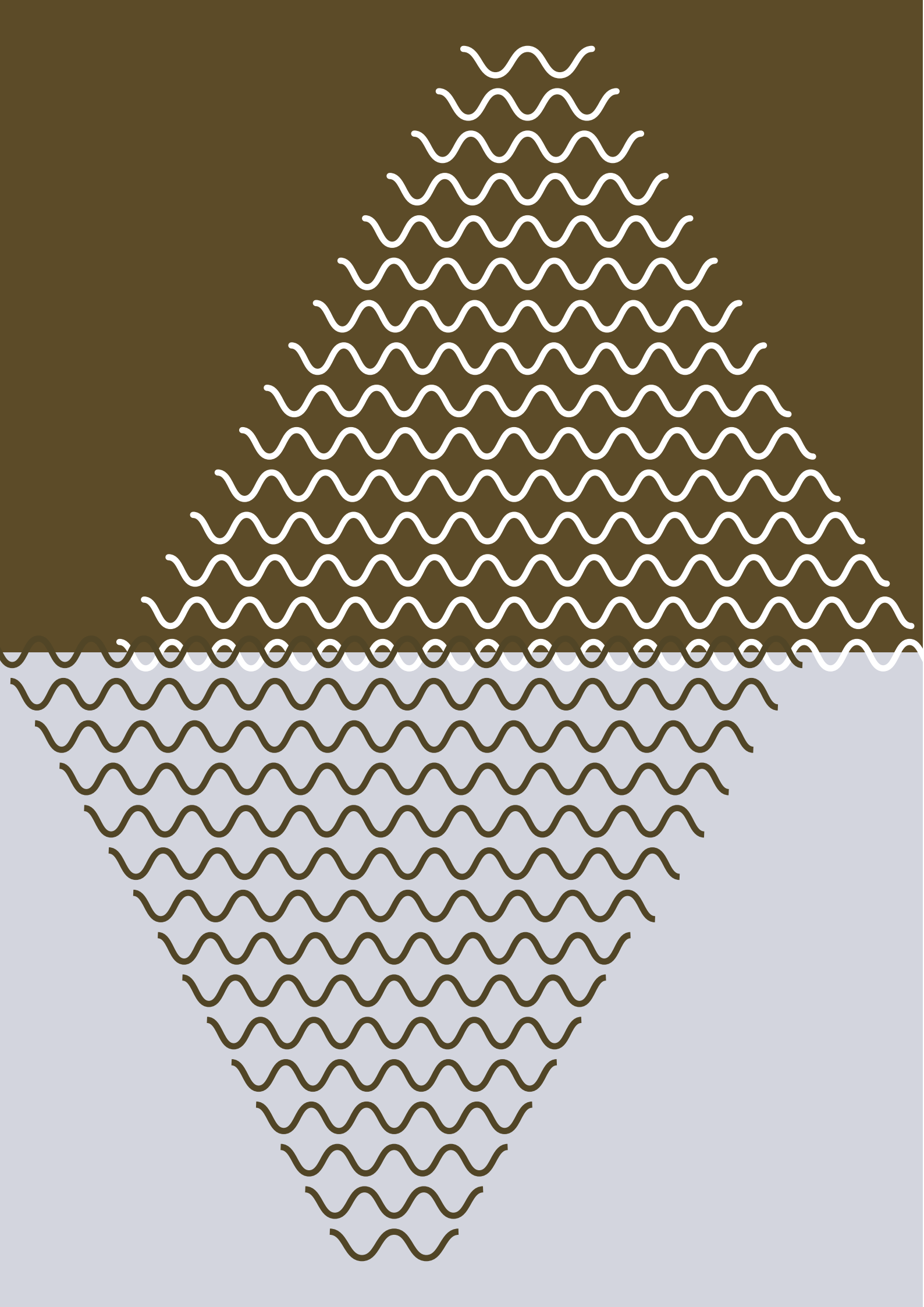
Watson, J.D. (1968). *La doppia elica. Il segreto della vita*. Trad. it. di B. Vitale e M. Attardo Marini. Garzanti.

Watson, J.D. & Crick, F.H. (1953). A Structure for Deoxyribose Nucleic Acid. *Nature*, 171 (4356), 737-738.

Wheeler, J.A. (1989). Information, Physics, Quantum: The Search for Links (309–366). In J.A. Wheeler (ed), *Proceedings III International Symposium on Foundations of Quantum Mechanics*.

Wiener, N. (1948). *La cibernetica*. Trad. it. di O. Beghelli. Bompiani.

Wittgenstein, L. (1921). *Tractatus logico-philosophicus e Quaderni 1914-1916*. Trad. it. di A.G. Conte. Einaudi.



Natura e artificio.
Il Linguaggio come Meta-Istituzione e Prassi sociale trasformativa

Nicola Garau

Laureato in Scienze filosofiche presso l'Università di Milano, ha il Baccellierato in Teologia alla Pontificia Facoltà Teologica della Sardegna. Si occupa di ermeneutica e pragmatismo. E' autore del saggio *Prassi e linguaggio. Apel interprete di Wittgenstein e Peirce* (2024).

garaunicola@gmail.com

The contribution offers a critical reinterpretation of the notion of language within the tension between nature and artifice, highlighting its institutional, performative, and historically situated dimensions. In a context dominated by biolinguistic, neuroscientific, and computational approaches, it argues for the need to suspend any reductionist impulse aimed at naturalizing linguistic functions, in order to restore language to its character as a complex symbolic praxis. The analysis begins with the paradox inherent in any attempt to determine the "nature" of language from an allegedly external standpoint: such neutrality is illusory, since language simultaneously constitutes the object, the medium, and the condition of its own description. This blind spot undermines many naturalistic and realist-metaphysical frameworks, which fail to recognize the self-implicating structure of linguistic enunciation. As an alternative, language is interpreted as a meta-institution: not a single social institution among others, but the practice that enables the formation, stabilization, and transformation of the entire institutional order. Meaning is thus understood as the effect of shared praxis, rather than the product of formal codes or purely neurocomputational processes. From this perspective, the distinction between natural and artificial language appears as a difference between evolutionary and designed forms of praxis. Artificial intelligence does not impose an algorithmic redefinition of language, but rather brings to light its historical, embodied, and co-constitutive nature as a generator of sense.

— LANGUAGE
— SYMBOLIC PRAXIS
— META-INSTITUTION
— ARTIFICIAL LANGUAGE

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 47 — 58

Introduzione

Definire la natura del linguaggio, la sua origine ed evoluzione è compito complesso ricorsivamente interpretato dalle scienze del linguaggio e dalle discipline che, a vario titolo, ne fanno oggetto di indagine. Dalla linguistica teorica alla biolinguistica, dalla neurolinguistica alle scienze cognitive, fino alla sociolinguistica e alle più recenti ricerche sull'intelligenza artificiale generativa, il linguaggio è alternativamente interpretato come funzione biologica, sistema formale, dispositivo cognitivo o prodotto culturale. Tale pluralità di approcci tende spesso a polarizzarsi attorno alla dicotomia, tanto intuitiva quanto teoricamente problematica, tra natura e artificio.

Infatti, mentre la biolinguistica e la neurolinguistica concentrano l'analisi sulle strutture biologiche e neurali che rendono possibile l'uso del linguaggio, indagandone al contempo le condizioni evolutive di insorgenza (Friederici 2011; 2017; Fedorenko, Ivanova, Regev 2024), e il generativismo pretende di individuare le regole naturali universali che sottenderebbero ogni grammatica umana postulando l'esistenza di una facoltà del linguaggio specie-specifica e relativamente autonoma rispetto alle pratiche comunicative concrete (Chomsky 1965; Chomsky 2000; Hauser, Chomsky, Fitch 2002), [1] gli approcci di matrice culturalista pongono, invece, maggiore attenzione alla dinamica comunicativa attraverso cui il linguaggio e la significatività vengono istituiti come pratiche e convenzioni sociali culturalmente situate (Vygotskij 1990; Hymes 1972; Halliday 1978).

[1] Per onestà va ricordato come Chomsky sia approdato nei suoi lavori più recenti a una visione evolutiva precisa dell'origine del linguaggio per quanto saltazionista (Berwick & Chomsky 2016)

Queste tendenze fondamentali informano il dibattito storico e continuano a determinare il dibattito attuale, anche se, è doveroso notare come sempre più autori approfondiscano l'intersezione tra biologia e cultura tematizzando la complessità dei due approcci con sempre più perspicuità e superando l'equivoco riduzionista che vedrebbe sovrapporre il termine naturale al termine biologico e il termine artificiale al termine convenzionale o culturale. [2]

Nella filosofia del Novecento, come è noto, il problema del linguaggio è centrale e, al di là degli approcci neopositivistici e analitici – che non tematizzano l'origine del linguaggio, ma perseguono l'ideale di un linguaggio formale puro o prediligono l'analisi del linguaggio ordinario – l'ermeneutica, il pragmatismo e specialmente la teoria dei giochi linguistici di Wittgenstein (2009) problematizzano l'origine del significato mettendone in questione ogni concezione indipendente dall'uso e dalla forma di vita. Da questo punto di vista l'intreccio tra dimensione biologica e culturale nel linguaggio appare inestricabile, per cui la complessità del fenomeno linguistico viene investigata dal punto di vista della prassi, cioè della tecnica come forma di vita peculiare dell'essere umano che lo porta a manipolare l'ambiente trasformandolo col suo operare e i suoi strumenti esosomatici tra cui, appunto, il linguaggio. [3]

[2] Ci riferiamo qui a Tomasello (2009), ma anche a quegli approcci neuroscientifici che accentuano la considerazione della dimensione incarnata, intersoggettiva e storicamente situata del linguaggio, interpretandolo come prodotto emergente di pratiche comunicative condivise come, per esempio, l'*embodied cognition* (Varela, Thompson, Rosch 1991; Lakoff & Johnson 1999; Gallese & Lakoff 2005).

Pur muovendosi entro una cornice filosofica, il contributo intende dialogare criticamente con approcci al linguaggio biologici, evolutivi e computazionali, mostrando come tali prospettive presuppongano pratiche simboliche e istituzionali che ne eccedono la spiegazione naturalistica.

[3] Seguendo Sini, per strumento esosomatico si intende l'estensione dell'azione corporea al di fuori del corpo, fondamento della conoscenza e della cultura. Il linguaggio, in quanto tecnologia che esternalizza azione e significato, rappresenta la forma compiuta dell'esosomatismo (Sini 2020).

Sul piano metodologico, il contributo procederà per critica ermeneutico-fenomenologica e ricostruzione semiotico-pragmatica, non per validazione empirica. L'obiettivo non è verificare o falsificare i risultati delle scienze cognitive e computazionali del linguaggio, ma interrogarne i presupposti filosofici impliciti, in particolare l'assunzione di un punto di osservazione neutro rispetto al linguaggio che si intende descrivere.

Nelle pagine che seguono si tenterà, quindi, di mettere in luce il limite e il paradosso di ogni discorso sulla natura del linguaggio a partire dal linguaggio, per avanzare l'idea che esso sia una meta-istituzione, cioè, una prassi simbolica che rende possibili tutte le istituzioni (Gehlen, Apel, Searle). La significatività, a partire da tale inquadramento, verrà, quindi, descritta secondo un approccio semiotico e pragmatista come l'effetto di prassi condivise (Wittgenstein, Peirce, Sini, Apel). Si mostrerà, per questo, come natura e artificio siano due termini incorporabili in una possibile definizione unitaria del linguaggio che trae origine dalla prassi e come, di conseguenza, sia possibile riformulare la distinzione tra linguaggio naturale e linguaggio artificiale come distinzione tra due forme di prassi: l'una involontaria e dinamicamente evolutiva, l'altra progettata, regolata e orientata alla simulazione funzionale (come nel caso dei *Large Language Models*). Anche il riferimento alla letteratura empirica recente su interpretabilità e analisi delle rappresentazioni interne dei LLM non avrà pertanto funzione probatoria, ma servirà a verificare se tali risultati confermino o complichino la tesi qui sostenuta. In quest'ottica, si mostrerà come l'intelligenza artificiale generativa non conduca a una ridefinizione del linguaggio in chiave algoritmica, ma come prassi regolata, divenga strumento di accrescimento del linguaggio che induce a ripensarne ancora una volta la complessità come forma storica, incarnata e istituzionale di co-costruzione del senso.

1. il paradosso dell'autoimplicazione di ogni discorso sulla natura del linguaggio

Ogni indagine sulla natura e l'origine del linguaggio si scontra con il limite strutturale di investigare e descrivere il linguaggio attraverso il linguaggio stesso. Ogni descrizione scientifica del fenomeno linguistico si trova, infatti, nell'inevitabile paradosso della coincidenza tra oggetto di studio e strumento di analisi. L'osservazione scientifica, invero, deve trasformarsi in descrizione linguistica per pervenire a una significatività che renda la materia accessibile e condivisibile come sapere. Ne consegue che qualunque discorso sulla natura del linguaggio risulta autoimplicato e dotato di una complessità non esauribile in un approccio autoproclamato come oggettivo (Sini 2021). Tale problematizzazione dovrebbe interessare qualunque scienza che si occupa del linguaggio, al fine di allontanare il rischio del riduzionismo che tende a considerare il linguaggio puramente come segno-rappresentazione universale di ogni oggetto e anche di sé stesso.

L'assunzione implicita di molte delle scienze sperimentali che si occupano anche del fenomeno linguistico – come ad esempio parte della biolinguistica e della neurolinguistica – muove dal presupposto metafisico che esista un mondo di fenomeni esterno alla mente e al linguaggio che vada assunto come fatto oggettivo da studiare empiricamente e che compito dello scienziato sia rilevarne le strutture, classificarne le manifestazioni,

descriverne le leggi. Questa impostazione non tiene conto dell'autoimplicazione dello sguardo del ricercatore in ogni descrizione e raggiunge un grado di limite strutturale quando l'oggetto di studio diviene proprio lo strumento di ogni descrizione, ossia il linguaggio. Carlo Sini, rievocando l'insegnamento di Steiner, mette a fuoco tale paradosso:

> Dobbiamo usare il linguaggio per studiare il linguaggio, dice Steiner ricordando una riflessione di Wittgenstein del 1915: «Ciò che si rispecchia nel linguaggio non posso esprimerlo con esso» (però *questo* lo ha espresso!). Che cosa è il linguaggio? domanda Steiner. Ma che senso avrebbe "usare il linguaggio" per rispondere? Quindi come sta la "cosa"? Il fatto è che, anzitutto, le "cose" sono due: secondo il presupposto che sorregge la domanda, ci dovrebbe essere quella "cosa" cui la parola 'linguaggio' allude o rinvia; nel contempo c'è il "problema", che l'uso del linguaggio comporta, nel chiedere di sé. Un problema che mette in crisi ogni tentativo di risposta, ogni pretesa di scavalcare il linguaggio verso una realtà "altra" da esso, sebbene, per altro verso, di continuo non si faccia che domandare e rispondere. [...] Prima dell'uso del linguaggio non ci sono "cose" da dire, né a proposito delle quali coltivare credenze, né sulle quali sollevare domande e problemi. Si badi: ho scritto "da dire"; non che non ci siano cose indipendenti dal linguaggio (sebbene di nuovo 'cose' sia una espressione linguistica, sicché di fatto il problema si ripresenta). [...] I discorsi, le parole, però, non dicono a caso. Rispecchiano e interpretano luoghi di incontro e occasioni di esperienza: decifrano il mondo, diceva Steiner, e noi assentivamo convinti. Dire 'albero' facendo segno a un albero non è immotivato. È un po' come tendere il braccio per stornare le fronde nel bosco o atteggiare la mano a coppa per prendere acqua dal lago: abiti gestuali "significativi". Questo però non significa che esistono *alberi*. Dietro la parola 'albero' sta un complesso "lavoro" analitico e comunicativo, sta il suo uso sociale, cioè la sua costitutiva associazione con l'azione del corpo e dei corpi viventi, e di questi corpi con gli strumenti esosomatici propri dell'esperienza intersoggettiva umana, a cominciare dallo strumento esosomatico della voce significativa ecc. ecc. Questa complessità, questo indistricabile intreccio di riferimenti che è al fondo di ogni parola Steiner ce lo ha mostrato molto bene. (Sini 2021, 196-197)

È, dunque, il linguaggio che, in quanto prassi discorsiva, nel complesso lavoro analitico e comunicativo, istituisce il mondo come insieme degli oggetti intelligibili e istituisce sé stesso come significatività del mondo. In questo senso il linguaggio costruisce il sapere, vale a dire, l'insieme dei giudizi che rendono una totalità di senso plausibile e condivisibile, dunque, una rappresentazione universalmente valida (Wittgenstein 2009) o, più precisamente, valida per la comunità storica della comunicazione in cui viene prodotta e accettata (Apel 1973). Ed è per questo che le scienze linguistiche dovrebbero considerare criticamente come la propria descrizione del fenomeno linguistico non rappresenti un punto di vista neutro ma una pratica di modellizzazione caratterizzata da una circolarità riflessiva in cui la natura del linguaggio viene presupposta e insieme si esplica non solo nella sostanza del discorso ma anche nella prassi discorsiva che la istituisce come narrazione scientifica.

2. Il Linguaggio come Meta-Istituzione delle Istituzioni

La natura del linguaggio si caratterizza come prassi sociale umana in cui si configurano soggetti, oggetti e significati. È prassi istituente la realtà umana, che ne sovrintende ogni agire e per questo può essere considerata propriamente un'istituzione. Il linguaggio, seguendo l'impostazione filosofica di Karl-Otto Apel, potrebbe essere considerato l'apriori della conoscenza del mondo; perciò la comunità della comunicazione – in cui il linguaggio si fa tutt'uno con la prassi intersoggettiva – coincide con il trascendentale della conoscenza della realtà (Apel 1973). In tal modo la natura del linguaggio potrebbe essere accostata non solo all'idea di istituzione, ma a quella di meta-istituzione che ordina e sovrintende ogni aspetto della vita umana, ogni pratica sociale condivisa e anche la sfera individuale della relazione con sé stessi in quanto istituzione del pensiero (Apel 1973).

Con l'idea di istituzione, in senso antropologico – secondo l'impostazione di Gehlen (1955; 1956) dalla quale trae spunto anche la riflessione di Apel (2003) – si intende in generale ogni «consolidamento e autonomizzazione del nostro scambio attivo con il mondo esterno e con gli altri, atto ad assegnare al nostro comportamento un abito obbligante» (Apel 2003, 93). Entro questa definizione ricadono tutte le pratiche umane consolidatesi in comportamenti sociali automatici, ossia, le forme sociali in cui si esprimono le diverse culture: dallo stato al mercato, dalle forme della vita pubblica alla concretizzazione nella vita privata. Le istituzioni secondo Gehlen hanno il preciso compito di incanalare i bisogni umani, sgravare l'essere umano dal peso delle pulsioni stabilendo comportamenti obbliganti, abitudini, convenzioni in cui l'uomo realizza sé stesso e crea il proprio spazio vitale.

Nel processo di creazione e rinnovamento delle istituzioni svolge un ruolo fondamentale la comunicazione e il linguaggio che è l'istituzione che sovrintende ogni altra istituzione perché precede e fonda la possibilità di ogni istituzione. Su questo tema appare rilevante anche il contributo di John Searle che partendo proprio dalla sua teoria degli atti linguistici (Searle 1969) giunge a formulare un'ontologia sociale che considera il linguaggio come il motore della fondazione di ogni oggetto sociale (Searle 1995). Searle, infatti, sostiene che il linguaggio sia quella istituzione di base che fonda ogni altra istituzione perché può nominare la realtà attribuendo ai suoi oggetti un significato diverso da quello che avrebbero considerati semplicemente per quello che sono (Searle 2009). Il linguaggio, cioè, può attribuire agli oggetti funzioni nuove che Searle chiama Funzioni di *Status*. Così, pertanto, si esprime:

> Il linguaggio certamente è un'istituzione nel senso banale in cui soddisfa la nostra idea intuitiva di istituzione: è collettivamente accettato, implica convenzioni, è diverso da cultura a cultura e non è dato direttamente dalla nostra biologia. Inoltre, come tutte le istituzioni, è pieno di poteri deontici: diritti, doveri, responsabilità. Gli atti linguistici portano con sé molti poteri deontici. Tuttavia ritengo che il linguaggio sia diverso dalle altre istituzioni, e questo in un senso fondamentale, perché se da un lato i poteri semantici del linguaggio non vanno oltre la semantica, dall'altro noi *usiamo* quei poteri semantici per creare poteri *non* semantici: per creare governi, proprietà privata, università. In tutti questi casi abbiamo poteri non semantici che sono stati creati dal linguaggio. Quindi penso che il linguaggio

> sia l'istituzione sociale *di base*: crea tutte le altre, anche se non è a sua volta creato dallo stesso meccanismo. […] Quello che è interessante, per capire il ruolo dell'attribuzione di funzioni nella costruzione della realtà sociale, è che il linguaggio ci permette di imporre funzioni che vanno oltre la struttura fisica degli oggetti interessati, le Funzioni di *Status* appunto. La realtà sociale è interamente costituita da *status* che rendono possibile che qualcosa o qualcuno svolga una certa funzione. (Searle 2009, 109-110)

Per Searle la realtà sociale, dunque, viene interamente creata dalle Funzioni di *Status* che il linguaggio attribuisce alle cose. Il linguaggio ha quindi il potere di creare la realtà sociale nella dinamica di negoziazione dei significati e dell'attribuzione di *status* agli oggetti, è una prassi condivisa che stabilisce collettivamente nuovi significati attribuendo funzioni nuove a oggetti già semanticamente determinati.

Senza approfondire ulteriormente la prospettiva di Searle, si potrebbe notare come il linguaggio assuma per il filosofo una funzione definibile come meta-istituzionale perché, come per Apel, rappresenta la condizione stessa della fondazione di ogni istituzione. È, dunque, l'istituzione delle istituzioni. Quel che manca alla prospettiva di Searle è, invece, una riflessione più radicale sulla natura del linguaggio inteso come prassi che istituisce gli oggetti stessi e che stabilisce la significatività non a partire dalla semantica referenziale, ma dalla prassi collettiva, dal lavoro entro il quale insieme alla significatività si consegna anche la realtà sociale (Sini 2021). Il linguaggio inteso come prassi, invero, può essere considerato in senso radicale come meta-istituzione perché sovrintende e genera continuamente significati sociali e, dunque, costituisce il supporto dinamico di ogni cultura che, nel sapere, nomina la realtà umana. Il linguaggio, infatti, come istituzione è innanzitutto prassi consolidata, ma in quanto prassi è sempre soggetto alla trasformazione e, pertanto, al rinnovamento di sé e dei propri significati.

3. Il linguaggio come prassi sociale trasformativa: una prospettiva semiotica

Affrontare la questione della natura della significatività, alla luce della prospettiva finora delineata, impone di uscire dal paradigma corrispondentista che stabilisce la verità di una proposizione in base alla corrispondenza tra l'oggetto e la sua rappresentazione linguistica. La significatività, infatti, non può essere ridotta alla referenzialità perché emerge dall'utilizzo intersoggettivo del linguaggio, cioè dalla prassi discorsiva che si istituisce nella comunità storica della comunicazione.

In questa direzione, alcuni risultati provenienti dalla psicologia dello sviluppo e dall'antropologia cognitiva forniscono un riscontro empirico rilevante alla tesi della genesi prassica e intersoggettiva del significato. In particolare, le ricerche di Michael Tomasello mostrano come l'emergere delle competenze linguistiche sia inseparabile da pratiche di cooperazione intenzionale, attenzione condivisa e coordinamento intersoggettivo, all'interno delle quali il significato non precede l'uso, ma si istituisce progressivamente come esito di attività sociali condivise (Tomasello 2009). Tali evidenze empiriche confermano che il linguaggio non può essere adeguatamente compreso come mero sistema formale o come facoltà isolata,

ma come prassi sociale incarnata che struttura l'accesso al mondo e agli oggetti. [4]

Da un punto di vista filosofico, su questa stessa linea, una prospettiva semiotica offre un modello di descrizione particolarmente efficace della determinazione dei significati tematizzando in modo radicale la natura pragmatica del linguaggio. [5]

Secondo la semiotica pragmatica di Peirce, invero, la significatività non è riducibile a una relazione binaria tra segno e oggetto perché un segno è tale solo entro una relazione triadica che implica anche un terzo elemento, l'interpretante, che media la relazione tra oggetto e segno (CP 2.228). L'interpretante è ciò che rende intelligibile un segno, ossia ciò che permette la conoscenza di quel determinato oggetto attraverso il segno. L'interpretante, però, è strettamente connesso alla componente pragmatica perché, secondo la massima pragmatica, il significato della conoscenza di un oggetto corrisponde alla somma degli effetti concepibilmente pratici riferibili all'oggetto (CP 5.402). Per questo la conclusione della catena semiotica non culmina nella rappresentazione dell'oggetto, ma nell'interpretante ultimo che corrisponde all'abito di risposta, alla prassi concreta determinata dalla relazione al segno. La semiosi infinita, ossia l'interpretazione illimitata di un segno che inanella segni e significati, non termina quindi nell'oggetto, ma nella disposizione ad agire. Carlo Sini, interpretando la prospettiva di Peirce accentuandone la dimensione prassica, sottolinea come la potenza del segno non conduce alla rappresentazione ontologica di un oggetto in sé inesistente, ma alla catena delle interpretazioni in cui gli oggetti prendono corpo come frutto di prassi condivise per cui, così si esprime:

> «Dietro» al segno non c'è un *oggetto*, nel senso della ovvietà naturalistica. Impiantare la scienza semiotica sulla base di questa ovvietà significa fare un passo indietro rispetto a Peirce. Non dobbiamo pensare come se ci fosse un oggetto «in sé», preso in una sua beata estraneità rispetto alle relazioni segniche e alle loro funzioni, e poi un modo (*ground*) in cui esso si rapporta col segno (cioè con un'altra «cosa naturale»). Bisogna partire dall'*interpretante* che vede nel *ground* una possibilità di comportamento (ovvero di «risposta»: si ricordi la «massima pragmatica»). In questo «vedere» (che è poi, in senso più proprio, un «fare», il fare di una struttura culturale definita, il suo modo di porsi in relazione segnica: questo significa in senso profondo l'interpretante) si radica la contemporanea possibilità dell'*oggetto* e del segno o *representamen*, cioè si fonda la loro relazione. (Sini 2007, 55)

Ogni oggetto della conoscenza si consegna, così, entro precise coordinate sociali e culturali che ne stabiliscono la forma e i confini, di per sé transeunti e in cammino entro una prassi trasformativa che coincide con il movimento stesso delle società umane (Sini 2012), le quali di volta in volta ristabiliscono, nella semiosi illimitata, nel discorso pubblico e nelle pratiche condivise nuove configurazioni e dunque altrettante interpretazioni nelle quali gli oggetti si stagliano come segni riconosciuti e accettati socialmente. La realtà, per questa ragione, appare come l'insieme complessivo dell'infinito processo della semiosi; a tal riguardo, così Sini scrive:

[4] Il riferimento a Tomasello non intende fornire una spiegazione empirica esaustiva dell'origine del linguaggio, ma indicare un punto di convergenza tra dati sperimentali e una concezione filosofica del significato come prassi intersoggettiva.

[5] Nello specifico si vorrebbe indicare in questa sede la prospettiva semiotica di Peirce secondo l'interpretazione di Sini (2007) e Fabbrichesi (1993).

> Se il reale dunque è un segno e reali in senso proprio sono i segni, in quanto veicoli del significato in cammino, reale, potremmo anche dire, è il significato sociale, ovvero un certo uso dei segni da parte della comunità o da parte delle menti (le quali [...] altro non sono che abiti di risposta, cioè atti interpretativi). La comunità sociale umana, dunque, non fa che tradurre un universo di fatti in un universo di segni dotati di significato, con l'avvertenza che anche i fatti dai quali si parte (la situazione «storica» di ogni interpretante incarnato, il modo «ermeneutico» del suo esistere) sono a loro volta segni, e cioè prodotti di una precedente attività segnico-interpretativa. (Sini 2007, 32-33)

La significatività del linguaggio si radica, pertanto, nella sua genesi eminentemente pratica, che lo rende per questo l'istituzione per eccellenza, vale a dire quella meta-istituzione che fissa le pratiche sociali in segni condivisi. Il linguaggio è così la prassi più propriamente umana: trasformazione e movimento continuo che nominando e rinominando la realtà genera nuovi segni e, con essi, nuovi oggetti e rinnovate configurazioni di senso, cioè nuove disposizioni ad agire.

Il linguaggio, pertanto, considerato nella sua componente trasformativa e attiva, seguendo Wittgenstein (2009), può essere anche descritto come forma di vita soggetta al mutamento. La sua natura prassica si caratterizza per la capacità di definire e ridefinire la mappa della realtà: un atto sociale che, chiamando in causa gli oggetti, ne istituisce l'esistenza pubblica e ne stabilisce la consistenza intersoggettiva.

Il linguaggio, allora, non può essere ridotto a sistema chiuso di segni né a codice formalizzabile una volta per tutte, perché è una prassi storica. Perciò ogni tentativo di descrizione, formalizzazione o modellizzazione del linguaggio non può che situarsi all'interno di tale prassi, come una sua configurazione derivata e storicamente determinata.

4. Dal linguaggio come prassi sociale trasformativa a prassi sociale regolata: il caso dei *Large Language Models*

Se il significato, come suggerisce Sini (2020), è l'effetto di prassi condivise e storicamente situate, allora anche i modelli formalistici e neurocomputazionali del linguaggio devono essere compresi come pratiche interpretative specifiche del fenomeno linguistico. Essi non costituiscono un accesso privilegiato o neutrale al linguaggio perché sono modelli che si sforzano di mapparne il funzionamento secondo schemi e regole che rispondono a significati unanimemente accettati all'interno di una comunità di pratiche. Il linguaggio, invero, non è un codice da decodificare né un dispositivo cerebrale isolato, ma un agire simbolico situato, mediato dal corpo, dalla tecnica, dal contesto storico e dalla relazione sociale.

Attraverso questa prospettiva può, dunque, essere superata la distinzione tra linguaggio naturale e linguaggio artificiale e riformulata come distinzione tra due forme di prassi: l'una involontaria e dinamicamente evolutiva, l'altra progettata, regolata e orientata alla simulazione funzionale. Non esiste, pertanto, un primato ontologico in senso assoluto del "naturale" sull'"artificiale", ma una differenza strutturale tra una prassi che evolve attraverso l'uso e il conflitto simbolico (il linguaggio naturale) e una prassi tecnico-formale che, per quanto avanzata, rimane vincolata alla riproduzione di pattern e alla computazione statistica. Se questo vale

per la modellizzazione della natura del linguaggio entro il generativismo (Chomsky 2000) e il neurocomputazionismo (Rumelhart & McClelland 1986) tanto più è rilevante se si considera il caso dell'intelligenza artificiale generativa e lo sviluppo dei *Large Language Models*. [6]

Gli attuali modelli di intelligenza artificiale generativa, i *Large Language Models* (LLM), sono in grado di simulare il linguaggio con prestazioni paragonabili a un parlante umano perché si fondano sulla ricombinazione statistica di parole e sintagmi la cui significatività è storicamente stratificata nel linguaggio umano. Si ancorano, pertanto, a una significatività percepibile dal fruitore umano perché rinvia alle conoscenze sedimentate in secoli di sviluppo del linguaggio. La creazione di LLM, per questo, non determina una ridefinizione del linguaggio in chiave algoritmica. L'intelligenza artificiale generativa va piuttosto compresa come una modellizzazione che prende le mosse da una precisa pratica progettuale e funzionale, il cui fondamento resta la significatività che si genera nei linguaggi naturali attraverso le pratiche. [7] Da questa prospettiva gli LLM possono essere intesi come un possibile potenziamento del linguaggio umano poiché simulano il comportamento inferenziale e sono in grado di generare catene estese di associazioni tra parole e sintagmi permettendo non solo di fornire informazioni, ma anche di articolare e sviluppare discorsi attraverso altri discorsi. [8] Gli LLM funzionano, dunque, come strumenti di ricombinazione statistica delle informazioni, che il fruitore umano riconosce come conformi a un sapere socialmente condiviso e riconosciuto. Proprio per questo possono operare come strumenti di approfondimento della conoscenza, in quanto generano configurazioni linguistiche conseguenti a conoscenze precedenti, la cui eventuale novità emerge solo nell'uso e nell'interpretazione umana. In tal senso, gli LLM sono strumenti tecnici che possono potenziare le possibilità del linguaggio umano mettendo in relazione significati non ancora esplorati, verificabili poi nell'esperienza pratica e, potenzialmente, capaci di incidere sulle pratiche sociali.

Occorre sottolineare che questa lettura del carattere non pienamente semantico delle relazioni istituite dagli LLM non è pacifica all'interno del dibattito. Piantadosi e Hill (2022), ad esempio, sostengono che negli LLM il significato emerga dalla rete di relazioni tra stati rappresentazionali interni (*conceptual role semantics*) senza richiedere alcun riferimento esterno e, dunque, senza richiedere quel rapporto tra forma e intento comunicativo che Bender e Koller (2020) ritengono, invece, condizione necessaria del significato linguistico. Anche la ricerca recente in *mechanistic interpretability* mostra come le rappresentazioni interne dei modelli siano spesso polisemiche e sovrapposte – una singola unità può codificare più concetti distinti anziché un significato

[6] Per una panoramica sull'evoluzione e il funzionamento dei principali *Large Language Models*, si veda: Naveed et al. (2023).

[7] Una prospettiva affine, sebbene sviluppata entro un differente quadro teorico, è rintracciabile nelle riflessioni di Luciano Floridi, che interpreta i sistemi di intelligenza artificiale come artefatti informazionali inseriti in pratiche socio-tecniche e normative, piuttosto che come entità cognitive autonome (Floridi 2017; 2022). La presente analisi si distingue tuttavia per l'accento posto sulla dimensione prassica e semiotico-istituzionale del linguaggio.

[8] Sul carattere simulativo – e non semantico in senso forte – dei processi inferenziali negli LLM si vedano Bender et al. (2021) e Mitchell (2019), che mostrano come le prestazioni linguistiche avanzate dipendano da operazioni statistiche su corpora testuali e non implichino comprensione intenzionale o autonoma del significato. La questione, nondimeno, resta controversa: sul versante opposto si veda Piantadosi & Hill 2022 su cui si tornerà. Accanto a questo dibattito, si possono inoltre segnalare tre filoni di ricerca rivolti al funzionamento interno dei modelli: la *mechanistic interpretability* (Bereska & Gavves 2024), le *activation analyses* (Belinkov & Glass 2019) e le tecniche di *explainable AI* (Doshi-Velez & Kim 2017; Rudin 2019), che, tuttavia, restano principalmente su un piano descrittivo-empirico e non toccano la questione filosofica di fondo (cfr. anche Williams et al. 2025).

stabile e autonomo (Elhage et al. 2022). Mollo & Millière (2026), tuttavia, pur muovendo da premesse ugualmente anti-scettiche, confermano – diversamente da Piantadosi & Hill – la necessità di un aggancio referenziale genuino al mondo, che nondimeno fondano su una storia di selezione (addestramento su corpora umani, *fine-tuning* tramite *feedback* umano) che è essa stessa prassi altrui sedimentata.

Anche quest'ultimo argomento, se letto alla luce della tesi qui sostenuta, ne conferma la sostanza. Infatti, anche ammettendo che gli LLM istituiscano relazioni interne complesse tra rappresentazioni – o, con Mollo e Millière, un aggancio referenziale genuino al mondo – resta il fatto che tali relazioni non nascono da una prassi condivisa e storicamente situata, ma da un addestramento su tracce già sedimentate di prassi altrui – la differenza, cioè, tra abitare la semiosi e ricombinarne gli esiti. Si può descrivere questa operazione come sovra-scrittura computazionale (Garau 2026), cioè una scrittura non originaria che amplifica le virtualità statistiche dell'archivio scritto senza accedere al suo statuto semantico. Per questo gli LLM non possono creare, in senso proprio, nuovi linguaggi capaci di rimettere in discussione i framework di addestramento che costituiscono la cornice di significato dei loro esiti perché, come si è sostenuto fin dalle prime battute di questo saggio, i linguaggi sorgono come effetto di prassi simboliche condivise e si istituiscono come forme sociali che cristallizzano pratiche e, con esse, significati, in quanto prassi sociali trasformative. Il caso dei *Large Language Models* consente così di chiarire in modo esemplare la differenza tra il linguaggio come prassi involontaria e dinamicamente evolutiva, che si trasforma storicamente nel conflitto simbolico delle pratiche sociali istituendo nuovi significati e nuovi oggetti, e il linguaggio artificiale come prassi progettata e regolata che, pur potenziando le possibilità di articolazione del discorso, rimane orientata alla simulazione funzionale di significatività già date e strutturalmente vincolata ai framework di senso che ne rendono possibile l'operatività.

Conclusione

I linguaggi artificiali e, in particolare, i *Large Language Models*, possono essere compresi in modo non riduzionistico come pratiche tecniche regolate, fondate sulla significatività già istituita nei linguaggi naturali. In quanto pratiche sono, quindi, destinati a incidere sui modi dell'uso linguistico potenziandone le possibilità combinatorie e trasformando indirettamente le pratiche sociali che attraverso il linguaggio si organizzano. Gli LLM non sono soggetti semantici, ma dispositivi tecnici che, inseriti nelle pratiche umane, contribuiscono a ridefinire il paesaggio simbolico entro cui il linguaggio continua il proprio movimento storico.

Assumere il linguaggio come prassi istituente e meta-istituzione consente così non solo di superare la dicotomia tra natura e artificio, ma di chiarire il senso e i limiti epistemici dell'intelligenza artificiale linguistica come momento interno – e non alternativo – alla storia delle pratiche simboliche umane.

Bibliografia

Apel, K.-O. (1973). *Transformation der Philosophie* (I-II). Suhrkamp.

Apel, K.-O. (2003). La filosofia delle istituzioni di Arnold Gehlen e la metaistituzione del linguaggio. *Discipline Filosofiche*, 13 (1), 93–114.

Belinkov, Y. & Glass, J. (2019). Analysis Methods in Neural Language Processing: A Survey. *Transactions of the Association for Computational Linguistics*, 7, 49–72.

Bender, E.M. & Koller, A. (2020). Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data (5185–5198). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Bender, E.M. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? (610–623). In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Bereska, L. & Gavves, E. (2024). Mechanistic Interpretability for AI Safety – A Review. *arXiv*.

Berwick, R.C. & Chomsky, N. (2016). *Perché solo noi. Linguaggio ed evoluzione*. Trad. it. A. De Palma. Bollati Boringhieri.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Doshi-Velez, F. & Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. *arXiv*.

Elhage, N. et al. (2022). Toy Models of Superposition. *arXiv*.

Fabbrichesi, R. (1993). *Introduzione a Peirce*. Laterza.

Fedorenko, E. et al. (2024). The Language Network as a Natural Kind Within the Broader Landscape of the Human Brain. *Nature Reviews Neuroscience*, 25 (4), 289–312.

Floridi, L. (2017). *La quarta rivoluzione*. Raffaello Cortina.

Floridi, L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*. Raffaello Cortina.

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, 91 (4), 1357–1392.

Friederici, A. D. (2017). *Language in Our Brain: The Origins of a Uniquely Human Capacity*. MIT Press.

Gallese, V. & Lakoff, G. (2005). The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge. *Cognitive Neuropsychology*, 22 (3–4), 455–479.

Garau, N. (2026). Scrittura e significato. Dal gesto definitorio alla sovra-scrittura computazionale. *Vita Pensata*, 34, 59–69.

Gehlen, A. (1955). *Der Mensch. Seine Natur und seine Stellung in der Welt*. Athenäum-Verlag.

Gehlen, A. (1956). *Urmensch und Spätkultur. Philosophische Ergebnisse und Aussagen*. Athenäum-Verlag.

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. Edward Arnold.

Hauser, M.D et al. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298 (5598), 1569–1579.

Hymes, D. (1972). On Communicative Competence (269–285). In J. Pride & J. Holmes (eds), *Sociolinguistics*. Penguin Books.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.

Mitchell, M. (2019). Artificial Intelligence Hits the Barrier of Meaning. *Information*, 10 (2), 51.

Mollo, D. C. & Millière, R. (2026). The Vector Grounding Problem. *Philosophy and the Mind Sciences*, 7 (1).

Naveed, H. et al. (2023). A Comprehensive Overview of Large Language Models. *arXiv*.

Piantadosi, S. T. & Hill, F. (2022). Meaning Without Reference in Large Language Models. *arXiv*.

Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1 (5), 206–215.

Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. MIT Press.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Searle, J. (1995). *The Construction of Social Reality*. Free Press.

Searle, J. (2009). *Coscienza, linguaggio, società*. A cura di U. Perone. Rosenberg & Sellier.

Sini, C. (2007). *Eracle al bivio: semiotica e filosofia*. Bollati Boringhieri.

Sini, C. (2012). Transito verità. In *Opere* (vol. V: Figure dell'enciclopedia filosofica). Jaca Book.

Sini, C. (2020). *Inizio*. Jaca Book.

Sini, C. (2021). *Idioma. La cura del discorso*. Jaca Book.

Tomasello, M. (2009). *Le origini della comunicazione umana*. Raffaello Cortina.

Varela, F. J. et al. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Vygotskij, L. S. (1990). *Pensiero e linguaggio*. A cura di L. Mecacci. Laterza.

Williams, I. et al. (2025). Mechanistic Interpretability Needs Philosophy. *arXiv*.

Wittgenstein, L. (2009). *Ricerche filosofiche*. A cura di M. Trinchero. Einaudi.

Artificial Language and Human Cognition. Towards a New Computational Functionalism?

Francesco Maria Cianciaruso

Graduated from the University of Florence in philosophy of mind and artificial intelligence, he works in the fields of epistemology, human cognition, semantic theories and artificial intelligence. His areas of interest range from analytics to ethical and regulatory issues related to the implementation of AI in society.

francescocia.99@gmail.com

Artificial neural networks show complex and adaptive structures that can generate representations and meaning. They can detect the unknown, build models of the world, and create meaningful contents. However, these features are not limited to AI: they offer a key to *rethinking intelligence*. Starting from these developments, this work advances a hybrid computational functionalism that integrates: 1) a top-down level, grounded in symbolic structures and formal rules, and 2) a bottom-up level, in which distributed learning and dynamic adaptation generate representational content. The concept of *network* is the cornerstone of this proposal: a relational and distributed structure that bridges biology and artefact, embodiment and formalism. In this perspective, semantics is not just bound to a biological body but emerges from the functional dynamics of complex systems, whether natural or artificial. This theoretical proposal transcends long-standing dichotomies (body/form, syntax/semantics, biological/artificial) and sketches a philosophy of mind capable of explaining how intelligence – *in all its forms* – constructs meaning through networks of structural interactions.

— SEMANTICS
— LANGUAGE
— NEURAL NETWORKS
— REPRESENTATIONS
— MEANING

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 59 — 72

Introduction: Minds, Language and Networks

What does it mean to understand a statement, represent a concept, or attribute meaning to an experience? What kind of cognitive structure makes all this possible? Philosophy and cognitive science have always revolved around these questions, ever since Plato's *Cratylus* or Aristotle's reflection on ρημα and ονομα and their relationship with νοηματα. With the emergence of the first intelligent machines and the convergence of axiomatics, intelligence theory, and axiomatic logic, the debate has oscillated between two major explanatory models. On the one hand, there was the symbolic paradigm of *classical cognitivism*, according to which mind operates as a system of syntactic manipulation of symbols. On the other hand, *embodied models* have argued for the eminently bodily, situated, and sensorimotor character of semantics.

Classical cognitivism is the first-born from this debate. Developed in the mid-twentieth century, it assumes that mind operates as a computational system that manipulates symbols according to explicit syntactic rules. Influenced by formal logic and symbolic artificial intelligence, this theoretical approach describes cognition as a series of regulated computations between representational mental states, in which each content has an abstract symbolic form that is autonomous from the body or the environment. Within this framework, *meaning* derives from the functional role of the *symbol* within a logical-syntactic system (Fodor 1975; Haugeland 1989). Linguistic competence, according to this view, is determined by generative and recursive rules, and can be modelled independently of perception or action.

In stark contrast, *embodied theories* inspired profoundly different cognitive models. Mind is not disjointed from the body, but radically embodied and situated. Concepts and mental representations are not abstract symbols, but neural simulations based on perceptual-motor data (Barsalou 1999, 2004, 2012). Semantics cannot be understood without reference to bodily experience, and language itself is the extension of a sensorimotor system interacting with the world (Gallese & Lakoff 2005). Conceptual metaphors, abstractions and even inferences would therefore be based on shared bodily schemas, which provide the experiential framework for cognition and language (Gibbs 2004, 2006; Desai 2011, 2022).

Both models have given fundamental contributions, but neither has proven theoretically sufficient. Classical cognitivism, while providing a rigorous logical-mathematical framework, fails to explain how the mind connects abstract symbols to concrete experience. Conversely, embodied theories, while anchoring meaning to a bodily and perceptual dimension, encounter difficulties in accounting for higher-level cognitive processes.

To complicate this description, other perspectives have emerged to overcome the rigid opposition between symbolic representations and sensorimotor systems. Instead of choosing between the symbolic abstraction and embodied simulation, these theories propose a functional and complementary coexistence of multiple representational formats within the cognitive system. One of these is the *Dual Coding Theory*, according to which information is represented in the mind through two distinct but interacting codes: (1) a verbal, abstract, linguistic and symbolic code, and (2) an embodied code, based on perceptual-motor and visual

representations. More recent neurocognitive evidence (Bi 2021; Wang et al. 2020) suggests that these two formats not only coexist but are selectively activated depending on the type of cognitive task and the nature of the content represented. In a pluralistic horizon, the relationship between syntax and semantics is no longer hierarchical or exclusive, but interdependent and dynamic. On the one hand, syntax is recognised as a system of autonomous generative rules, capable of organising language and supporting abstract processing. Semantics, however, does not depend solely on syntax, but is also constructed through the involvement of sensory, emotional, motor and contextual formats. The result is a model in which syntax provides the structural framework, but semantics is distributed across different formats: some embodied, others linguistic, abstract and disembodied.

Meanwhile, after a *dark age*, artificial intelligence has returned to the centre stage. Deep neural networks (DNNs) trained for multiple tasks have demonstrated surprising capabilities: they are able to recognise and generalize new data, construct abstract representations and learn complex conceptual structures (Johnston & Fusi 2023). Some models even simulate a rudimentary form of epistemic awareness, distinguishing between what they *know* and what exceeds their distribution (Bendale & Boult 2016). These behaviours are not simply the result of expanded syntax: *they suggest an emerging ability of AIs to model the world in a representational way*.

This paper takes these findings seriously and considers them not as engineering anomalies, but as philosophical stimuli. In this sense, contemporary AI does not function here as an external case study added to an already established theory of mind. Rather, it serves as a comparative model that makes visible certain structural features of representation: abstraction, generalisation, relational organisation and adaptation to novelty. These features allow the dialogue between philosophical theories of cognition and computational models to become theoretically productive. The hypothesis guiding the work is that AIs should not only be *explained* but can also help to *explain* cognition (Cichy et al. 2019). In other words, the question is not (only) how AI represents, but:

What does AI tell us about how representations work in the human mind?

The relevance of this question is not limited to AI. If artificial systems force us to distinguish between embodiment, symbolic structure and representational function, they also provide a test case for rethinking the architecture of human cognition itself. The philosophical issue is therefore not whether machines simply reproduce human understanding, but whether their behaviour reveals structural conditions that may also be relevant to natural cognition. To answer this question, we propose a new theoretical model that overcomes the symbolic/embodied dualism and embraces the idea that meaning can emerge from a dynamic network of structured interactions. The arguments that follow take the form of a *hybrid computational functionalism*, capable of bringing together: (1) a top-down level, in which the symbolic and syntactic structure organises higher cognitive functions, and (2) a bottom-up level, in which representations emerge from distributed learning and internal simulations.

The unifying key is the concept of the *network*, not as a vague metaphor, but as an operational and philosophical principle. DNNs, brain

synaptic networks, conceptual language networks – all show that intelligence, rather than being localised, is an emergent property of connections. It is in the relationships between nodes, weights, stimuli and feedback that models of the world are constructed, and meaning is generated.

This paper will be structured as follows: (1) we will firstly show the limits of the classical cognitive paradigms and the explanatory potential of dual coding theories and their neural evidence. Then, (2) AI's philosophical implications will be emphasized. We will focus specifically on DNN's ability to create meaning from the structure and without body. (3) Starting from this evidence, we will propose a revised computational functionalism, at the same time bottom-up and top-down, in which *network* is the key concept.

This work is an attempt to describe the contours of a new theoretical framework. It is not a question of establishing a theory, but suggesting an intermediate structure in which embodiment, symbolism and distributed learning are no longer antagonistic paradigms, but interconnected functional components. From this perspective, mind – natural or artificial – is not a place of meaning, but a reticular process that continuously constructs it. This work does not merely integrate existing theories: it proposes a change of framework. A *new functionalism* that does not start from the symbol or the body, but from the network.

Syntax, Semantics and Coding Theories

Almost no one other analogy was more successful than the *mind/computer* one. The idea was as simple as effective: mind (and AI) is a formal system that manipulates symbols according to syntactic rules. Linguistic statements were reduced to formal structures, and meaning was defined by the role that a symbol occupied within the system. Inspired by the optimism of the 1960s, Marvin Minsky, one of the pioneers of artificial intelligence, went so far as to say that: «within a generation [...] few compartments of intellect will remain outside the machine's realm — the problems of creating artificial intelligence will be substantially solved» (Minsky 1967). However, the efforts of classical cognitivism to understand what intelligence and language manipulation have only clarified what they are not. This view had the strength of mathematical clarity, but also the fragility of a theory that leaves meaning *suspended in a vacuum*: a system capable of manipulating words without knowing what it is talking about (Searle 1980).

It was precisely on this point that the *embodied turn* took place. If classical cognitivism was guilty of having *disembodied* the mind, embodiment set out to bring it back into the body and, with it, into the world. Following embodied theories, mental representations are not abstract and formal symbols, but neural simulations that reactivate perceptual-motor patterns associated with sensory experience (Barsalou 1999). Understanding a word does not mean manipulating an arbitrary symbol but activating a motor-perceptual content in a sensorimotor system. Embodied cognition theories have showed how the very structure of thought (including abstract concepts and metaphors) is rooted in bodily experience (Lakoff & Johnson 1980; Gibbs 2004, 2006; Gallese & Lakoff 2005; Borghi 2014, 2016). Indeed, linguistic metaphors reflect a conceptual

architecture that maps the abstract domain onto a concrete, sensory and physical one. Experimental evidence has supported this framework. With fMRI studies, Desai et al. (2022) showed that understanding action metaphors (for example, *grasping an idea*) activates the same motor areas involved in physical action, even when there is no actual movement. [1] These findings reinforce the hypothesis that meaning is a neural extension of perception and action.

However, this theory has a catch. Despite solid empirical support, embodiment does not provide a complete explanation of linguistic cognition. It shows difficulties in dealing with high-level processes such as the manipulation of purely formal concepts. Furthermore, not all mental representations seem to have a perceptual-motor correlate. Some would be non-sensory and require autonomous symbolic processing. Neuroscientific evidence confirms this asymmetry. Indeed, a classic study (Mahon & Caramazza 2008) showed that patients with lesions in motor areas can maintain intact comprehension of action verbs and abstract concepts, suggesting that bodily simulation is not always necessary. Other studies (Moro 2004, 2013) seem to demonstrate that syntax behaves as an independent mechanism, with its own neural specificity. It has been shown that language cortex was selectively activated only when subjects processed possible grammatical structures, while remaining silent with *unnatural* ones (Moro 2004). Following this theoretical framework, syntax, and therefore the symbolic and formal dimension of thought, is not a product of meaning, nor a reflection of the body or perception, but an organisational principle, rooted in human neural architecture.

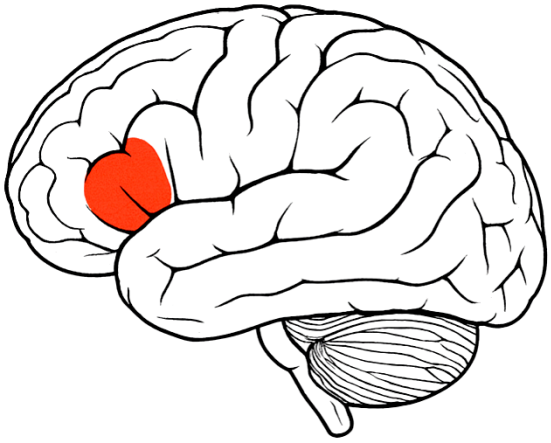
[1] In Desai (2022), a wide range of metaphorical expressions (MET; for example *The man grasped the idea*) are examined with abstract or non-sensory motor control phrases (ABS; for example *He had a bad day*). Literal action expressions (LIT; for example *The man grabbed the handle*) are also used in the research. The prediction is that sensorimotor areas should be activated by METs, and to a greater extent than ABS. In the case of metaphors involving, for example, concepts of action, neuroimaging studies would confirm the prediction in the superior motor areas associated with complex action planning and hand-object interaction in the left anterior inferior parietal lobe.

To fill this gap, pluralistic frameworks have emerged. Dove (2009) proposes a model in which mind uses a plurality of representational formats — sensory, linguistic, symbolic, emotional — each with specific roles. In this perspective, semantics is the product of the interaction between heterogeneous formats, and syntax retains functional autonomy as a mechanism for combining and organising content. More recent research has provided refined empirical evidence. Wang et al. (2020) identified two distinct forms of knowledge representation in the human brain: one sensory, anchored in perceptual areas; the other non-sensory, located in linguistic and associative areas. This dual pathway explains how we can process concrete and abstract concepts using partially separate circuits. On the other hand, Bi (2021) proposes a dual neural coding model, in which the brain represents knowledge through two parallel and complementary systems: a sensory code, rooted in the perceptual-motor areas and activated mainly in the processing of concrete concepts, and a non-sensory, amodal and linguistic code, which supports the understanding of abstract concepts or those without a direct experiential correlate. Neuroimaging data have shown that, during conceptual processing, the brain activates distinct areas and patterns: from sensory areas for perceptive concepts to associative and linguistic areas for abstract concepts. Some studies (Striem-Amit et al. 2018; Bottini 2020; Wang 2020) reveal a functional division in the anterior temporal lobule (ATL): its dorsal portion responds mainly to imperceptible concepts with object referents, while the lateral portion

is more active for abstract concepts without any referents. Conversely, the medial portion of the ATL shows a preference for perceptual concepts. This organisation suggests that the ATL does not represent meaning uniformly, but modulates its activity based on two key dimensions: sensory perceptibility and object referentiality. In other words, distinct parts of the ATL specialise in representing concepts that range from concrete and perceptual to disembodied but structured. [FIG. 1] These findings are echoed by Peelen and Caramazza (2012), who provided fMRI evidence that the ventral anterior temporal lobule encodes not only perceptual features but also abstract conceptual properties of objects. Their work reinforces the view that semantic representation in the brain is not grounded exclusively in sensory input but rather relies on internal structural properties that transcend modality — precisely the kind of representational flexibility exhibited by neural networks trained across multiple tasks.

These studies offer a fundamental contribution to the debate on the origins of semantics: even in the absence of direct sensory experience, the brain develops conceptual representations based on structured and functional foundations. This reinforces the idea that meaning does not depend exclusively on bodily experience, but may emerge in a partially disembodied way, subject to precise and layered neural organisational principles.

This evidence seems to indicate that the two formats not only exist but operate in relative functional independence. Bi (2021) also emphasises that encoding cannot be reduced to single modalities, but a multimodal

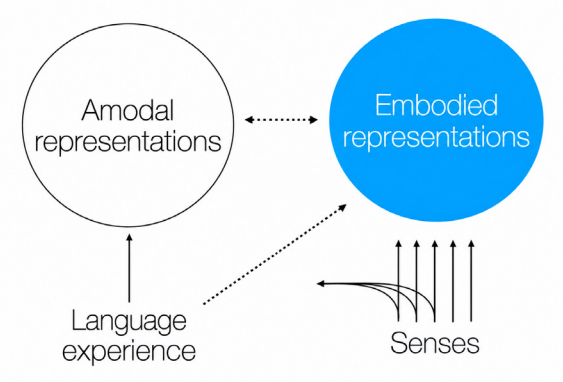


[FIG. 1] Representation of dATL. Studies such as Striem-Amit et al. (2018) indicate that: (1) the dorsal ATL responds to imperceptible concepts with an object referent, (2) the lateral portion responds to abstract concepts without a referent, (3) the medial portion responds to concepts that are perceptible to the senses. Bi (2021) suggests that the dATL is activated by perceptual stimuli (for example, colour) in sensory-deprived subjects, asserting that linguistic-symbolic areas play a primary role in representational processes, alongside sensorimotor areas.

convergence that allows the brain to construct flexible and transferable representations even in the absence of direct sensory stimuli.

The theoretical implications of these evidence are significant. The embodied model, taken alone, would not explain the complexity of cognitive processes. At the same time, purely symbolic theories fail to capture the power rooted in bodily experience. This proposal integrates these two extremes, showing that semantics does not belong exclusively to the body or the abstract mind, but emerges from the interaction of two complementary systems that coexist in the human brain. In this sense, the dual coding model provides a unified yet pluralistic framework capable of explaining the flexibility and complexity of human cognition. [FIG.2]

The path from the failure of classical cognitivism to the emergence of embodiment, and then to representational pluralism, traces a clear trajectory: no single theory seems to capture the entire spectrum of cognitive functions. The challenge today is to articulate a framework in which autonomous syntactic structure coexists with embodied semantic processes and purely abstract forms of representation. It is in this fertile and still largely unexplored theoretical space that the proposal for a *hybrid computational functionalism* fits in, capable of explaining the plurality of formats without sacrificing the unity of the cognitive system. This transition is crucial for the following analysis of neural networks. Dual coding and representational pluralism already show that semantic cognition cannot be reduced to a single format. DNNs radicalise this problem. Indeed,



[FIG. 2] Schematic representation of the interaction between symbolic and perceptual formats in a hybrid model.

they display forms of abstraction, generalisation and internal organisation in systems that are not biologically embodied. For this reason, they do not replace neurocognitive evidence but extend the philosophical question concerning the structural conditions under which representations become functionally meaningful.

This opens the theoretical space we want to explore: a paradigm that conceives the mind as a dynamic network, in which form and function, symbol and body, cease to be in opposition.

Meaning and Structure: The Case of DNNs

If artificial networks can produce coherent representations, answer novel questions, infer patterns, translate expressions and generalise across contexts, then at least some functional properties associated with semantic cognition may not be exclusive to embodied experience. This does not mean that DNNs possess meaning in the same way human subjects do. Rather, their behaviour suggests that semantic-like organisation can emerge from structured relations among internal states, training constraints and task demands. The evidence that artificial models can simulate semantically meaningful behaviour raises a crucial question: *is it really the body that establishes meaning, or is it the structure that generates it?* In this chapter, we will not discuss AI as a simple computational tool. We will treat it as a philosophical lens, a cognitive probe that forces to recognise that perhaps the mind is not flesh or machine. But *network*. If DNNs can construct representations without a body, then semantics can emerge from *structure*. This is not just an engineering achievement: it is a philosophical clue.

Neural networks developed for multiple recognition tasks have shown the emergence of artificial representations (Yang et al. 2019; Johnston & Fusi 2023). Johnston and Fusi (2023) have demonstrated this empirically: by simply learning to manage multiple tasks, DNNs develop abstract representations, a type of distributed coding that allows the network to *generalise* to new situations based on very few examples. These models capture general characteristics of the data through layers of linear/non-linear functions that extract hierarchical patterns from the raw data. In this context, abstract representations would allow the production of potentially new examples of existing classes. These representations would therefore be crucial for effective generalisation and extraction of task-relevant features. They correspond to a functional geometry of mental space, in which latent variables (such as colour, shape, and functionality of the object) emerge in a linear and independent manner. This abstractness is not design but emerges spontaneously from the multiplication of behavioural constraints (Johnston & Fusi 2023). This internal, almost topological geometry is *structure* rather than *content*. And if the network can construct abstract and operational representations without bodily constraints, then maybe semantics does not reside solely in embodied experience. It would be an emergent property of complex structures, capable of articulating contexts, tasks and relationships autonomously. In these networks, meaning is not *imported* but constructed by structural convergence. The emergence of abstract representations in a neural network does not by itself demonstrate the presence of understanding, intentionality or phenomenological

experience. However, it shows that some functional properties commonly associated with semantic representation (for example, abstraction, generalisation, context-sensitive discrimination and transfer across tasks) can arise from distributed relational dynamics without being explicitly encoded as symbolic rules. The point is therefore not that DNNs possess human-like semantics, but that they weaken the claim that semantic organisation must depend exclusively on sensorimotor embodiment.

Among the most intriguing developments in the current wave of neural network-based artificial intelligence are DNNs designed for visual recognition. Their applications range from autonomous driving to embodied robotics (Sünderhauf 2018), from biometric identification to computational models of the visual cortex within cognitive neuroscience (Güçlü 2015; Yamins & DiCarlo 2016). At the core of many of these systems lies the *Softmax* activation function, a widely adopted method for producing probability distributions over a set of known classes. [2] Its appeal lies in its simplicity: it transforms raw activations into an interpretable ranking of class probabilities, enabling classification tasks across diverse domains. However, the world is not a closed set. Systems routinely encounter out-of-distribution (OOD) data, inputs that deviate from any class the model has seen during training (Bendale & Boult 2016). Here, SoftMax's rigid architecture shows its limits: it cannot discern when its prediction is wrong. To address this limitation, *OpenMax* was developed: an activation function built for open-world recognition. Unlike Softmax, which always forces a decision among known categories, OpenMax incorporates a *meta-recognition* framework that enables the model to assess the confidence of its own classifications (Bendale & Boult 2016). [3] It does so by analysing the internal activation patterns for each known class and constructing a probability distribution. This structure allows the model to detect when a stimulus may belong to an unknown class. Using Weibull distribution to estimate these confidence margins, OpenMax evaluates the distance between a new input and the learned activation distributions. The further the distance, the more likely the stimulus represents a semantic outlier — a class the system has never encountered. In this way, the network begins to trace the boundaries of its own knowledge. From a cognitive perspective, this is no small feat. The system is not just classifying data but developing a primitive epistemic reflexivity. It distinguishes between what it knows and what it does not know and adjusts its behaviour accordingly. Open-set recognition thus mirrors a foundational capacity of human cognition: the ability to operate in uncertain, dynamic environments, where the categories are not fixed, and the unexpected is inevitable. This ability to detect and respond to the unknown (and to draw epistemic boundaries without reliance on prior embodied experience) offers strong support for the theoretical model advanced in this work. If a network can construct, adjust, and defend its own representational architecture in the face of novelty, then the foundations of cognition must lie not merely in sensory grounding, but in the dynamical architecture of relational structures.

[2] *Softmax* function is defined as:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_i}}$$

where z_i is the value of the network parameters for class i, e^{z_i} is their exponential, and $\sum_{j=i}^{n} e^{z_i}$ is the sum of the exponentials of all parameters. In this way, the activation function acts on the network parameters, transforming them into a probability distribution, thus allowing the neural architecture to perform the classification task correctly.

[3] *OpenMax* is defined as:

$$\hat{z}_i = z_i \bullet \mathrm{Weibull}(z_i; l_i, k_i)$$

where $\mathrm{Weibull}(z_i; l_i, k_i)$ is the probability that the parameter values z_i belong to the distribution model generated by the network, thus allowing us to calculate whether the data is an anomaly with respect to the emerging model in the system.

What these findings exemplifies is a mode of intelligence that does not simulate embodiment but reconstructs meaning through structural reflexivity. In this sense, open-set models reinforce the core premise of a new computational functionalism: cognition is not the product of fixed semantic contents, nor of sensorimotor processes alone, but of systems capable of orienting themselves within a fluid, open-ended space. Semantics, here, is not given, but is inferred, negotiated, and continually revised. This is not a weaker form of intelligence: it is rather a different grammar of meaning born from structure.

A New Computational Functionalism?

The discussion so far has highlighted a tension that runs through much of contemporary research on the mind and language: on the one hand, the need for a perceptual and bodily anchor in the construction of meaning; on the other, the evidence of autonomous cognitive structures capable of operating on concepts, rules and abstract categories. Embodied theories have brought the role of the body and sensorimotor experience back to the core of the debate (Barsalou 1999; Lakoff & Johnson 2005; Borghi 2014; Desai 2022), while symbolic and recursive approaches have continued to demonstrate the power of grammatical architectures and syntactic manipulation (Moro 2004).

However, when viewed in the light of the latest neuroscientific and computational discoveries, these positions do not seem to contradict each other but rather describe *different components of the same cognitive process*. Clinical and fMRI evidence shows that the brain hosts several distributed and specialised representational formats that coexist and activate depending on the task (Striem-Amit et al. 2018; Wang 2020; Bi 2021). Similarly, DNNs (from GPT to OpenMax) demonstrate that artificial systems can construct abstract representations, even in the absence of sensory anchoring (Johnston 2023).

This heterogeneity highlights a crucial point: no single theory seems sufficient to explain the entire spectrum of *meaning*. Embodied theories work well for perceptual and concrete concepts but struggle to account for the semantic flexibility of abstract or symbolic concepts. Symbolic models, in turn, account for grammatical structure but ignore the dynamics of situated learning. [FIG.3]

THEORY	KEY CONCEPTS	EXPECTED ACTIVATIONS
Embodied	Meaning arises from the activation of perceptual and motor circuits	Premotor and motor cortex, superior parietal cortex
Formalist	Meaning is the manipulation of symbols and syntax	Language network: Inferior Frontal Gyrus (IFG / Broca's area), posterior Middle Temporal Gyrus (pMTG), dATL
Dual Coding	Two systems coexist: one verbal/symbolic and one imaginative/perceptual	dATL + IFG with sensorimotor areas

[FIG. 3] Summary of the different theorical frameworks that focus on meaning emergence and representational formats.

What we propose is not an *eclectic* mediation, nor a simple sum of pre-existing approaches. Rather, it is the hypothesis that meaning emerges from the complexity of the system itself, from the relationship between heterogeneous representational formats that interact and organise themselves into an architecture dense enough to generate internal coherence.

Considering the evidence provided, what is needed is *not a new form of dualism* or a theory that simply distinguishes between two codes of representation. Faced with growing empirical convergence between neuroscience, computational linguistics, and theory of mind, there is a need for a *new theoretical framework*, capable not only of recognising the plurality of representational forms, but also of explaining their coherent and generative interaction.

From this perspective, what can serve as a unifying principle is a functional structure: a *network*. Not as a simple neural architecture, but as an organisational form, a relational topology capable of hosting multiple dynamics and different levels of abstraction. The network, understood as a self-organising internal structure, becomes the point of convergence between embodied learning, symbolic representation and computational abstraction.

— Hybrid Hypothesis (HH): Meaning emerges from sufficiently complex structure. [4] The components and representational formats constitute different and complementary neural strategies of the intelligent systems to build meaning from the structure.

[4] By *sufficiently complex structure*, I do not mean complexity in a merely quantitative sense, such as the number of nodes, parameters or connections. I refer instead to an organised relational architecture capable of integrating heterogeneous representational formats, stabilising internal differences, generalising across contexts and revising its states in response to novelty. Complexity is therefore relevant only when it becomes functionally organised.

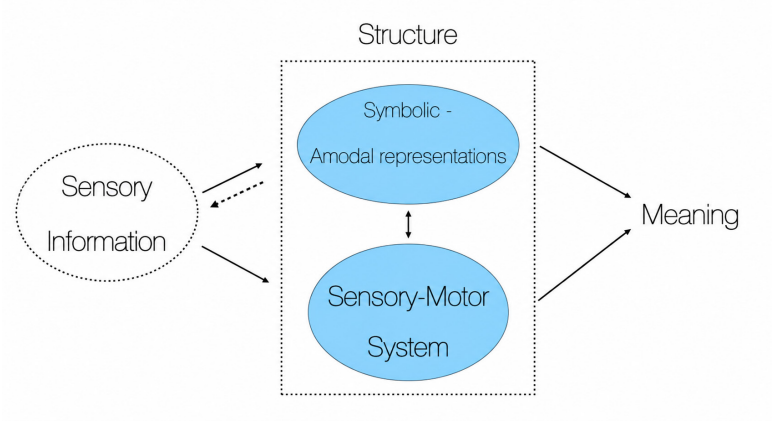
Within this argument, DNNs represent a theoretical basis and a philosophical lens for supporting the idea that meaning emerges from sufficiently complex structures, and that this feature may be common among different intelligent systems. The ability of neural models to produce abstract representations and generalise new data through statistical *models of the world*, allows us to assert the emergence of meaning within their computational units, although through formats and processes that differ from those required in biological systems.

The idea here proposed is that it is the structure and complexity of the model that allows meaning to emerge. It is precisely the concept of structure and network that could allow different formats, representations and neural processes to be brought together in a single coherent and hybrid theoretical framework. [FIG.4]

Through the philosophical concepts of *network* and *structure*, it would thus be possible to bring together models and experimental data from different research perspectives into a single coherent theoretical framework, providing a consistent explanation of the emergence of meaning from complex systems in one unified model.

Concluding Remarks

In this paper, we explored the theoretical possibility of reconciling different perspectives and cognitive models that explain the emergence of



[FIG. 4] Schematic Representation of HH

meaning and linguistic processing within a single, coherent and unified theoretical framework. Evidence from cognitive science seems, on the one hand, to confirm the idea that the sensorimotor system plays a central role in the encoding of concepts and language. On the other hand, the role of syntactic and symbolic formats seems inevitable, especially about high-level processes. Numerous theories have emerged to fill the gap between these different perspectives. What we have claimed is that the concepts of structures and networks would justify this convergence. These concepts are borrowed from current knowledge on artificial intelligence and deep learning. Artificial models seem capable of producing abstract representations and models of the world. This *semantic* competence would seem to suggest a central role for the structure of the system itself, whose complexity determines the manipulation of meaning and linguistic competence.

Our theoretical proposal is intended as a suggestion for research, a new solution to the problem of meaning that requires experimental support. Only new empirical evidence could confirm or refute this convergence, but research continues.

Bibliography

Barsalou, L.W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22 (4), 577–660.

Barsalou, L.W. (2012). The Human Conceptual System. *The Cambridge Handbook of Psycholinguistics*, 239–258.

Barsalou, L. W. & Solomon, K. O. (2004). Perceptual Simulation in Property Verification. *Memory & Cognition*, 32 (2), 244–259.

Bendale, A. & Boult, T. E. (2016). Towards Open Set Deep Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1563–1572).

Bi, Y. (2021). Dual Coding of Knowledge in the Human Brain. *Trends in Cognitive Sciences*, 25 (10), 883–895.

Borghi, A.M. & Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. Springer.

Bottini, R. et al. (2020). Brain Regions Involved in Conceptual Retrieval in Sighted and Blind People. *Journal of Cognitive Neuroscience*, 32, 1009–1025.

Caruana, F. & Borghi, A. (2016). *Il Cervello in Azione: Introduzione alle Nuove Scienze della Mente*. il Mulino.

Cichy, R.M. & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences*, 23 (4), 305–317.

Desai, R.H. (2022). Are Metaphors Embodied? The Neural Evidence. *Psychological Research*, 86 (8), 2417–2433.

Desai, R.H. et al. (2011). The Neural Career of Sensory-Motor Metaphors. *Journal of Cognitive Neuroscience*, 23 (9), 2376–2386.

Dove, G. (2009). Beyond Perceptual Symbols: A Call for Representational Pluralism. *Cognition*, 110 (3), 412–431.

Fang, Y. et al. (2018). Semantic Representation in the White Matter Pathway. *PLoS Biology*, 16, e2003993.

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Gallese, V. & Lakoff, G. (2005). The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge. *Cognitive Neuropsychology*, 22 (3–4), 455–479.

Gibbs Jr, R.W. (2006). Metaphor Interpretation as Embodied Simulation. *Mind & Language*, 21 (3), 434-458.

Gibbs Jr, R.W., Lima, P. L. C. & Francozo, E. (2004). Metaphor Is Grounded in Embodied Experience. *Journal of Pragmatics*, 36 (7), 1189–1210.

Güçlü, U. & Van Gerven, M.A. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations Across the Ventral Stream. *Journal of Neuroscience*, 35 (27), 10005–10014.

Haugeland, J. (1989). *Artificial Intelligence: The Very Idea*. MIT Press.

Johnston, W.J. & Fusi, S. (2023). Abstract Representations Emerge Naturally in Neural Networks Trained to Perform Multiple Tasks. *Nature Communications*, 14 (1), 1040.

Lakoff, G. & Johnson, M. (1980). The Metaphorical Structure of the Human Conceptual System. *Cognitive Science*, 4 (2), 195–208.

Lambon Ralph, M.A. et al. (2016). The Neural and Computational Bases of Semantic Cognition. *Nature Reviews Neuroscience*, 18, 42–55.

Mahon, B.Z. & Caramazza, A. (2008). A Critical Look at the Embodied Cognition Hypothesis and a New Proposal for Grounding Conceptual Content. *Journal of Physiology-Paris*, 102 (1–3), 59–70.

Minsky, M.L. (1967). *Computation: Finite and Infinite Machines*. Prentice-Hall.

Moro, A. (2004). Autonomia della Sintassi e Tecniche di Neuroimmagine. *Lingue e Linguaggio*, 3 (1), 133–170.

Moro, A. et al. (2013). Syntax Without Language: Neurobiological Evidence for Cross-Domain Syntactic Computations. In *The Equilibrium of Human Syntax* (317–341). Routledge.

Peelen, M.V. & Caramazza, A. (2012). Conceptual Object Representations in Human Anterior Temporal Cortex. *Journal of Neuroscience*, 32, 15728–15736.

Searle, J.R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3 (3), 417–424.

Striem-Amit, E. et al. (2018). Neural Representation of Visual Concepts in People Born Blind. *Nature Communications*, 9, 5250.

Sünderhauf, N. et al. (2018). The Limits and Potentials of Deep Learning for Robotics. *The International Journal of Robotics Research*, 37 (4–5), 405–420.

Wang, J. et al. (2010). Neural Representation of Abstract and Concrete Concepts: A Meta-Analysis of Neuroimaging Studies. *Human Brain Mapping*, 31 (10), 1459–1468.

Wang, X. et al. (2020). Two Forms of Knowledge Representations in the Human Brain. *Neuron*, 107 (2), 383–393.

Yamins, D.L. & DiCarlo, J.J. (2016). Using Goal-Driven Deep Learning Models to Understand Sensory Cortex. *Nature Neuroscience*, 19 (3), 356–365.

Yang, G.R. et al. (2019). Task Representations in Neural Networks Trained to Perform Many Cognitive Tasks. *Nature Neuroscience*, 22, 297–306.

San Francesco e l'IA. Dalla zoosemiotica all'intelligenza artificiale relazionale*

Daria Arkhipova

Ha conseguito il dottorato di ricerca presso l'Università di Torino, Italia, e l'Università di Tartu, Estonia. La sua attività di ricerca si concentra sulla significazione dei sistemi di intelligenza artificiale, esplorando come vengono progettati e utilizzati. Partecipa al progetto "San Francesco e l'intelligenza artificiale" dell'Università degli Studi di Torino, e al progetto AITA della Fondazione Bruno Kessler.

daria.arkhipova@unito.it

Nicola Zengiaro

Ricercatore post-dottorale Marie Skłodowska-Curie Actions Global (MSCA) presso la Charles University di Praga e l'Universidad EAN di Bogotá. Coordina il progetto VITAE (*Vital Interpretations of Trans-ecologies, Animacy and Epistemologies*), indagando empiricamente come il concetto di "vita" venga costruito e protetto nelle politiche ambientali globali. È Associate Editor della rivista *Biosemiotics* e Managing Editor di *Linguistic Frontiers* e *Animal Studies*.

nicola.zengiaro@fhs.cuni.cz

This contribution, developed within the framework of the "San Francesco e l'IA" project, critically interrogates the computational approaches currently applied to the study of animal vocalizations, situating this inquiry at the intersection of semiotics, animal communication and artificial intelligence. Drawing on zoosemiotics and biosemiotics, fields that challenge the reduction of animal communication to anthropocentric linguistic categories, the study analyzes recent attempts to "translate" animal vocalizations via LLM-based architectures, including cetacean-focused projects such as CETI, ESP and DolphinGemma. These approaches, we argue, rest on problematic assumptions: they treat animal vocalizations as structured language awaiting decoding, conflicting with biosemiotic accounts of animal communication as contextual and multisensory, rather than primarily informational. These approaches overlook that animal signs are indexical, iconic and pragmatically embedded in ecological and social networks that shape their meaning. The contribution thus exposes the epistemological limitations of "decoding" paradigms and proposes a shift toward interpretive and relational AI, understood not as a translation tool but as a technology of mediation that refines human perception without reducing animal communication to human categories. This reframes AI as an interpretive technology attuned to the semiotic complexity of living systems, advancing an ecological non-reductionist understanding of interspecies interaction within the interdisciplinary debate among semiotics, cognitive ethology and AI studies.

*Questo progetto ha ricevuto un finanziamento dalla Fondazione CRT (LEOM_CRT_25_01 – 110992 / 2024.1420 – *San Francesco e l'IA*, PI: Massimo Leone).

Attribuzione dei contributi: i Capitoli 4 e 5 sono a cura di Daria Arkhipova; i Capitoli 2 e 3 sono a cura di Nicola Zengiaro; introduzione e conclusione sono stati scritti insieme.

— BIOSEMIOTICA
— COMUNICAZIONE ANIMALE
— INTELLIGENZA ARTIFICIALE
— TRADUZIONE

1. Introduzione

Il progetto "San Francesco e l'IA", condotto nel 2025 presso l'Università di Torino, si colloca all'intersezione tra semiotica, comunicazione animale e intelligenza artificiale, proponendo un'indagine transdisciplinare che interroga criticamente le strategie computazionali applicate allo studio delle vocalizzazioni non umane. Il presente contributo ne illustra i presupposti teorici e metodologici che ne hanno guidato lo svolgimento. Il lavoro si articola lungo due direttrici complementari: da un lato, un'analisi storico-teorica della zoosemiotica e della biosemiotica come discipline che hanno problematizzato la riduzione della comunicazione animale a categorie linguistiche antropocentriche; dall'altro, un esame critico dei recenti tentativi di "traduzione" delle vocalizzazioni animali attraverso sistemi di intelligenza artificiale. La ricerca si fonda sulla convinzione che l'impiego dell'intelligenza artificiale nello studio della comunicazione animale, lungi dall'essere un'operazione neutra, implichi precisi presupposti epistemologici e semiotici che necessitano di un'attenta valutazione critica. In particolare, il progetto intende verificare se e in che misura gli approcci computazionali dominanti si basino sull'assimilazione delle vocalizzazioni animali a un linguaggio strutturato, decodificabile e trasferibile in termini umani. Questa ipotesi diffusa nei settori dell'IA e delle scienze cognitive risulta problematica alla luce dei principi della biosemiotica.

L'impiego dell'intelligenza artificiale nello studio della comunicazione animale ha acquisito negli ultimi anni un rilievo senza precedenti, generando un nuovo orizzonte epistemologico in cui si prospetta la possibilità di "tradurre" e "decodificare" i segnali emessi da specie non umane in forme accessibili all'interpretazione umana (Andreas *et al.* 2022; Rutz *et al.* 2023). Progetti ambiziosi come CETI (Cetacean Translation Initiative), ESP (Earth Species Project), DolphinGemma e ISPA (Inter-Species Phonetic Alphabet) propongono di applicare tecniche di Natural Language Processing e Deep Learning alle vocalizzazioni dei cetacei e altri animali, immaginando modelli computazionali capaci di riconoscere e generare strutture linguistiche analoghe a quelle umane (Hagiwara, Miron & Liu 2024; Beguš *et al.* 2025a). Tuttavia, questo entusiasmo tecnologico solleva interrogativi fondamentali di natura semiotica, epistemologica ed etica: è legittimo assumere che la comunicazione animale possa essere modellata secondo le categorie strutturali del linguaggio umano? Quali presupposti ideologici sottendono l'operazione di "traduzione" interspecifica? E quali forme di riduzionismo antropocentrico rischiano di occultare la complessità semiotica dei sistemi viventi? Quali strumenti di intelligenza artificiale possono rivelarsi efficaci per integrare gli approcci biosemiotici e zoosemiotici nella traduzione della comunicazione animale?

Il presente contributo si inserisce in questo dibattito critico attraverso un'indagine transdisciplinare che incrocia semiotica, biosemiotica, etologia cognitiva e studi sull'intelligenza artificiale. I primi risultati dell'indagine evidenziano come gli approcci computazionali dominanti si basino su un'assimilazione problematica delle vocalizzazioni animali a un linguaggio strutturato, decodificabile e trasferibile in termini umani. Un'ipotesi che, sebbene ricorrente nella storia delle scienze cognitive, contrasta con i principi fondamentali della biosemiotica. Secondo quest'ultima prospettiva, la comunicazione animale costituisce un processo relazionale,

ecologicamente contestuale e multisensoriale, in cui i segni prodotti dagli animali si configurano spesso come indicali e iconici, inscritti in una rete ecologica e sociale che ne determina il senso. La funzione comunicativa, in questa cornice teorica, non è necessariamente finalizzata alla trasmissione di contenuti informativi discreti, ma può assumere carattere performativo, affettivo, identitario o relazionale. Il caso dei capodogli risulta particolarmente illuminante, poiché le loro vocalizzazioni e *codas* non costituiscono una mera forma di vocalizzazione da tradurre, ma una prassi comunicativa immersa in dinamiche di riconoscimento identitario, coesione clanica, costituzione dell'ecologia ambientale e codificazione culturale, all'interno di un sistema sociale cooperativo e intergenerazionale.

Questo articolo si propone di esaminare l'importanza delle capacità dell'intelligenza artificiale nell'ambito dei fondamenti zoosemiotici e biosemiotici, analizzando il quadro teorico relativo alla traduzione e all'interpretazione della comunicazione animale. Si approfondisce la sfida della "traduzione", esplorando in che modo un approccio semiotico possa contribuire a una comprensione più articolata dei segnali animali. Il lavoro prosegue esplorando le potenzialità dell'intelligenza artificiale applicata alla comunicazione tra esseri umani e animali, con un'analisi sull'uso attuale dei Large Language Models (LLM) e sulle modalità con cui tali strumenti possano essere migliorati per essere integrati nei sistemi AI, con l'obiettivo di decodificare e tradurre in modo più competente i segnali comunicativi interspecie.

2. Fondamenti Zoosemiotici e Biosemiotici

Il framework teorico della zoosemiotica si costituisce storicamente come una rottura epistemologica fondamentale rispetto alle concezioni meccanicistiche del comportamento animale, proponendo una sintesi originale tra la semiotica e le scienze della vita che trova il suo atto di nascita formale nel 1963. In quell'anno, Thomas A. Sebeok (1963), semiologo ungherese naturalizzato statunitense, all'interno di una recensione seminale pubblicata sulla rivista *Language*, coniò il termine per indicare un campo di indagine dedicato alla comprensione scientifica della comunicazione negli animali non umani, denunciando lo stato "tassonomicamente parrocchiale" (Sebeok 1963, 465) degli studi precedenti, frammentati in descrizioni aneddotiche e privi di una visione comparativa sistematica. La periodizzazione di Maran (2010) individua due fasi principali della zoosemiotica: la prima (1962-1969), che potrebbe però iniziare già nel 1960 con il saggio di Sebeok a Burg Wartenstein, e la seconda che comincia nel 1975 e che, secondo una rilettura di Cobley (2014), andrebbe idealmente estesa fino agli anni Novanta per ricomprendere gli studi sul caso "Clever Hans" (iniziati col convegno del 1980) e le ricerche sul gatto domestico pubblicate nel 1994 e riedite nel 2001. Tali approcci si inseriscono nel cuore della "svolta cognitiva" degli anni Sessanta, trovando momenti di cristallizzazione teorica proprio nei simposi di Burg Wartenstein organizzati dalla Wenner-Gren Foundation — in particolare *Comparative Aspects of Human Communication* (1960) [1] e *Animal Communication* (1965) —, dove studiosi del calibro di Gregory Bateson, Charles Hockett e Peter Marler iniziarono a interrogarsi sulle proprietà distintive del linguaggio

[1] Si veda https://wennergren.org/symposium-seminar/comparative-aspects-of-human-communication/ e https://wennergren.org/symposia-seminars/ (visitato il 10/01/2026).

umano in rapporto ai sistemi di segnalazione animale. La novità radicale introdotta da Sebeok (1998), influenzato dalla logica di Charles S. Peirce e dalla fenomenologia biologica di Jakob von Uexküll, risiede nell'estensione della semiosi — il processo di produzione e interpretazione dei segni — a tutto il mondo vivente, sottraendo il concetto di significato al dominio esclusivo della parola umana e ridefinendo la comunicazione non come un mero scambio meccanico di segnali, ma come un'attività interpretativa complessa e situata nel mondo della vita.

Il pilastro filosofico su cui poggia gran parte degli approcci iniziali della disciplina è il concetto di Umwelt, elaborato da von Uexküll (1921 [1909]) all'inizio del XX secolo, il quale postula che ogni organismo abiti un mondo soggettivo modellato dalle proprie specifiche capacità percettive (*Merkwelt*) e operative (*Wirkwelt*) in relazione all'ambiente circostante. In questa prospettiva, la semiosi animale è intrinsecamente legata alla biologia del soggetto: un segnale non è un dato astratto, ma un elemento che assume valore di segno solo all'interno della relazione dinamica tra l'organismo e il suo ambiente (Deely 2001). Tale impostazione permette di superare il dualismo natura-cultura, suggerendo che l'evoluzione biologica sia anche un'evoluzione della capacità di interpretare segni in modo sempre più complesso e meno deterministico (Sebeok 1997, 436), un concetto che il biosemiologo danese Jesper Hoffmeyer (1996, 61) avrebbe più tardi definito come "libertà semiotica" (Hoffmeyer 2008, 185). La zoosemiotica non cerca dunque negli animali una grammatica speculare a quella umana, ma indaga le modalità con cui gli organismi costruiscono modelli della realtà attraverso codici che, a differenza dell'arbitrarietà simbolica del linguaggio verbale, sono prevalentemente iconici e indicali, mantenendo una connessione fisica o di somiglianza con il proprio referente.

L'architettura metodologica della disciplina si è strutturata attorno alla tripartizione proposta da Charles Morris (1946): sintattica, semantica e pragmatica, adattata con rigore alle specificità del mondo non umano (Martinelli 2010, 66-68). La zoosintattica si occupa dello studio della struttura dei segnali e delle regole di combinazione, affrontando la sfida di identificare unità discrete all'interno di flussi comunicativi spesso continui e multimodali, come i canti dei cetacei o le danze delle api. La zoosemantica indaga il rapporto tra i segni e i loro referenti, analizzando come i segnali veicolino informazioni su stati interni, elementi ambientali o identità sociali. Infine, la zoopragmatica esamina le condizioni di produzione e gli effetti dei segni sul comportamento, radicando la semiosi in una dimensione sociale e relazionale. Questo apparato metodologico (Sebeok 1990) ha permesso di affrontare uno dei dibattiti più accesi della disciplina, ossia quello della contrapposizione tra il modello informazionale e il modello manipolativo o dell'influenza. Se l'etologia classica di Lorenz e Tinbergen concepiva la comunicazione come un trasferimento di informazioni finalizzato al beneficio della specie, la svolta impressa da Richard Dawkins e John Krebs negli anni Settanta ha introdotto una visione più strategica, secondo cui i segnali sono strumenti selezionati evolutivamente per influenzare o "manipolare" il comportamento del ricevente a vantaggio dell'emittente (Davies, Krebs & West 1981). In termini generali Sebeok (2001, 156) ha spesso argomentato che la semiosi è la capacità di una specie di produrre e comprendere i tipi specifici di modelli di cui ha bisogno per elaborare e codificare gli stimoli percettivi a modo suo. La zoosemiotica

contemporanea integra queste visioni sostenendo un modello di comunicazione-come-influenza che risuona con le teorie pragmatiche umane, dove l'atto comunicativo mira a modificare la rappresentazione del mondo del destinatario, rendendo la semiosi un'arena di negoziazione costante (Maran, Martinelli & Turovski 2011; Kull 2014).

Un elemento di ulteriore complessità è rappresentato dalla multimodalità costitutiva della comunicazione animale, la quale sfrutta diversi canali sensoriali, come quello acustico, visivo, chimico, tattile, elettrico e vibrazionale, ciascuno vincolato da precise restrizioni fisiche ed ecologiche. La scelta del canale influenza profondamente la struttura del segno: la comunicazione acustica, ad esempio, è soggetta a vincoli di propagazione che variano tra aria e acqua, mentre la comunicazione chimica, basata su feromoni, offre una persistenza temporale che compensa la scarsa direzionalità. La comprensione di come questi canali si integrino in un sistema olistico è fondamentale per evitare riduzionismi tecnologici, specialmente nel contesto attuale in cui l'intelligenza artificiale tende a privilegiare i dati audio e video a scapito della semiosi chimica o tattile (Zengiaro 2026). Lo sviluppo della disciplina ha inoltre portato alla distinzione tra relazioni intraspecifiche, regolate da codici filogeneticamente condivisi, e relazioni interspecifiche, dove la comunicazione si fonda su processi di interpretazione più aperti e situati, come avviene nelle simbiosi o nelle interazioni tra animali domestici e umani (Sebeok 1990, 45). In questo ambito si inserisce l'antropozoosemiotica, che analizza le rappresentazioni culturali e i saperi che l'uomo produce per interpretare l'alterità animale (Zengiaro & Jaramillo 2025), evidenziando come il confine tra le specie sia costantemente mediato da testi, immagini e tecnologie.

Negli sviluppi più recenti, la zoosemiotica è confluita nella biosemiotica e nell'ecosemiotica (nonostante la zoosemiotica, fino al 1981, si identificasse con la "biosemiotica generale"), [2] discipline che estendono l'analisi dei segni a tutti i livelli dell'organizzazione biologica e studiano come le relazioni semiotiche siano modellate dai cambiamenti ambientali e dall'antropizzazione. Sebeok (1990, 108) ha illustrato, ad esempio, come le variabili ambientali possano influenzare la comunicazione, notando che i richiami dei grilli maschi variano in base alla temperatura, richiedendo al sistema interpretativo della femmina di tenere conto dei fluttuanti fattori esterni. Ciò dimostra non solo la flessibilità, ma anche la sofisticatezza evolutiva dei codici animali, la cui funzione non è rigidamente fissata ma risponde a condizioni contingenti. Studi contemporanei sul cambiamento climatico hanno dimostrato che l'acidificazione degli oceani altera il mezzo attraverso il quale le balene trasmettono i segnali, influenzando così la loro comunicazione (Sehgal, Tumar & Schönwälder 2010, 3).

Nonostante ciò, storicamente, la disciplina ha dovuto difendersi dall'accusa di antropomorfismo (Gensini 2018), ma la forza del framework biosemiotico risiede proprio nel riconoscimento che, sebbene gli animali non possiedano il linguaggio verbale, essi partecipano a una semiosi incarnata e situata che rifiuta il riduzionismo meccanicistico dell'animale come

[2] La nozione stessa di "biosemiotica", che era stata utilizzata con parsimonia durante i primi anni Sessanta, inizia lentamente a sostituire il termine conosciuto dalla comunità semiotica di "zoosemiotica". Lo stesso Sebeok nell'articolo collettivo apparso sulla rivista Semiotica del 1984, il quale delineava il programma di ricerca della biologia semiotica, esitava ancora a utilizzare questo termine (Anderson, et al., 1984). Allo stesso modo nel suo dizionario del 1986 aveva deciso di evitare il termine "biosemiotica" (Sebeok T. A., 1986). Infatti, da quando Sebeok (1963) coniò il termine di zoosemiotica nel 1963, ebbe inizio un continuo allargamento delle frontiere rispetto alla soglia inferiore della semiotica, ridestando interesse nel coniare nuove forme di semiosi all'interno del regno biologico (si veda Zengiaro 2024a). L'idea originale di Sebeok che legava i processi semiotici ai viventi con un sistema nervoso, venne messa in discussione da Martin Krampen che, nel 1981, pubblicò un articolo sull'ipotesi di indagare la semiosi nel mondo vegetale.

macchina stimolo-risposta (Zengiaro 2024b). Il tentativo di Sebeok (1965) di inquadrare i segnali animali attraverso le funzioni jakobsoniane segnò il culmine della spinta unitaria della zoosemiotica, pur riconoscendo l'impossibilità di rintracciare tratti metalinguistici o poetici al di fuori della specie umana. Il volume collettaneo *Animal Communication* (1968) rappresenta la sintesi di questa stagione, integrando contributi di etologi, ornitologi e biologi del calibro di Stuart A. Altmann, Peter Marler e E. O. Wilson. Ciononostante, l'influenza della rivoluzione chomskiana portò Sebeok a mettere in dubbio la validità di un approccio metodologico unico. Questo mutamento di prospettiva determinò la frammentazione del campo d'indagine, separando definitivamente lo studio dei sistemi comunicativi animali dalla riflessione sul linguaggio verbale, ormai orientata verso i principi della grammatica generativa.

È in questo contesto che il termine zoosemiotica diviene inadeguato e limitativo rispetto alle nuove ricerche in corso. Come aveva predetto a suo tempo anche Greimas, tuttavia, la zoosemiotica era destinata a «formare un vero e proprio campo semiotico, autonomo e promettente» (Greimas & Courtés 2007, 384). Sebeok, dunque, decide di allargare ufficialmente il campo di studi identificandolo con la "biosemiotica", un termine che aveva proposto Juri Stepanov (1971) nel 1971, ma che era già apparso con le sue declinazioni moderne in un articolo di Friedrich Rothschild (1962) nel 1962 e prima ancora dal chimico Vincent Kletzinsky nel 1855 (Kull 2022). Un'altra motivazione rispetto all'uso del termine "biosemiotica" in sostituzione a quello di zoosemiotica deriva dal fatto che gli studi di Sebeok nella comunicazione animale portarono con sé una risposta insoddisfacente da parte degli etologi. L'approccio zoosemiotico, come nota Kull (2003), non era stato colto come sufficientemente operativo da parte di chi si occupava di etologia; inoltre, gli studiosi del comportamento linguistico delle scimmie ritenevano che le teorie e le critiche di Sebeok non fossero del tutto pertinenti. Queste sono state sicuramente delle ragioni per le quali la ricerca di Sebeok si è poi diretta verso dei principi più fondamentali della biosemiotica in generale.

In particolare, come abbiamo osservato, il framework teorico della zoosemiotica stabilisce che la comunicazione animale deve essere intesa come un'attività interpretativa, relazionale e profondamente radicata nella *Umwelt*. Questa consapevolezza costituisce la base indispensabile per qualsiasi tentativo di integrazione con l'intelligenza artificiale. La sfida non risiede nella semplice "decodifica" di un codice presunto statico, bensì nella comprensione di una semiosi dinamica che non può essere ridotta a una grammatica senza vita o a un linguaggio senza soggetto. Solo attraverso il rispetto dell'opacità e l'apertura della semiosi non umana (Zengiaro 2025b), così come dell'alterità del vivente, la tecnologia può trasformarsi da strumento di traduzione antropocentrica in un dispositivo di mediazione ecologica capace di affinare la nostra percezione delle intelligenze non umane.

3. La sfida della "Traduzione"

I recenti sviluppi nell'ambito dell'intelligenza artificiale e delle tecnologie di rilevazione bioacustica hanno riacceso l'interesse scientifico per l'analisi sistematica della comunicazione animale non umana (Bermant *et al.* 2019).

In particolare, i progetti che mirano alla cosiddetta "traduzione" delle vocalizzazioni dei cetacei sollevano questioni epistemologiche e metodologiche che vanno ben oltre il riconoscimento automatico di pattern acustici, investendo direttamente i concetti di significato, intenzionalità e interpretazione (CETI Project 2024). La sfida centrale non consiste unicamente nel decodificare segnali, bensì nel chiarire in che senso tali segnali possano essere considerati atti comunicativi. In questo quadro, l'approccio semiotico offre una prospettiva critica fondamentale, in quanto consente di problematizzare l'analogia linguistica, di mettere in evidenza la natura relazionale del significato e di delineare con maggiore precisione i limiti e le potenzialità delle applicazioni di intelligenza artificiale nello studio della comunicazione interspecifica.

In questa prospettiva, il Progetto CETI si propone di affrontare in modo sistematico lo studio della comunicazione dei cetacei attraverso la raccolta su larga scala di dati multimodali di natura acustica e comportamentale. A tal fine, il progetto impiega un'infrastruttura tecnologica articolata che comprende una rete estesa di idrofoni subacquei sincronizzati, tag non invasivi applicati agli individui (D-tag), droni aerei e veicoli subacquei autonomi (Andreas *et al.* 2022; Wessling 2025).

Una delle principali sfide metodologiche riguarda la gestione, l'integrazione e l'allineamento dei flussi di dati eterogenei provenienti da sensori differenti, nonché la loro annotazione e preparazione per l'analisi tramite tecniche di intelligenza artificiale. Su questa base, il progetto mira allo sviluppo di modelli avanzati di machine learning, inclusi modelli linguistici specificamente progettati per le vocalizzazioni dei cetacei (cosiddetti Whale Language Models o NatureLM-audio (Hagiwara, Miron, Liu 2024)), con l'obiettivo di individuare strutture ricorrenti nella comunicazione e di metterle in relazione con i comportamenti osservati. La validazione empirica delle ipotesi e dei modelli sviluppati avviene mediante esperimenti di playback interattivi, accuratamente progettati per testare le risposte comportamentali degli animali a specifici stimoli acustici.

Un risultato particolarmente rilevante emerso di recente dall'analisi di migliaia di codas — sequenze di click prodotte dai capodogli — consiste nell'identificazione di un insieme finito di unità acustiche ricorrenti, interpretabili come una sorta di "alfabeto fonetico" della specie (Stowell 2022). Tale scoperta è stata resa possibile dall'impiego di algoritmi di riconoscimento e classificazione dei pattern applicati a grandi volumi di dati acustici.

Un tratto comune a iniziative quali DolphinGemma, CHAT (Cetacean Hearing and Telemetry) e ISPA (Inter-Species Phonetic Alphabet), sviluppato nell'ambito dell'Earth Species Project (ESP), è l'applicazione di tecniche di Natural Language Processing (NLP) e Deep Learning allo studio delle vocalizzazioni dei cetacei. Questi progetti condividono l'obiettivo di costruire sistemi di trascrizione e modellazione ispirati ai paradigmi della linguistica computazionale, ipotizzando l'esistenza di strutture combinatorie e regolarità formali che possano essere descritte mediante modelli linguistici capaci di riconoscere e, in prospettiva, generare sequenze "whalish" (Goldwasser *et al.* 2023, 2).

L'analisi della comunicazione animale mediata dall'intelligenza artificiale si fonda prevalentemente su ampie raccolte di vocalizzazioni registrate in ambiente naturale, incluse emissioni acustiche associate ad azioni

ripetitive e interazioni sociali all'interno dei gruppi. I modelli di IA vengono addestrati secondo procedure analoghe a quelle impiegate per gli LLM, al fine di individuare pattern statistici utili al riconoscimento degli atti comunicativi. In questo contesto, assumono particolare rilevanza anche le questioni relative all'agency degli animali coinvolti e alle implicazioni etiche connesse alla raccolta, all'interpretazione e all'uso dei dati (Rodríguez-Garavito 2024; Küçükuncular 2025).

Le critiche rivolte all'analisi degli atti comunicativi tra animali non umani derivano in larga misura dalle profonde differenze che li separano dal linguaggio umano (Eco 2018). La ricerca scientifica sulla comunicazione animale ha spesso assunto il linguaggio verbale umano come modello di riferimento implicito, orientando l'indagine verso la ricerca di strutture analoghe e privilegiando una concezione informazionale della comunicazione. I primi etologi, tra cui Konrad Lorenz, Niko Tinbergen e Karl von Frisch, concepirono la comunicazione come un trasferimento di informazioni funzionale al benessere della specie (Merler 1979). Come abbiamo accennato precedentemente, il decennio 1950-1960 ha segnato una transizione cruciale verso una visione integrata della comunicazione, in cui il confine tra linguistica e biologia si è fatto poroso. Attraverso i contributi di George Miller, Ray Birdwhistell e Edward Hall, i fenomeni comunicativi smettevano di essere un'esclusiva umana per diventare un campo d'indagine trans-specifico, trovando, come suggeriva l'antropologa Margaret Mead, nell'etichetta della "semiotica" la sua definizione epistemologica definitiva (Gensini & Palazzolo 2019, 234). In questo contesto, il modello dei *design features* di Charles Hockett (1958) ha rappresentato il tentativo più compiuto di ricondurre la varietà dei segnali biologici entro una matrice teorica comune, permettendo di analizzare il linguaggio umano non più come un'entità isolata, ma come una configurazione specifica di tratti condivisi con altre specie.

Questa impostazione è stata successivamente problematizzata da prospettive alternative, in particolare da Dawkins e Krebs (1979), che hanno proposto una visione strategica della comunicazione animale. Secondo tale approccio, i segnali non sono selezionati per informare il ricevente, bensì per influenzarne il comportamento a vantaggio dell'emittente. Un esempio emblematico è costituito dal richiamo del maschio della rana túngara (*Physalaemus pustulosus*), che arricchisce il segnale base (*whine*) con componenti aggiuntive (*chucks*) per aumentare l'attrattività nei confronti delle femmine. Tali componenti sembrano sfruttare bias sensoriali preesistenti, senza apportare benefici diretti alle femmine e, anzi, aumentando il rischio di predazione da parte di pipistrelli attratti da segnali più complessi. Tale spostamento teorico dalla trasmissione di informazioni alla manipolazione strategica e alla coevoluzione tra segnale e risposta evidenzia la complessità dei sistemi comunicativi animali.

La complessità della comunicazione animale rappresenta al contempo una sfida significativa e un'opportunità per lo sviluppo di modelli di intelligenza artificiale. I sistemi che aspirano a "decodificare" o "tradurre" la comunicazione animale non possono limitarsi a una semplice corrispondenza tra segnale e significato, o tra vocalizzazione e comportamento, ma devono integrare dimensioni contestuali, dinamiche evolutive, elementi di teoria dei giochi e aspetti cognitivi e percettivi del ricevente (Zengiaro 2025a).

Il dibattito sulle relazioni evolutive e funzionali tra i sistemi comunicativi umani e non umani ha inoltre messo in luce le profonde differenze tra i dispositivi di codifica del linguaggio verbale e quelli della comunicazione cinesica e paralinguistica. Tuttavia, è stato osservato come esistano significative affinità tra i codici della cinesica e del paralinguaggio umano e quelli dei mammiferi non umani (Bateson 1979).

Alla luce di queste considerazioni, appare necessario progettare sistemi di intelligenza artificiale capaci di tener conto delle specificità della comunicazione animale, evitando un'applicazione diretta e non critica dei modelli del linguaggio naturale umano (Adornetti 2024). In questo solco si inseriscono le recenti frontiere della ricerca bioacustica, che tentano di rintracciare analogie strutturali ancora più profonde tra la fonetica umana e i sistemi di segnalazione dei cetacei. Studi recenti sulle vocalizzazioni dei capodogli (Beguš *et al.* 2025a, 2025b) hanno infatti evidenziato la presenza di pattern spettrali sistematici all'interno dei codas, identificando variazioni nelle frequenze formanti che ricalcano la configurazione delle vocali e dei dittonghi umani. Tale scoperta ha alimentato l'ambizione di sviluppare modelli traslativi complessi, come il Whale Alphabet Model (WhAM) (Paradise *et al.* 2025), volti a discretizzare il flusso sonoro animale in un "alfabeto" computabile attraverso l'intelligenza artificiale.

Tuttavia, l'applicazione di questi modelli solleva interrogativi critici che richiamano la necessità di un approccio biosemiotico. Sebbene l'individuazione di una fonologia nelle balene offra una straordinaria svolta nello studio della comunicazione animale, il tentativo di tradurre tali sequenze in unità linguistiche discrete rischia di incorrere in un riduzionismo antropocentrico. Dalla prospettiva della zoosemiotica, infatti, il significato non risiede in una corrispondenza biunivoca tra segno e referente, ma nel principio della biotraduzione (Kull, Torop 2003): un processo interpretativo incarnato in cui il segnale non viene semplicemente "decriptato", ma tradotto in azione e coordinazione vitale all'interno dello specifico Umwelt della specie o in spazi di sovrapposizione tra Umwelten.

Affinché i modelli di intelligenza artificiale risultino scientificamente solidi e non meramente proiettivi, essi devono integrare la consapevolezza che la comunicazione non umana non è una Stele di Rosetta latente in attesa di essere tradotta. La sfida non risiede nel mappare i segnali animali su una sintassi antropocentrica, quanto nel riconoscere una semiosi che opera in una catena evolutiva e un'interpretazione biologica. Solo superando il bias della traduzione come "conversione di codici" in favore di una comprensione dei processi di biotraduzione sarà possibile evitare che la tecnologia riduca l'alterità animale a un riflesso del logos umano, perdendo di vista la specificità dei sistemi di significazione non verbali.

4. I meccanismi interni dell'IA per la comunicazione tra animali e umani

L'intelligenza artificiale presenta un potenziale significativo nell'analisi di grandi quantità di dati e nell'identificazione di pattern rilevanti attraverso l'impiego di tecniche statistiche e logiche avanzate. In particolare, i sotto-ambiti dell'IA quali il machine learning (ML) e il deep learning (DL) mettono a disposizione strumenti particolarmente efficaci per superare molte delle limitazioni proprie dei metodi di analisi tradizionali. A titolo

esemplificativo, studi recenti hanno applicato tecniche di IA all'analisi del comportamento dei topi, dimostrando la capacità di tali sistemi di identificare e classificare con elevata accuratezza comportamenti complessi (Guo *et al.* 2024). Tutti questi avanzamenti nello sviluppo dell'IA sono resi possibili da una specifica classe di tecnologie di intelligenza artificiale – Large Language Models (LLM).

Tuttavia, è necessario chiarire preliminarmente quale tipo di tecnologia siano gli LLM e quali modalità operative essi adottino nel contesto della comunicazione umana, nonché valutare criticamente le possibili applicazioni nell'ambito della comunicazione interspecifica. Gli LLM rappresentano un approccio del machine learning volto alla modellazione dei linguaggi attraverso la scomposizione dei loro elementi costitutivi, trattati come parametri, e l'individuazione di pattern complessi a partire da dataset di dimensioni massicce. Tale capacità consente a questi modelli di svolgere compiti quali la risposta a domande, la traduzione, la sintesi e la generazione di contenuti, operando essenzialmente come sofisticati motori predittivi basati sul testo (Glazkov & Makarov 2024; Yildirim & Paul 2024; Vats *et al.* 2026).

Gli LLM analizzano sequenze di token, che spesso corrispondono a parole o loro parti, assegnando i parametri in funzione delle relazioni statistiche che tali token intrattengono con altri elementi della sequenza. In sistemi come i linguaggi naturali umani, regolati da strutture semantiche e sintattiche relativamente stabili, questo approccio consente di identificare tali relazioni con un elevato grado di accuratezza. Proprio su questa base si fonda anche la possibilità di traduzione automatica tra lingue diverse, presupponendo una sostanziale condivisibilità delle dimensioni sintattiche e semantiche del linguaggio umano. Ad esempio, concetti come caldo e freddo vengono rappresentati come opposizioni all'interno di spazi vettoriali costruiti dai parametri, analogamente a coppie come tempo caldo e tempo freddo. Tuttavia, nel tentativo di estendere questi modelli algoritmici a specie non umane, diviene imprescindibile considerare il ruolo dell'Umwelt, *Merkwelt* e *Wirkwelt* proprie di ciascuna specie. A titolo esemplificativo, per una specie dotata di caratteristiche fisiche e fisiologiche differenti da quelle umane, come le lucertole, la nozione di tempo caldo potrebbe corrispondere a una fase di maggiore attività, mentre il tempo freddo potrebbe essere associato a rallentamento o letargo invernale. Questo esempio evidenzia come i presupposti semantici impliciti nei modelli linguistici umani non siano direttamente trasferibili ad altri sistemi di vita e come, di conseguenza, gli LLM debbano essere profondamente ripensati e riconfigurati per tener conto delle specificità specie-specifiche, della loro relazione con il mondo fisico e della complessità intrinseca dei processi di traduzione o decodifica automatizzata della comunicazione interspecifica.

Gli LLM sono architettati per processare linguaggi umani dotati di proprietà specifiche: discretezza (separazione in unità minime identificabili), ricorsività (capacità di generare infinite strutture da un set finito di regole), arbitrarietà del segno (relazione convenzionale tra significante e significato) e componibilità sintattica (combinazione di elementi secondo regole grammaticali). Queste qualità permettono ai modelli di operare attraverso meccanismi di predizione statistica su sequenze discrete, apprendendo correlazioni tra pattern linguistici da enormi dataset testuali. I modelli Generative Pre-trained Transformers (GPT) rappresentano

attualmente una delle classi degli LLM più affermate ed efficaci nel campo del Natural Language Processing (NLP). L'architettura dei Transformer, introdotta nel 2017 e formalizzata da Vaswani *et al.* (2017), è stata inizialmente concepita per compiti di *sequence transduction*. Tuttavia, nel corso degli anni, essa si è affermata come architettura di riferimento per un'ampia gamma di applicazioni nel trattamento automatico del linguaggio naturale (Kim *et al.* 2023) ed è stata estesa con successo anche ad altri ambiti, tra cui la *computer vision* (Khan *et al.* 2022) e l'elaborazione del parlato, *speech processing* (Latif *et al.* 2023). Questi modelli, utilizzati nel campo della computer vision e nell'analisi dello *speech processing*, rivestono un'importanza particolare nell'analisi dei dati bioacustici. In particolare, il modello Transformer standard, basato sul meccanismo di self-attention, continua a prevalere nello sviluppo degli LLM (Touvron *et al.* 2023), grazie alla sua elevata capacità di parallelizzazione nel processamento delle sequenze linguistiche e alla scalabilità che ne consente l'addestramento su corpora testuali di dimensioni molto estese (Zhao *et al.* 2023). È possibile discutere l'impiego dei modelli come GPT nelle applicazioni di comunicazione uomo–animale facendo riferimento, in particolare, alle pratiche di playback, come nei progetti DolphinGemma e CETI. La funzione principale dei GPT risiede infatti nei compiti di sequence transduction. Con il termine "sequence transduction" si intendono quei compiti di ML che prevedono la trasformazione di una sequenza di input in una sequenza di output, entrambe potenzialmente variabili in lunghezza e struttura. Tale concetto è centrale in numerose applicazioni, in particolare nello speech processing. I modelli di sequence transduction sono progettati per gestire dati sequenziali di lunghezza variabile, consentendo una comprensione contestuale delle relazioni interne alle sequenze (Chen 2023).

Tuttavia, nel contesto dell'analisi dei dati bioacustici, questo approccio può risultare particolarmente problematico, pur essendo in grado di produrre output credibili che imitano la comunicazione animale o che simulano tentativi di traduzione sincrona delle vocalizzazioni in linguaggio naturale. Il limite principale risiede nel fatto che la rigidità grammaticale dei linguaggi naturali umani consente di eseguire il processo di traduzione attraverso l'elaborazione di sequenze formali, senza necessariamente considerare il messaggio nella sua globalità e nel suo contesto d'uso. Tale impostazione risulta tuttavia antitetica alla comunicazione animale, la quale è profondamente situata, fortemente dipendente dal contesto e intrinsecamente legata alla dimensione pragmatica dell'interazione. Nel contesto della comunicazione interspecifica, come evidenziato nei capitoli precedenti, la dimensione zoopragmatica assume un ruolo centrale. La funzione e il significato dei segnali animali emergono infatti in larga misura dalle condizioni d'uso, dalle interazioni situate e dagli effetti comportamentali prodotti sui riceventi. Di conseguenza, la progettazione di modelli LLM in applicazioni di comunicazione uomo–animale richiede un'integrazione esplicita della dimensione pragmatica, senza la quale i processi di modellazione, generazione e playback dei segnali rischiano di risultare parziali o epistemologicamente inadeguati rispetto alla complessità della comunicazione animale.

Come già discusso, progetti quali CETI, DolphinGemma e ISPA propongono l'applicazione dei LLM alle vocalizzazioni dei cetacei, assumendo

che i principi computazionali sviluppati per il linguaggio umano, basati su apprendimento automatico, riconoscimento di pattern e modellazione probabilistica, possano essere trasferiti alle produzioni vocali non umane. Tuttavia, poiché gli LLM operano tramite predizione sequenziale elemento per elemento, essi mostrano una limitata capacità di attribuire il significato in modo pienamente olistico e tendono a imporre una griglia interpretativa antropocentrica, che presuppone l'esistenza di un codice stabile e decodificabile, astratto dal contesto ecologico e sociale dei segnali (Yovel & Rechavi 2023). Ne consegue che tale approccio non può essere considerato uno strumento universale per la comunicazione interspecifica, nemmeno quando applicato a una singola specie come i cetacei.

Le criticità emerse alla luce dei fondamenti della biosemiotica e della zoosemiotica evidenziano possibili bias intrinseci nell'architettura stessa degli LLM, derivanti da assunzioni implicite proprie del trattamento computazionale del linguaggio umano. Per superare il mero riconoscimento di pattern e avanzare verso l'analisi di dimensioni più complesse, quali sintassi, semantica e soprattutto pragmatica, risulta cruciale integrare lo studio della comunicazione con quello del comportamento, della cognizione e degli stati emotivi e fisiologici degli animali, tenendo conto del loro specifico Umwelt. In tale prospettiva, se sviluppati in modo mirato per gruppi specifici, caratterizzati da pratiche comunicative ricorrenti e da una limitata variabilità ambientale, gli approcci statistici basati su LLM possono risultare efficaci entro confini ben definiti.

Ciononostante, riconosciuti i limiti strutturali degli attuali modelli di intelligenza artificiale, diventa fondamentale interrogarsi su come tali tecnologie possano contribuire a superare i limiti percettivi e fisici dell'essere umano, arricchendosi dei contributi teorici della semiotica, della biosemiotica e della zoosemiotica, al fine di sviluppare sistemi di IA più sensibili alla pluralità dei segni, ai contesti ecologici e alla complessità relazionale della comunicazione interspecifica.

5. Verso un'IA Relazionale

Anziché concepire l'IA come strumento di traduzione automatica, si dovrebbe proporre di svilupparla come una tecnologia di mediazione interpretativa, capace di affinare la percezione umana rispetto alle vocalizzazioni animali senza imporre schemi linguistici precostituiti. Questo approccio richiede l'abbandono degli LLM tradizionali in favore di architetture computazionali progettate per riconoscere pattern non solo bioacustici, ma anche visivi, chimici, tattili, vibrazionali, ambientali, sociali in relazione al contesto ecologico e comportamentale. Evidenziare regolarità senza assumere strutture sintattiche predefinite e integrare dati multimodali per una comprensione ecosistemica della comunicazione può aiutare a superare i limiti presenti nelle applicazioni di IA di oggi.

Tra i principali vantaggi dell'IA si annoverano l'efficienza, la scalabilità, l'aumento dell'oggettività e la possibilità di generare scoperte innovative. In particolare, l'IA consente l'elaborazione rapida di enormi volumi di dati, inclusi i dati raccolti con le tecnologie dotate di IA capaci di superare i limiti dell'Umwelt umano (Arkhipova 2023).

Strumenti avanzati, come quelli basati sulla computer vision, consentono la raccolta di dati sugli animali nel loro ambiente naturale senza

interferire con l'espressione spontanea dei comportamenti. Queste tecnologie si distinguono per l'efficacia nella gestione di dataset di grandi dimensioni, nella rilevazione di pattern sottili e complessi difficilmente osservabili dall'essere umano e nell'elaborazione di informazioni provenienti da modalità sensoriali eterogenee, quali i segnali acustici (nell'ambito della bioacustica computazionale) e le informazioni visive derivate dall'analisi video del comportamento (Delahaye 2021). La capacità dell'IA di elaborare dati a velocità e su scale senza precedenti apre nuove prospettive nella decodifica dei segnali animali, nella comprensione dei comportamenti sociali e nel monitoraggio della biodiversità. Studi recenti hanno documentato l'impiego di tecniche di machine vision e di analisi basate sull'IA per il monitoraggio del comportamento di primati, felini, roditori, delfini e balene, sia in contesti naturali sia in condizioni di cattività (Vogg *et al.* 2024).

L'automazione dei processi analitici riduce la soggettività tipica dell'osservazione umana, favorendo risultati più standardizzati, coerenti e riproducibili tra diversi studi e gruppi di ricerca. Inoltre, gli algoritmi di IA sono in grado di individuare pattern sottili o altamente complessi che spesso sfuggono alla percezione umana, inclusi segnali al di fuori della gamma uditiva umana (come ultrasuoni e infrasuoni) o segnali estremamente deboli immersi nel rumore ambientale. L'intelligenza artificiale, pertanto, non si limita ad accelerare l'analisi dei dati, ma rende possibile l'individuazione di strutture e relazioni precedentemente ignote, come dimostrato, ad esempio, dalla scoperta della complessità combinatoria delle vocalizzazioni dei capodogli. Essa consente inoltre la progettazione di nuove tipologie di esperimenti, quali quelli basati su playback generati dall'IA, aprendo così orizzonti di ricerca finora inesplorati.

Studi precedenti confermano che le ricerche più efficaci sui processi comunicativi animali sono state condotte da studiosi che hanno raccolto i dati vivendo a stretto contatto con i gruppi di animali oggetto di studio. Tale prossimità ha consentito di andare oltre la mera analisi dei dati raccolti, permettendo ai ricercatori di diventare, in una certa misura, membri "inermi" o integrati del gruppo sociale osservato (Kummer & Goodall 1985; Sapolsky & Share 2004). I recenti avanzamenti nell'IA rendono oggi possibile elaborare e ampliare questo tipo di approccio personalizzato allo studio delle comunità animali, consentendo di tenere conto delle specificità locali, delle caratteristiche ambientali e dei particolari fattori di stimoli a cui i gruppi sono esposti. Tale prospettiva permette di individuare pattern di somiglianza tra gruppi differenti, di analizzare le differenze e di osservare le trasformazioni dei sistemi comunicativi nel tempo.

Si propone lo sviluppo di modelli IA biosemioticamente informati che incorporano i principi fondamentali della zoosemiotica: la dimensione pragmatica e situata della comunicazione, la varietà dei tipi di segno (iconici, indicali, simbolici), e la funzione performativa e relazionale delle vocalizzazioni. Questi modelli dovrebbero operare come amplificatori percettivi che consentano ai ricercatori di esplorare l'Umwelt delle specie studiate senza ridurlo a categorie antropocentriche.

Alla luce dei principi fondamentali che costituiscono l'ossatura teorica della disciplina (Sebeok 1986, 1179), la progettazione di sistemi di IA per la traduzione della comunicazione animale dovrebbe essere impostata come un'architettura computazionale multilivello, capace di modellare in

modo integrato le diverse fasi del processo comunicativo. In particolare, il sistema dovrebbe includere moduli dedicati alla modellazione dei processi di formulazione e codifica dei segnali da parte dell'emittente, catturando la dinamica temporale e la strutturazione dei messaggi; alla gestione della trasmissione, con l'esplicita rappresentazione dei canali comunicativi, delle condizioni ambientali e, nei sistemi multicanale, delle regole di coordinamento, combinazione o commutazione tra canali differenti; e alla decodifica e interpretazione lato ricevente, tenendo conto dei vincoli percettivi e cognitivi specifici della specie modellata. L'architettura dovrebbe inoltre incorporare una rappresentazione formale del repertorio comunicativo, includendo la tipologia dei segnali, le loro proprietà acustiche, visive o multimodali, le regole di composizione sequenziale e le strutture informative emergenti, nonché una caratterizzazione esplicita delle proprietà strutturali del codice comunicativo. Ulteriori componenti chiave riguardano la modellazione del contesto, inteso come insieme di variabili ambientali, sociali e situazionali che condizionano la produzione e l'interpretazione dei segnali; l'integrazione di modelli evolutivi in grado di simulare la trasformazione di comportamenti non comunicativi in segni funzionali attraverso processi di ritualizzazione; e, infine, la rappresentazione dell'ontogenesi comunicativa, mediante modelli di apprendimento che descrivono lo sviluppo dei sistemi di segni nel corso del ciclo di vita dell'individuo. Un simile approccio consente di allineare i principi teorici della comunicazione animale con pratiche di progettazione computazionale, favorendo la costruzione di sistemi di IA interpretabili, scalabili e sensibili alle specificità biologiche delle specie considerate.

6. Conclusione

In conclusione, il presente contributo, come parte della ricerca svolta nel progetto "San Francesco e l'IA", ha messo in evidenza le sfide epistemologiche, semiotiche e pratiche legate all'impiego delle tecnologie di intelligenza artificiale nello studio della comunicazione animale, con un particolare focus sui tentativi di "traduzione" delle vocalizzazioni animali, come nel caso dei cetacei. I risultati ottenuti suggeriscono che l'approccio computazionale tradizionale, che cerca di assimilare le vocalizzazioni animali a un linguaggio strutturato e decodificabile (Beguš *et al.* 2025b), presenta limiti significativi. In particolare, tale approccio tende a ridurre la complessità e la relazionalità della comunicazione animale, ignorando la sua dimensione ecologica e contestuale. Il lavoro, basato sugli studi derivanti dal progetto condotto all'Università di Torino, ha dimostrato che la biosemiotica, concependo la comunicazione come un processo dinamico e situato, fornisce una base teorica per un'interpretazione più accurata e rispettosa delle modalità comunicative degli animali, che vanno oltre la mera decodifica dei segnali.

Il progetto ha quindi proposto un cambiamento di paradigma, spostando l'attenzione dall'idea di "traduzione" a quella di una mediazione ecologica, dove l'IA non è un semplice strumento di decodifica, ma una tecnologia interpretativa che rispetta la complessità dei sistemi semiotici viventi. Si è sottolineato come la progettazione di sistemi IA che tengano conto della pluralità e della varietà dei segni animali, dei contesti in cui si inseriscono e delle relazioni interspecifiche, possa condurre a modelli di comunicazione più equi e scientificamente rigorosi. Infine, il lavoro ha

evidenziato la necessità di sviluppare tecnologie in grado di interagire con la percezione animale, rispettando l'opacità e l'alterità dei segnali, senza ridurre la comunicazione interspecie a un mero processo informativo. Il progetto suggerisce così la direzione verso un'intelligenza artificiale relazionale ed ecologica, che possa contribuire a una comprensione più profonda e rispettosa delle comunicazioni non umane, favorendo forme di coesistenza interspecifica.

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 73 — 89

Bibliografia

Adornetti, I. (2024). Animal Signalling Between Informing and Influencing: Setting the Stage for a Pragmatic-Rhetorical Model of Communication. In I. Adornetti & E. Ferretti (eds), *Introducing Evolutionary Pragmatics* (23–38). Routledge.

Anderson, M. et al. (1984). A Semiotic Perspective on the Sciences: Steps Toward a New Paradigm. *Semiotica*, 52 (1–2), 7–47.

Andreas, J. et al. (2022). Toward Understanding the Communication of Sperm Whales. *iScience*, 25 (6), 104393.

Arkhipova, D. (2023). Materia Digitale: L'Impatto dei Social Media Basati Sull'IA Sulla Dimensione Materiale degli Utenti. *EIC*, 38, 296–304.

Bateson, G. (1979). *Mind and Nature: A Necessary Unity*. Dutton.

Beguš, G. et al. (2025a). The Phonology of Sperm Whale Coda Vowels. *bioRxiv*.

Beguš, G. et al. (2025b). Vowel – and Diphthong-Like Spectral Patterns in Sperm Whale Codas. *Open Mind*, 9, 1849–1874.

Bermant, P. C. et al. (2019). Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics. *Scientific Reports*, 9, 12588.

CETI Project. (2024). Sperm Whale Phonetic Alphabet Proposed for the First Time. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

Chen, Z. (2023). Attention Is Not All You Need Anymore. *arXiv*.

Cobley, P. (2014). Codes and Coding: Sebeok's Zoosemiotics and the Dismantling of the Fixed-Code Fallacy. *Semiotica*, 198, 33–45.

Davies, N. B., Krebs, J. R. & West, S. A. (1981). *An Introduction to Behavioural Ecology*. Blackwell.

Dawkins, R. & Krebs, J. R. (1979). Animal Signals: Information or Manipulation? In J. R. Krebs & N. B. Davies (eds), *Behavioral Ecology: An Evolutionary Approach* (282–309). Blackwell.

Deely, J. (2001). Umwelt. *Semiotica*, 134 (1–4), 125–135.

Delahaye, P. (2021). Use of Quantitative Measures in Zoosemiotics: How Machines Are Becoming a New Pair of Ears and Eyes for Researchers. *Biosemiotics*, 14, 287–294.

Eco, U. (2018). Animal Language Before Sebeok. *Sign Systems Studies*, 46 (2–3), 365–377.

Gensini, S. (2018). A Proposito di Zoosemiotica: L'Inizio della Storia. In M. W. Bruno et al. (a cura di), *Linguistica e Filosofia del Linguaggio. Studi in Onore di Daniele Gambarara* (263–282). Mimesis.

Gensini, S. & Palazzolo, G. (2019). Simbolicità e Emotività negli Animali Non Umani. *Rivista Italiana di Filosofia del Linguaggio*, 230–241.

Glazkov, N. & Makarov, I. (2024). Utterance-Aware Adaptive Data Labeling and Summarization. *IEEE Access*.

Goldwasser, S. et al. (2023). A Theory of Unsupervised Translation Motivated by Understanding Animal Communication. *arXiv*.

Greimas, A. J. & Courtés, J. (2007). *Semiotica: Dizionario della Teoria del Linguaggio*. Mondadori.

Guo, C. et al. (2024). A Survey on AI-Driven Mouse Behavior Analysis Applications and Solutions. *Bioengineering*, 11 (11), 1121.

Hagiwara, M., Miron, M. & Liu, Y. (2024). ISPA: Inter-Species Phonetic Alphabet for Transcribing Animal Sounds. In *ICASSPW* (828–832).

Hockett, C. F. (1958). *A Course in Modern Linguistics*. IBH Publishing.

Hoffmeyer, J. (1996). *Signs of Meaning in the Universe*. Indiana University Press.

Hoffmeyer, J. (2008). *Biosemiotics: An Examination Into the Signs of Life and the Life of Signs*. University of Scranton Press.

Khan, S. et al. (2022). Transformers in Vision: A Survey. *ACM Computing Surveys*, 54 (10s).

Kim, S. et al. (2023). Big Little Transformer Decoder. *arXiv*.

Krampen, M. (1981). Phytosemiotics. *Semiotica, 36* (3–4), 187–209.

Kull, K. (2003). Thomas A. Sebeok and Biology: Building Biosemiotics. *Cybernetics and Human Knowing*, 10 (1), 47–60.

Kull, K. (2014). Zoosemiotics Is the Study of Animal Forms of Knowing. *Semiotica*, 198, 47–60.

Kull, K. (2022). The Term 'Biosemiotik' in the 19th Century. *Sign Systems Studies*, 50 (1), 173–178.

Kull, K. & Torop, P. (2003). Biotranslation: Translation Between Umwelten. In S. Petrilli (ed.), *Translation Translation* (313–328). Rodopi.

Kummer, H. & Goodall, J. (1985). Conditions of Innovative Behaviour in Primates. *Philosophical Transactions of the Royal Society B*, 308 (1135), 203–214.

Latif, S. et al. (2023). Transformers in Speech Processing: A Survey. *arXiv*.

Maran, T. (2010). Why Was Thomas A. Sebeok Not a Cognitive Ethologist? *Biosemiotics*, 3 (3), 315–329.

Maran, T., Martinelli, D. & Tirovski, A. (2011). *Readings in Zoosemiotics*. Mouton de Gruyter.

Martinelli, D. (2010). *A Critical Companion to Zoosemiotics*. Springer.

Marler, P. (1979). Ethology and the Study of Animal Communication. In A. H. McGrew (ed.), *Animal Behavior* (212–225). Cambridge University Press.

Morris, C. W. (1946). *Signs, Language, and Behavior*. Prentice Hall.

Paradise, O. et al. (2025). WhAM: Towards a Translative Model of Sperm Whale Vocalization. *arXiv*.

Rothschild, F. S. (1962). Laws of Symbolic Mediation in the Dynamics of Self and Personality. *Annals of the New York Academy of Sciences, 96*, 774–784.

Sapolsky, R. M. & Share, L. J. (2004). A Pacific Culture Among Wild Baboons. *PLoS Biology*, 2 (4), e106.

Sebeok, T. A. (1963). Review of Communication Among Social Bees. *Language*, 39 (3), 448–466.

Sebeok, T. A. (1965). Animal Communication. *Science*, 147 (3661), 1006–1014.

Sebeok, T. A. (1968). *Animal Communication: Techniques of Study and Results of Research*. Indiana University Press.

Sebeok, T. A. (1986). *Encyclopedic Dictionary of Semiotics* (Vols. 1–3). Mouton de Gruyter.

Sebeok, T. A. (ed.) (1990). *Essays in Zoosemiotics*. Toronto Semiotic Circle.

Sebeok, T. A. (1997). The Evolution of Semiosis. In R. Posner, K. Robering & T. A. Sebeok (eds.), *Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture* (Vol. 1, 436–446). Walter de Gruyter.

Sebeok, T. A. (1998). The Estonian Connection. *Semiotica, 126*, 20–41.

Sebeok, T. A. (2001). *Global Semiotics*. Indiana University Press.

Sehgal, A., Tumar, I. & Schönwälder, J. (2010). Effect of Climate Change and Anthropogenic Ocean Acidification on Underwater Acoustic Communication. In *Proceedings of OCEANS'10 IEEE Sydney* (1–6). IEEE.

Stepanov, J. S. (1971). *Semiotika*. Nauka.

Stowell, D. (2022). Computational Bioacoustics With Deep Learning. *PeerJ*, 10, e13152.

Touvron, H. et al. (2023). LLaMA 2: Open Foundation and Fine-Tuned Chat Models. *arXiv*.

Uexküll, J. von. (1921 [1909]). Umw*elt und Innenwelt der Tiere*. Springer.

Vaswani, A. et al. (2017). Attention Is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (6000–6010).

Vats, S. et al. (2026). How Do LLMs Work?: A Deep Dive Into Transformer Models. In A. Kumar et al. (eds), *Digital Watermarking in Cloud Environments For Copyright Protection* (86–106). IGI Global Scientific Publishing.

Vogg, R. et al. (2024). PriMAT: A robust multi-animal tracking model for primates in the wild. *bioRxiv*, 1-17. https://doi.org/10.1101/2024.08.21.607881

Wessling, E. G. et al. (2025). Pathways from theory to practice: applying animal cultures to conservation. *Wildlife conservation*, *1*, 2.

Yildirim, I. & Paul, L. A. (2024). From Task Structures to World Models: What Do LLMs Know? *Trends in Cognitive Sciences*, 28 (5), 404–415.

Yovel, Y. & Rechavi, O. (2023). AI and the Doctor Dolittle Challenge. *Current Biology*, 33 (15), 783–787.

Zengiaro, N. (2024a). En Búsqueda de los Límites de la Semiótica: Desde el Umbral Inferior de Eco Hasta la Ecosemiótica Materialista. *Metatheoria*, 14 (2), 75–92.

Zengiaro, N. (2024b). Vibrant Worlds: An Artistic Interpretation of Material Intelligence in the Spider's Umwelt. *Biosemiotics*, 17, 671–691.

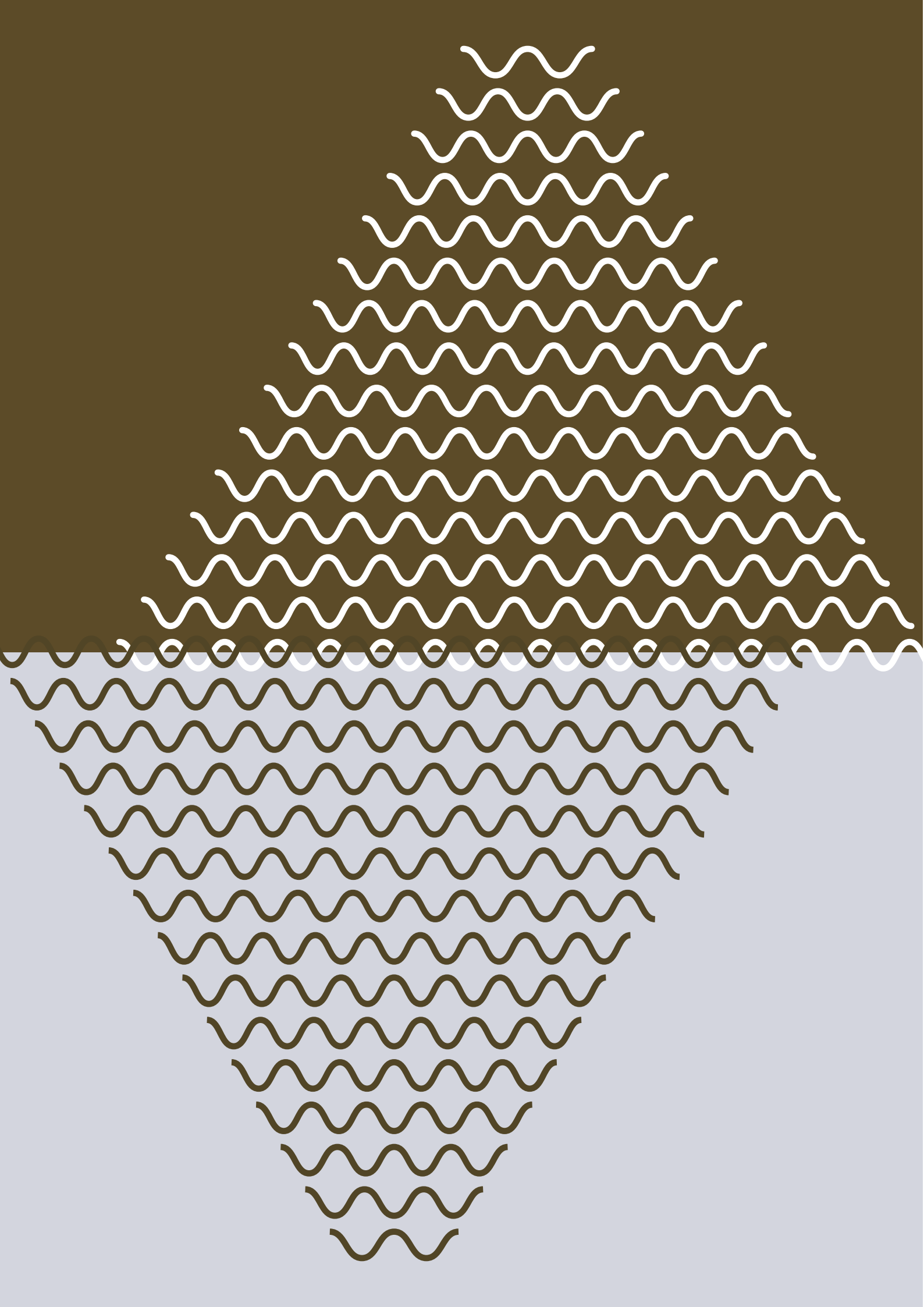
Zengiaro, N. (2025a). Translating the Wild: AI, Semiotics, and the Future of Animal Communication. *Digital Age in Semiotics & Communication*, 8, 178–205.

Zengiaro, N. (2025b). For a Semiotic of Opacity: The Role of Biosemiotics Between AI and Animal Communication. *Chinese Semiotic Studies*, 21 (4), 567–594.

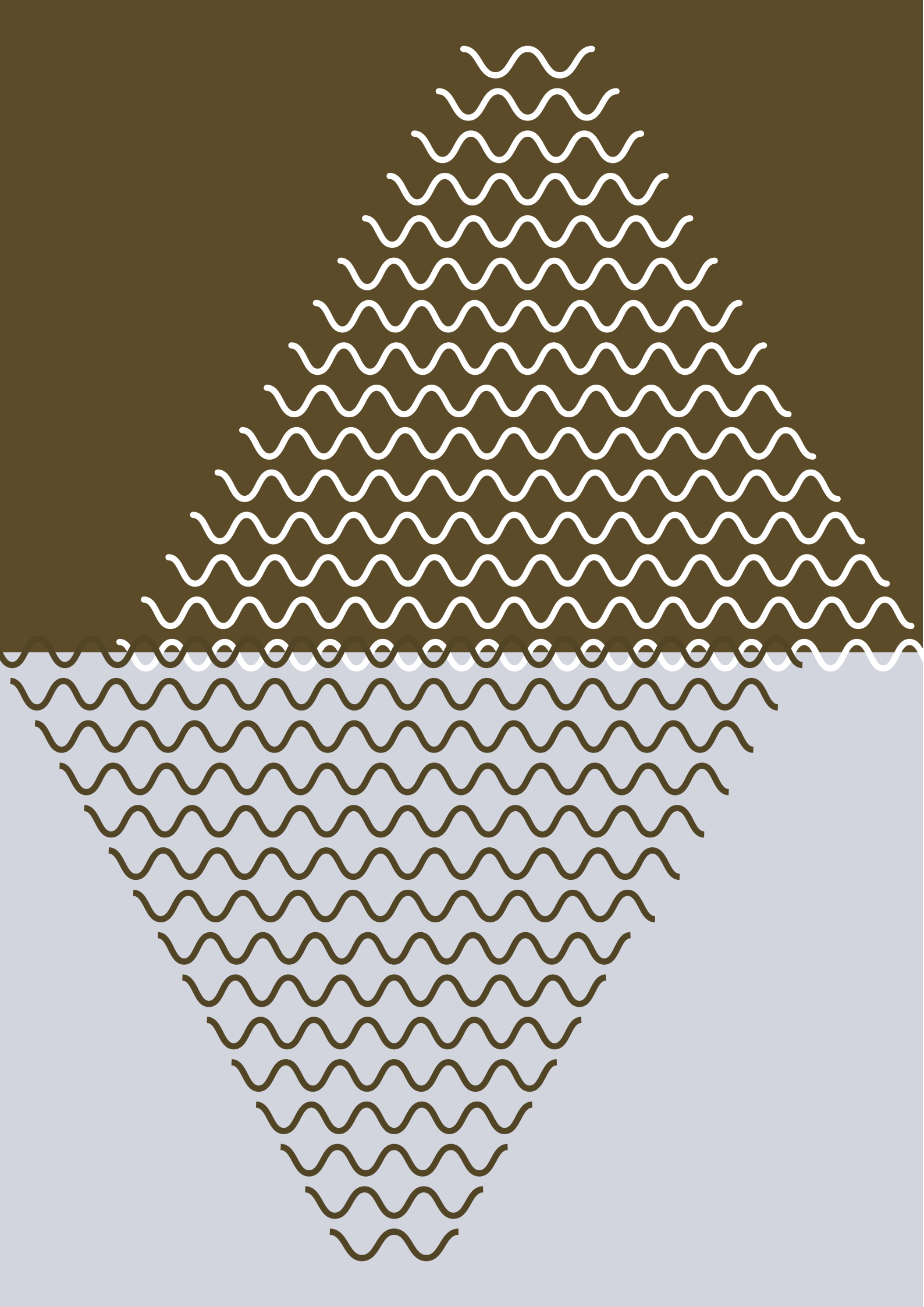
Zengiaro, N. (2026). AI and Animal Communication. A Generative Zoosemiotics Perspective. *Linguistic Frontiers*, 9 (1), 1–15.

Zengiaro, N. & Jaramillo, T. J. (2025). Mythology and Zoosemiotics. Exploring Snake Narratives in Greek, Aztec, and Amazonian Cultures. *Biosemiotics*, 18, 1–15.

Zhao, W. X. et al. (2023). A Survey of Large Language Models. *arXiv*.



Parte II



Why Stories Enchant the Brain. Splendors and Miseries of an Extraordinary Human Faculty*

Francesco Ferretti

Full Professor at Roma Tre University, where he teaches Philosophy and Cognitive Sciences of Language and directs the Cosmic Lab. His research interests encompass the origin and evolution of language and the mind, as well as cognitive pragmatics, fields in which he conducts both theoretical and experimental studies.

francesco.ferretti@uniroma3.it

Humans stand out from other animals due to their pervasive ability to tell stories. Taking a persuasive rather than informative approach to human communication, I argue that humans invented stories to enhance their persuasive skills. From an evolutionary perspective, storytelling stems from narrative thought—cognitive systems enabling us to project ourselves into time, space, and other minds—which existed long before verbal expression. Stories exert a powerful persuasive effect because they captivate the human brain, offering a natural way to represent reality. However, this power is a double-edged sword: while it is adaptive and strengthens group identity for cooperation, it can easily blur the line between persuasion and manipulation, forming the basis of propaganda. Because the qualities that empower stories can be used both positively and maliciously, studying human narrative abilities is essential to developing defense strategies against manipulation. Ultimately, the persuasive power of stories represents both the glory and the misery of human communication.

* This article is based on a review of the topics presented at the conference "Mind matters: navigating neuroethics in the 21st century", 10th Conference of the Italian Society for Neuroethics (SINe) and 15th International Scientific Conference on Neuroethics, Rome, June 5, 2024.

— NARRATIVE
— PERSUASION
— COGNITION
— BRAIN
— MANIPULATION

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 93 — 105

1. *Homo Narrans*

In this article, I will argue that the ability to tell stories distinguishes human communication from that of other animals. More precisely, I will contend that humans invented stories to make communication more persuasive (Ferretti 2022; 2025). The first step to understanding what makes stories such effective persuasive tools is to ask where the human ability to tell stories comes from.

According to a widely held view, stories are the product of a social group's cultural tradition. In defining Homo narrans, Niles (1999) maintains that

> oral narrative is and for a long time has been the chief basis of culture itself. …it is chiefly through storytelling that people possess a past. It is through prized stories, often in a ritual context, that a complex religious dimension is added to life. It is through storytelling, very often, that people articulate their cherished values and… make possible a future that differs from what now exists.

This quote suggests that the ability to tell stories is a defining feature of the cultural history of Homo sapiens. Niles' thesis sits well within Bruner's cultural constructivist framework. In Bruner (1991) opinion, in fact, narrative is the hallmark of the representation of human experience, because the representational capacities made possible by narrative do not appear reducible to any other form of representation.

According to Bruner's cultural constructivist perspective, the stories through which people represent reality are the product of an internalization process in the individual mind of the 'external' stories told within the social group. Indeed, stories are stored in individuals' narrative practices within a community, becoming part of their psychological heritage through educational and learning processes. Bruner writes:

> When we enter human life, it is as if we walk on stage into a play whose enactment is already in progress – a play whose somewhat open plot determines what parts we may play and toward what denouements we may be heading. Others on stage already have a sense of what the play is about, enough of a sense to make negotiations with the newcomer possible. (…) I am proposing (…) that it is culture, not biology, that shapes human life and the human mind, that gives meaning to action by situating its underlying intentional states in an interpretative system (Bruner 1990, 34-35).

In line with Bruner's cultural constructivism, Hutto (2009) considers stories to be products of socio-culturally embedded traditions. The case of Folk Psychology (FP) exemplifies this perspective. When we interpret the behavior of others by attributing mental states to them, no particular mental process is at work. According to Hutto, FP is entirely governed by the Narrative Practice Hypothesis (NPH), which he considers to be a strong and radical hypothesis. In his view, in fact,

> The NPH must be distinguished from the softer claim that narrative engagements merely add finishing touches and refinement to pre-existing mindreading capacities (…). We don't just refine our pre-existing *FP* understanding by means of

> narrative engagements – on the contrary, we don't begin to exhibit *folk* psychology skills proper until we've had the right sorts of encounters with the right sorts of narratives. Children are not FP competent until they have mastered certain narrative skills. Engaging in narrative practice is the *source* of our FP understanding (Hutto 2009, 13).

An interesting point to note about this constructivist perspective is that it aligns well with Scalise Sugiyama's thesis that language is a prerequisite for storytelling. According to Scalise Sugiyama (2001), humans could not have used stories to exchange information prior to the emergence of language. The Language First Hypothesis states that our ancient ancestors must have already had language in order to begin telling stories. Specifically, verbal language is necessary for storytelling, as it distinguishes it from other forms of expression. The ability to tell stories is, in fact, «involves communication: fundamentally verbal (and, for the majority of its existence, oral)» (Scalise & Sugiyama 2005, 180) .

NPH captures an important facet of the role of storytelling in the human psyche. However, even if we accept that the cultural constructivist perspective seizes important aspects of human narrative abilities, it should be emphasized that this thesis is only valid in the context of *Homo sapiens'* current evolutionary stage. Considering stories as the basic prerequisite for explaining everything else, including individuals' mental lives and their communicative abilities, NPH says nothing about the evolutionary origin of stories. In my opinion this is serious negligence, since reflecting on the origins of stories provides a theoretical basis for understanding the persuasive power of stories and their evolutionary rationale. Without analyzing the evolutionary history of human narrative abilities (i.e. the formation of the narrative practices presupposed by cultural constructivism), it is impossible to explain why humans invented stories for communication.

At this point, I would like to offer an alternative perspective to that of the proponents of NPH on the role of stories in human communication and thought. My proposal begins with a reversal of the traditional intuitive view of the relationship between language and narrative.

2. Stories at The Origin of Language

Before continuing, I would like to make an important clarification. In this article, I do not intend to deny the role of social practices in human representational abilities. Nor do I intend to deny the role of language in these abilities. Rather, my intention is to re-evaluate the idea that cognitive systems hardwired into the brain are as important as the social narrative practices described by proponents of cultural constructivism. Without this re-evaluation, it is impossible to explain why stories are so persuasive or why humans invented them to improve communication. The first step in this direction is a reversal of perspective in the study of the relationship between language and narrative. My thesis, the Narrative First Hypothesis, is that language depends more on narrative than vice versa (Ferretti 2022; 2025).

The first piece of evidence to support my argument comes from a quote by Michael Corballis in *The Wandering Mind*. This quote highlights

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 93 – 105

a key point for my proposal: the idea that language is parasitic on storytelling rather than being an evolutionary precondition of it. According to Corballis (2015), in fact, language was invented specifically for storytelling: «if there is anything that defines our species as unique, it is the telling of stories and the invention of language as a means of doing so» (107). The important thing about this quote is not so much the idea that stories make us unique — a hypothesis supported by many other authors, as we have seen — but rather that language evolved from stories. In other words, stories are not the product of language; rather, they are the evolutionary platform that led to language's origin. Therefore, human narrative ability must be attributed to a factor other than language, at least in terms of its origins. So, how can we tell stories without language?

My proposal aligns well with models of language origin that view verbalization as a late stage in a process beginning with gesture. I hypothesize that language originated in pantomime (Ferretti 2022, 2024b; Ferretti et al. 2022). Through pantomimic storytelling, our ancestors developed a form of human communication long before verbal language emerged. McNeill (2005) offers the most fitting definition of the narrative basis of pantomime, describing it as a 'dumb show', or a sequence of gestures conveying a narrative, with a story to tell, produced without speech.

For the purposes of this paper, it is not important that McNeill is unwilling to consider pantomime as a form of protolanguage. Its definition remains important to us, regardless of what McNeill makes of it in terms of the origins of language. The idea that pantomime was the means of expression used by our archaic ancestors, which then gave rise to human language, is a widely accepted theory among scholars of the origin and evolution of language (Żywiczyński, et al. (eds), for a collection of recent studies). I specifically propose that pantomime was chosen by our ancient ancestors as a means of storytelling in the absence of verbal language (Ferretti 2022, 2024b, 2025; Ferretti et al., 2022). That said, the idea that pantomime could be considered the first way in which humans began to tell stories in the absence of language leaves open the question of why (the evolutionary reasons) humans chose to adopt pantomimic narrative form of communication from the beginning. The idea that humans have been distinguished from other animals by their ability to tell stories since the earliest stages of language leads us to consider what selective pressures led humans to invent this form of communication. As previously mentioned, my basic idea is that the ability to tell stories is closely related to the extraordinary persuasive power of stories. However, before proceeding further in this direction, it is important to clarify one point.

The idea that humans invented stories for the purpose of communication is challenged by the widely held view that stories are a cognitive rather than a communicative tool. Following the Aristotelian tradition, several authors argue that storytelling is a human invention designed to enhance cognitive abilities, such as the capacity for imaginative thinking. Corballis's thesis, which is also supported by other authors (Gottschall 2012; Brooks 1984), is that the ability to tell stories evolved in response to selective pressures favoring the enhancement of counterfactual reasoning. Contrary to the claims of proponents of the Aristotelian perspective, I would argue that stories emerged at the dawn of human communication in response to selective pressures related to enhancing communication

rather than knowledge. Stories enhance communication because all communication is oriented towards persuasion rather than the exchange of information, and stories are an extraordinarily effective persuasive tool. My position on this matter is primarily based on evolutionary reasoning: the reference to persuasion is a key factor in ensuring that human communication is highly similar to that of other animals. To understand the evolutionary reasons that led humans to communicate by telling stories, it is helpful to outline a more general framework of communication.

3. Persuasive Communication

Since the publication of the famous 1978 article by Dawkins and Krebs, 'Animal Signals: Information or Manipulation', it has become common practice to distinguish between models that view communication as an act of informing someone about something, and models that view communication as an act of trying to influence someone to act in a certain way. I believe that persuasive intentions primarily drive communication in both animals and humans. In line with this persuasive model, Kennedy (1992; 1998) proposed the thesis that rhetoric is a natural phenomenon found not only in humans of different cultures, but also in non-human animals. Considering rhetoric in these terms provides a framework of continuity, in which human characteristics cannot be viewed as a qualitative leap. Inspired by the Darwinian tradition, Kennedy (1992, 4) argues that «rhetoric is manifest in all animal life and existed long before the evolution of human beings».

For Kennedy, insisting on the existence of a 'deep natural rhetoric' means attributing persuasive abilities to animals that lack the capacity for language. It also suggests that rhetoric may have preceded human language in a framework that identifies persuasion as the primary function of language rather than communication. Indeed, as Kennedy (1998, 4-5) argues, «communication would not take place without a rhetorical impulse to drive it. There is no 'zero degree' of rhetoric in any utterance because there would be no utterance without a rhetorical impulse».

Among the basic evolutionary reasons for taking the persuasive route of animal communication is related to the fact that persuasion is a way to avoid reactive aggression: since fights can be costly to both aggressors and defenders, «any mechanism that can settle disputes without resorting to violence would seem to be of mutual benefit» (Parrish 2014, 40). The attempt to convince others, instead of provoking a conflict, can, therefore, be considered a highly adaptive strategy. As Kennedy (1998, 14) claims, «it seems clear that nature has encouraged the evolution of rhetorical communication as a substitute for physical encounters».

As I have already said, my idea is that all forms of communication have a persuasive foundation. To argue that animal communication has a rhetorical foundation represents an evolutionary basis for drawing a line of continuity with human communication. In the first book of the *Rhetoric*, Aristotle argues that humans tend by nature, as well as by culture, to persuade. The idea that the inclination to persuade is a fact of nature establishes a line of continuity between animal and human communication. That said, what makes human communication specific, and probably unique?

4. Narrative Persuasion

My proposal is that humans differ from other animals by inventing a specific way of communicating. Humans invented stories to enhance their ability to persuade and thus to enhance their ability to communicate. The thesis of the narrative foundation of language encompasses both elements of continuity, since persuasion is a feature of both human and animal communication, and elements of specificity, because stories are a typical and exclusive feature of human communication. To maintain that human communication is distinguished from that of other animals by the ability of humans to tell stories leads us to address two important theoretical issues: the first is to define stories by analyzing the properties that distinguish narrative from other forms of communication; the second addresses the analysis of what makes a story such a powerful tool of persuasion.

One way to characterize stories is to distinguish them from other persuasive strategies used by humans in communication, such as the ability to argue and counter-argue, which is typical of conversational exchanges (Ferretti, 2024a; Ferretti and Adornetti, 2021). The most effective way to define stories is to examine their distinctive features. There is a strong debate in narratology about which properties characterize stories. Despite different theoretical perspectives, there is agreement that character, plot and time are essential story properties. Different narratological perspectives are characterized by the different weights given to these properties. For example, Fludernik (1996) is a proponent of the radical hypothesis that character comprehension (i.e. the states of mind that govern a character's actions in a story) is all that is needed to understand a story's plot. In contrast to this emphasis on character, Brooks (1984, 5) explicitly references Aristotle to argue for the primacy of plot, stating that it characterizes any form of narrative as it would be impossible to move through the incidents, episodes, actions of a story without a principle of interconnection and intention. In his perspective, *plot* is not just an essential element of the structure of narrative but mostly a structural and dynamic constituent of human thought.

I completely agree with Brooks. When it comes to studying narrative, I am not interested in formal or structural analyses of texts (for example, in the manner of Barthes 2004). Rather, I am interested in narrative thinking and the processes by which humans construct narrative representations of reality. Referencing the narrative structure of representation is an important step towards my proposal in this paper. My thesis is that the narrative structure of storytelling originates from the narrative structure of thought (Ferretti 2021). In this context, the reference to the time factor is crucial.

The relationship between plot and temporality lies at the heart of Ricoeur's (1983) book *Time and Narrative*. According to Ricoeur, the narrative plot is inextricably linked to the temporal nature of human experience:

> One presupposition commands all the others, namely, that what is ultimately at stake in the case of the structural identity of the narrative function as well as in that of the truth claim of every narrative work, is the temporal character of human experience. The world unfolded by every narrative work is always a temporal

> world. Or, as will often be repeated in the course of this study: time becomes human time to the extent that it is organized after the manner of a narrative; narrative, in turn, is meaningful to the extent that it portrays the features of temporal experience (Ricoeur 1983, 3).

As is well known, Ricoeur's proposal aims at combining the theory of Augustine (time without narrative) and the theory of Aristotle (narrative without time), who both reflected on the two horns of the problem without proposing a concordance. The intertwining of time and plot, as outlined by Ricoeur, has been confirmed from a neuroscientific perspective. In this paper, I embrace a tradition very distant from that of Ricoeur. My way of analyzing the dependence of narrative communication on narrative thinking is to study the cognitive architectures that make story processing possible. Indeed, it is in the study of these systems that lies the explanation of why stories are such effective persuasion tools. In my opinion, stories enchant the brain because stories are the way that the cognitive systems wired in the brain enable a proto-narrative representation of reality: from this point of view, we can say that the brain recognizes stories as a kind of natural food.

5. What Cognitive Systems Underlie Narrative Processing?

A good starting point for thinking about this question is offered by Corballis (2015, 111), who states that «the same constructive process that allows us to reconstruct the past and construct possible futures also allows us to invent stories». Combining Corballis's suggestion on cognitive architecture with the importance attributed to the time factor in constructing a plot supported by Ricoeur, my first look is at the cognitive devices that allow humans to navigate through time.

It is well known that people can travel through time in their minds. They can relive past events and envisage future scenarios by projecting themselves backwards and forwards in time (Suddendorf & Corballis 1997, 2007). It is important to emphasize that reference to Mental Time Travel (MTT) does not simply reflect the ability to extract particular meanings or knowledge from memory; rather, it involves actively retrieving and reconstructing past and potential experiences. But how can we provide evidence that this cognitive mechanism is one of the systems underlying the comprehension and production of narrative plot? To answer this question, we designed an experimental protocol to analyze the narrative skills of children with ASD at the Cosmic Lab of Roma Tre University together with Bambino Gesù Hospital in Rome. Previous independent studies had shown that some children with ASD have difficulties with both temporal projection and narrative skills. Our studies were based on the hypothesis that these difficulties are correlated. Specifically, we hypothesized that temporal projection difficulties would underlie global coherence difficulties in story production. The results of our research supported this hypothesis. Children with ASD who had greater difficulty with temporal projection also had greater difficulty with global coherence in stories (Ferretti et al. 2018; Marini et al. 2019). Our experiments suggest that MTT is one of the cognitive devices involved in processing story plots. A case similar to systems that enable navigation through time during the

processing of global coherence is that of systems that guide navigation through space (Mental Space Travel, or MST), as demonstrated by the studies of Race and his colleagues (2015) on patients with hippocampal amnesia.

Additionally, it is important to note that MTT and MST are not the only cognitive devices involved in processing stories. Mindreading is another cognitive device that plays a significant role in this process (Mar 2004; 2011). A large body of literature suggests that fictional relationships with story characters activate the same mindreading mechanisms involved in social relationships with real people. For example, Yuan and colleagues (2018) have demonstrated that mindreading systems are activated by visual stories even when there are no depicted characters, proving the character-driven nature of story creation, as the authors argue.

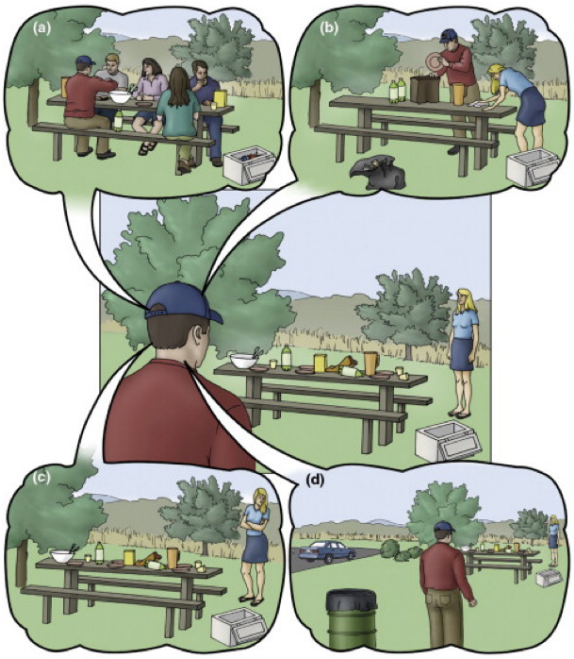
From the considerations so far, it can be concluded that narrative processing involves a macrosystem comprising three different cognitive systems: those governing projection in space (MST) and time (MTT), and the system governing projection into the minds of others. These three systems (the Triadic System of Grounding and Projection) underlie the construction of 'distended' forms of representation (Ferretti 2022, 2025). It is the distended nature of narrative that distinguishes it from other forms of representation, which are governed by global coherence.

In a highly influential paper, Buckner and Carroll (2007) argued that the devices responsible for mindreading, spatial navigation, and temporal navigation emerge at a similar age and share a common functional anatomy. This anatomy includes frontal and medial temporal areas that are traditionally associated with planning, episodic memory, and default (passive) cognitive states. In my view, the brain's self-projection network, which enables the perceiver to extend the perceived scene in terms of time, space and the social world, must be considered the macrosystem responsible for the proto-narrative nature of human representations of experience. [FIG. 1]

The concerns raised so far lead us to consider two things. Firstly, as well as referring to the narrative practices of the social group, it is plausible to argue that individuals' narrative abilities depend on a cognitive macrosystem ('the narrative brain') capable of constructing 'proto-narrative' representations of reality. Secondly, it is reasonable to argue that stories' extraordinary persuasive power primarily stems from their reliance on narrative processing systems in the brain. For the remainder of this paper, I will examine some traits of the persuasive power of stories and demonstrate that the splendors and miseries of storytelling are closely related to the activities of the 'narrative brain'.

6. The Persuasive Power of Emotions

One of the fundamental features of devices that render stories persuasive is their ability to trigger emotions (Bilandzic & Busselle, 2013). For instance, numerous studies have demonstrated the positive impact of stories on social cognition, particularly with regard to empathy (Castano 2024; Castano et al. 2024). In our research at CosmicLab at the University of Rome Tre, we examined the relationship between storytelling and emotions in more detail by analyzing the role of autobiographical stories in



[FIG. 1] Forms of self-projection: a) recollection of the past – episodic memory; b) anticipation of a future scenario – episodic future thinking; c) simulation of another person's perspective – theory of mind; d) simulation of a mental map of the environment – mental space travel (figure taken from Buckner, Carroll 2007, 51).

modifying emotional experiences, particularly empathic processes related to emotion recognition (Altavilla et al. 2022). For this study, we used EEG to measure brain activity while a group of 29 undergraduate students recognized emotional expressions. The students were then divided into two groups: an introspective group and a control group. The control group was asked to keep a diary of daily events for one week, while the introspective group was asked to keep a diary of daily emotions, thoughts and bodily sensations experienced during the week. After seven days, all participants underwent another EEG assessment of emotional expression recognition. The results revealed changes in neural activity associated with recognizing emotional expressions in the Introspection Group (IG) only. Specifically, the introspection group showed a greater P300 amplitude at T1 than at T0 in the left and right centre-frontal montage. This suggests that introspective storytelling can modulate brain activity during a social cognition task (i.e. an emotion recognition task). This finding is consistent with numerous studies showing that storytelling (e.g. reading novels) has a positive effect on empathy and, consequently, on social relationships.

Stories are a powerful tool for building and reinforcing group identity, consolidating the prestige of leaders and persuading audiences to cooperate in common activities due to their persuasive power. The emotional connection to stories also suggests that the origin of narrative can be considered in the context of sexual selection as well as natural selection (Miller 2000). For example, stories play a significant role in transmitting religious beliefs and thus serve as a powerful means of social cohesion in the context of sexual selection (Ferretti & Adornetti 2022). It could be argued that stories are to humans what the peacock's tail is to that animal; the ability to tell stories is what makes human courtship the defining

feature of sexual selection (Ferretti 2022). This is enough to recognize the importance of stories in human activities. It is time to conclude our analysis of the dark side of the moon.

7. The Poisonous Principle of Stories

In his book *The Story Paradox* (2021), Jonathan Gottschall shows that, in addition to their undeniable virtues, stories have a dangerous negative power due to their persuasive nature. They are both a source of social cohesion within a group and a source of radical division between groups. Stories can act as magnets, drawing people into cohesive tribes. However, they can also be the thing that most alienates one tribe from another, like magnets turned inside out. For this and other reasons, Gottschall believes «that storytelling is the "poisonous principle" of humanity. Toxic principles are often essential substances for human survival, but at the same time they are deadly» (23).

The poisonous principle of stories depends largely on the fact that stories conceal the persuasive intent of the storyteller. The persuasive power of narrative depends on the fact that stories are a means of inhibiting the listener's defensive strategies, because stories have the power to *fly under the radar*, that is, stories allow the storyteller to achieve her persuasive goal by hiding her persuasive intent (Dal Cin, Zanna, Fong 2004). In this way, stories can bypass the audience's natural defensive reaction. This is a way of achieving persuasion that relies on mechanisms of transportation and engagement. According to the Transportation Imagery Model proposed by Green and Brock (2000, 703), in fact «the reader loses access to some real-world facts in favor of accepting the narrative world that the author has created». In this way, while the reader is immersed in the story, she may be less aware of the real-world facts and this undermines her ability to counter-argue, making her more willing to accept the story claims and thus more easily swayed.

Unlike argumentative persuasion, where the intent to persuade is explicit and the audience can use defensive strategies governed by epistemic vigilance mechanisms (Sperber et al. 2010; Sperber & Mercier 2017), much of the persuasive power of stories comes from the ability to conceal intent. Storytelling is a persuasive strategy that leaves the listener unable to defend themselves because stories prevent the activation of epistemic vigilance mechanisms, thus stopping the listener from enacting defensive strategies.

Given the close relationship between storytelling and empathy that we have discussed, it is plausible to imagine that the negative effects of stories might also spill over into empathy itself. Paul Bloom devoted his book *Against Empathy* to the negative aspects of empathy:

> Empathy has its merits. It can be a great source of pleasure, involved in art and fiction and sports, and it can be a valuable aspect of intimate relationships. And it can sometimes spark us to do good. But on the whole, it's a poor moral guide. It grounds foolish judgments and often motivates indifference and cruelty. It can lead to irrational and unfair political decisions, it can corrode certain important relationships, such as between a doctor and a patient, and make us worse at being friends, parents, husbands, and wives (Bloom 2017).

The case of empathy is a good example of what Gottshall calls *The Story Paradox*. As we have seen, empathy is undoubtedly a fundamental factor in the social cohesion generated by emotional contagion. At the same time, however, while the emotional contagion of stories strengthens relationships within the social group, it also strengthens divisions with those outside the social group. Empathy distorts our moral judgments in the same way that prejudice does. In other words, because it is much easier to empathize with in-group members than with out-group members, the main effect of empathy generated by stories may not be to blur the lines between "us" and "them," but to strongly reinforce them. (Gottshall 2021, 139-140).

8. Conclusions

According to my hypothesis, stories are tools invented by humans to enhance the persuasive nature already present in animal communication. The persuasive power of stories is important in group activities, as it promotes social identity and cooperation; at the same time, however, the ability to conceal persuasive intent that storytelling provides makes stories a dangerous communication tool. The crux of the problem is that the very same features that make stories useful tools of persuasion also make them dangerous. Therefore, the splendors and miseries of storytelling simultaneously represent the pinnacle of socialization between individuals and the means by which the most violent conflicts are generated.

The question of how to harness the positive aspects of stories while avoiding the issues arising from concealing the speaker's persuasive intent is an important one, but it goes beyond the scope of this paper.

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 93 – 105

Bibliography

Adornetti, I. & Ferretti, F. (eds). (2024). *Introducing Evolutionary Pragmatics*. Routledge.

Altavilla, D. et al. (2022). Introspective Self-Narrative Modulates the Neuronal Response During the Emphatic Process: An Event-Related Potentials (ERPs) Study. *Experimental Brain Research*, 240 (10), 2725-2738.

Barthes, R. (2004). Introduction to The Structural Analysis of Narratives (65-116). In M. Bal (ed.) *Narrative theory: Critical Concepts in Literary and Cultural Studies*. Routledge.

Bilandzic, H. & Busselle, R. (2013). Narrative Persuasion (200-219). In J.P. Dillard & L. Shen (eds), *The SAGE Handbook of Persuasion: Developments in Theory and Practice*. SAGE Publications.

Bloom, P. (2017). *Against Empathy: The Case For Rational Compassion*. Random House.

Brooks, P. (1984). *Reading For The Plot*. Harvard University Press.

Bruner, J. (1990). Culture And Human Development: A New Look. *Human Development*, 33 (6), 344-355.

Bruner, J. (1991). The Narrative Construction Of Reality. *Critical Inquiry*, 18 (1), 1-21.

Buckner, R.L. & Carroll, D.C. (2007). Self-Projection And The Brain. *Trends in Cognitive Sciences*, 11 (2), 49-57.

Castano, E. (2024). Less Is More: How The Language Of Fiction Fosters Emotion Recognition. *Emotion Review*, 16 (2), 73-83.

Castano, E. et al. (2024). On The Complexity Of Literary And Popular Fiction. *Empirical Studies of the Arts*, 42 (1), 281-300.

Corballis, M.C. (2015). *The Wandering Mind: What The Brain Does When You're Not Looking*. University of Chicago Press.

Dal Cin, S. et al. (2004). Narrative Persuasion And Overcoming Resistance (175-192). In *Resistance And Persuasion*. Psychology Press.

Dawkins, R. & Krebs, J.R. (1978). Animal Signals: Information Or Manipulation (282-309). In J. R. Krebs & N. B. Davies (eds), *Behavioural Ecology: An Evolutionary Approach*. Blackwell.

Ferretti, F. (2021). The Narrative Origins Of Language. In N. Gontier et al. (eds), *The Oxford Handbook Of Human Symbolic Evolution*. Oxford University Press.

Ferretti, F. (2022). *Narrative Persuasion. A Cognitive Perspective On Language Evolution* (VII). Springer Nature.

Ferretti, F. (2024a). Agonistic Conversation: A Cognitive-Interactive Perspective On The Origin Of Grammar (124-149). In I. Adornetti & F. Ferretti (eds), *Introducing Evolutionary Pragmatics*. Routledge.

Ferretti, F. (2024b). Narrative And Pantomime At The Origin Of Language (78-99). In P. Żywiczyński et al. (eds), *Perspectives On Pantomime*. John Benjamins.

Ferretti, F. (2025). Stories Come First. The Origins Of Human Communication From A Naturalised Perspective Of Language. *Rivista internazionale di Filosofia e Psicologia*, 16 (1), 1-16.

Ferretti, F. & Adornetti, I. (2021). Persuasive Conversation As A New Form Of Communication In Homo Sapiens. *Philosophical Transactions of the Royal Society B*, 376 (1824), 1-9.

Ferretti, F. & Adornetti, I. (2022). Dual Inheritance Theory: Religion, Narrative, and Selection. In *The Routledge Handbook of Evolutionary Approaches to Religion* (243-260). Routledge.

Ferretti, F. et al. (2018). Time And Narrative: An Investigation Of Storytelling Abilities In Children With Autism Spectrum Disorder. *Frontiers in Psychology*, 9, 944.

Ferretti, F. et al. (2022). Narrative Pantomime: A Protolanguage For Persuasive Communication. *Lingua*, 271.

Fludernik, M. (1996). *Towards a 'Natural Narratology'*. Routledge.

Gottschall, J. (2012). *The Storytelling Animal: How Stories Make Us Human*. Houghton Mifflin Harcourt.

Gottschall, J. (2021). *The Story Paradox: How Our Love Of Storytelling Builds Societies And Tears Them Down*. Hachette UK.

Green, M.C. & Brock, T.C. (2000). The Role Of Transportation In The Persuasiveness Of Public Narratives. *Journal of Personality and Social Psychology*, 79, 701-721.

Hutto, D. (2009). Folk Psychology As Narrative Practice. *Journal of Consciousness Studies*, 16 (6-7), 9-39.

Kennedy, G.A. (1992). A Hoot In The Dark: The Evolution Of General Rhetoric. *Philosophy & Rhetoric*, 25 (1), 1-21.

Kennedy, G.A. (1998). *Comparative Rhetoric: An Historical And Cross-Cultural Introduction*. Oxford University Press.

Mar, R. A. (2004). The Neuropsychology Of Narrative: Story Comprehension, Story Production And Their Interrelation. *Neuropsychologia*, 42 (10), 1414-1434.

Mar, R.A. (2011). The Neural Bases Of Social Cognition And Story Comprehension. *Annual Review of Psychology*, 62 (1), 103-134.

Marini, A. et al. (2019). Episodic Future Thinking And Narrative Discourse Generation In Children With Autism Spectrum Disorders. *Journal of Neurolinguistics*, 49, 178-188.

McNeill, D. (2005). Introduction. In D. McNeill (ed.), *Language and gesture* (1-12). Cambridge University Press.

Mercier, H. & Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

Miller, G. (2000). *The Mating Mind: How Sexual Choice Shaped The Evolution Of Human Nature*. Anchor.

Niles, J.D. (1999). *Homo Narrans: The Poetics And Anthropology Of Oral Literature*. University of Pennsylvania Press.

Parrish, A.C. (2013). The (Instinctual) Art Of Persuasion. *Evolutionary Review*, 4 (1), 57-66.

Parrish, A.C. (2014). *Adaptive Rhetoric. Evolution, Culture, and The Art of Persuasion*. Routledge.

Race, E. et al. (2015). Sharing Mental Simulations And Stories: Hippocampal Contributions To Discourse Integration. *Cortex*, 63, 271-281.

Ricoeur, P. (1983). *Temps et récit* (I). Éditions du Seuil.

Sperber, D. et al. (2010). Epistemic Vigilance. *Mind & Language*, 25 (4), 359-393.

Suddendorf, T. & Corballis, M.C. (1997). Mental Time Travel And The Evolution Of Human Mind. *Genetic, Social, and General Psychology Monographs*, 123 (2), 133-167.

Suddendorf, T. & Corballis, M.C. (2007). The Evolution Of Foresight: What Is Mental Time Travel, And Is It Unique To Humans? *Behavioral & Brain Sciences*, 30 (3), 299-313.

Sugiyama, M.S. (2001). Food, foragers, and folklore: The role of narrative in human subsistence. *Evolution and Human Behavior*, 22 (4), 221-240.

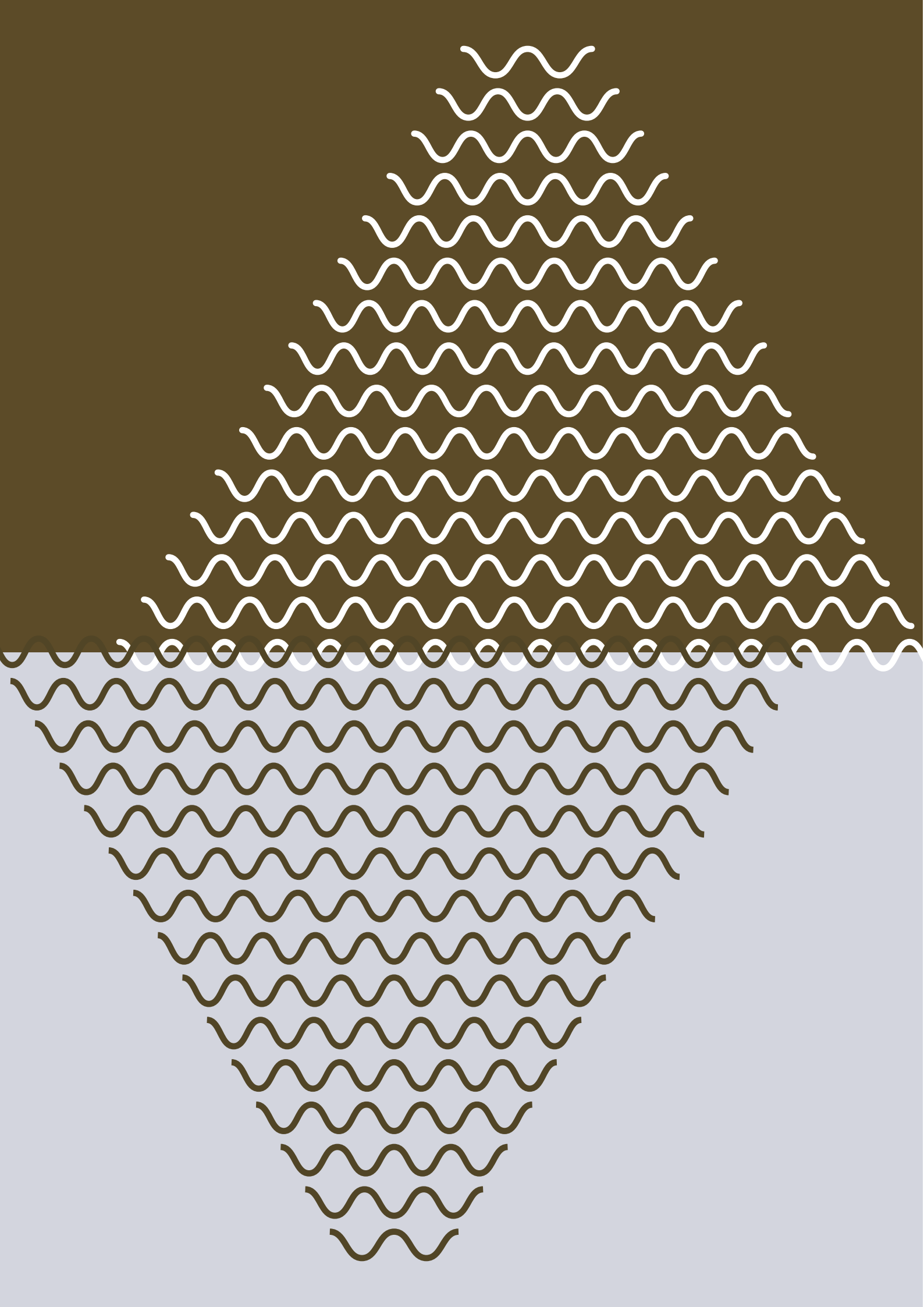
Sugiyama, M.S. (2005). Reverse-Engineering Narrative: Evidence Of Special Design (177-196). In J. Gottschall, D.S. Wilson (eds.) (2006). *The Literary Animal: Evolution And The Nature Of Narrative*. Northwestern University Press.

Yuan, Y. et al. (2018). Storytelling Is Intrinsically Mentalistic: A Functional Magnetic Resonance Imaging Study Of Narrative Production Across Modalities. *Journal of Cognitive Neuroscience*, 30 (9), 1298-1314.

Żywiczyński, P. et al. (eds). (2024). *Perspectives on Pantomime*. John Benjamins.

Żywiczyński, P. et al. (2018). Defining Pantomime for Language Evolution Research. *Topoi*, 37 (2), 307-318.

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 93 — 105



Alice in Worderland.
The Language(s) of Architecture in the Age of AI

Carlo Deregibus

Architect and PhD, is a tenure-track Assistant Professor of Architecture and Urban Design at the Politecnico di Torino. He is the Architectural Advisor of the atheneum and design manager of the Masterplan Team. He was awarded the VI International Hangai Prize and the NIB Prize 2018, selecting the top 10 emerging architectural firms in Italy.

carlo.deregibus@polito.it

The language of architecture has always been multidimensional, oscillating between visual and verbal forms, with its evolution never linear. Over time, graphical representation has been transformed by increasingly intuitive digital interfaces, while verbal discourse has shifted from communication to conceptual legitimation and, lately, has dissolved into diffuse storytelling. Once again, Artificial Intelligence unsettles this fragile balance by forging a new relationship between the visual and the verbal, thanks to a (seemingly) unprecedented level of accessibility. Even more, it exposes the ontological linguistic nature of architectural practice itself by revealing a latent systemic interface that can be navigated through keywords and iterative translations – an operative spacetime we'll call *Worderland*. Within it, architects operate on a super-language that binds multiple special languages while remaining irreducible to any of them.

To investigate this hypothesis, a first historical and theoretical section identifies a set of keywords that permeate architecture and design. Through two experimental case studies, the second part examines how architectural language becomes performative, inter-semiotic and interactive. Then, the third part points out a series of takeaways by recalling the keywords of the first one. Eventually, the paper reframes the nature of architectural language in the age of artificial intelligence: no longer as a symbolic system or representational code but as an effective, inter-semiotic, and interactive device.

— LANGUAGE IN ARCHITECTURE
— MEANING IN ARCHITECTURE
— ARTIFICIAL INTELLIGENCE
— ARCHITECTURAL DESIGN
— ARCHITECTURAL CONCEPT
— PARAMETRIC DESIGN
— KEYWORDS
— DESIGN AGENCY

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 107 — 124

1. Down the Rabbit-Hole

> Down, down, down. Would the fall NEVER come to an end!
> (Carroll 1865, 4).

Over time, language has taken on very different roles in architecture, undergoing profound transformations and assuming narrative, explanatory, generative, functional and promotional functions. To grasp the issue in its contemporary configuration, we will adopt a systemic and exploratory approach, moving recursively between historical reconstruction, theoretical reflection and experimental practice, allowing each to continuously inform the others. We will first retrace some critical stages in the evolution of language in architecture, from which a set of theoretical coordinates progressively emerges. We will then read two independent experiments through those coordinates, allowing the recurring patterns that emerge – should they decide to come out of the Rabbit-Hole – to further articulate and refine the theoretical framework.

De architectura by Vitruvius, generally regarded as the conceptual foundation of Western architectural history, was written in a form of vulgar Latin, specialist by virtue of its many technical and colloquial terms (what today we would call site jargon), and unadorned in relation to its dedicatee (Romano 1997). Indeed, one of the first and most important examples of the uneasy coexistence between *communicative* language and *special* languages, or subcodes «characterised by a specialised lexicon, in relation to particular extralinguistic domains and the corresponding spheres of meaning» (Berruto 2020, 178). Without engaging with the nuances that have shaped recent linguistic debate, this discrepancy lies at the heart of Renaissance treatises – with Bernardino Baldi's *De verborum vitruvianorum significatione* (Baldi 1612) explicitly foregrounding its significance. Over time, the special lexicon became *specialist*, contributing decisively to the recognisability of groups, guilds, trades and, ultimately, professions.

For five centuries, the special lexicon was refined and expanded in scope and typology. Architects began to write essays, articles and volumes, developing a critical stance beyond the strictly constructive and manual sphere. The transition between the eighteenth and nineteenth centuries is particularly significant (Kruft 1987, 1988): the architect assumes an artistic and intellectual dimension increasingly separated from that of the builder (Biraghi 2019). The notion of the architectural conception emerges (now evolved into the architectural "concept"). Manifestos are written, and battles erupt between opposing camps – with the emblematic case of London's Crystal Palace undermining academic certainties. Increasingly forceful revolutions culminate in the Modern Movement: stylistic upheavals that, however, do not alter the idealistic framework sustaining the relation between words and works (Collins 1965).

This remains the case until the post-war crisis, when the failure of the Modern reveals architectural thought's increasing detachment from the meanings of its time. Architecture's "language" is thus redefined: not as a way to describe architectural actions, but as a conceptual and organisational matrix. From Giovanni Klaus Koenig's studies (1964) to Umberto Eco's contribution, semiotics distinguishes syntactic and semantic codes and, further, elements and typological genres (Eco 1971). The structuralist

explosion is brief yet intense, but its influence persists, with continued references to architectural languages tied to more or less defined canons (Olmo 2024). [1] This necessity arises because the end of the Modern marks the impossibility of fixed meanings and intentions.

A period of experimentation follows, seeking to reconstruct a stable ground of exchange between forms and their meanings. This occurs, partly, through linguistic experimentation – Postmodernism proper, which works precisely on the discrepancy between syntax and semantics (Venturi 1966); partly, by viewing architecture as a system of communication rather than of artistic production. The lexicon and concepts of Ferdinand de Saussure and Roland Barthes are taken up by De Fusco (1967), who attempts to move beyond modernist autonomy by elaborating on the signifier–signified relationship and on the associative relations among elements – indeed, underestimating their potentially infinite extension. Peter Eisenman (1963) draws on Noam Chomsky's generative–transformational linguistics to define a combinatorial logic of the architectural vocabulary (volume, mass, surface and movement) through compositional syntax, only to discover that pure formal manipulation requires further legitimation.

For fifty years, the axiom that linguistic construction legitimises architecture holds, for lack of alternatives. Meaning, syntactically constructed, becomes increasingly self-referential: many languages are created, yet not *language*. Then something changes. The texts that crowded post-war journals were essays of rare complexity, with extensive graphic apparatuses and few photographs: authorial writings that generated a genre of their own (Cartocci 2022). Today, extensive photographic spreads are paired with schematic drawings and short, non-argumentative texts: a visualist turn that subordinates conceptions to commercial display. These symptoms reflect broader transformations and lead to a decline in the quality of architectural writing, alongside its quantitative increase and the growing divide between academia and practice (Wiseman 2014).

At the same time, the generative use of language – that is, the generative quality of narrativity applied to project morphogenesis – also changes. If designing means creating syntactic systems, architecture can be transformed *tout court* into a radical communicative system, in which many (or rather, infinite) parameters contribute through a computational, evolutionary, and evaluative process (Schumacher 2012). Architectural textuality absorbs contemporary themes and promotes new, essential *selling words* (Iglesias-Vázquez 2025) – e.g., sustainability, co-design, resilience, heritage. Their communicative and commercial exploitation relies on a different language, with storytelling experts working alongside designers during participatory processes (Shaw 2022). Too often, this is the contemporary role of those who were once critics.

[1] This position also relates to a time when figures such as Manfredo Tafuri, Asor Rosa and Massimo Cacciari articulated a Marxist interpretation of architecture as superstructure (see Mometti 2012).

2. Taking a breath

> « It's enough to drive one crazy! » (Carroll 1865, 80).

Given this evolution, defining language in architecture appears quite difficult. The theme has undergone so many permutations, often addressed

with great methodological rigour, that any univocal definition would require an ideological stance – or, conversely, a retreat into vagueness. Yet traces of *all* these permutations persist in contemporary practices.

Across architectural practices, the relaxed interplay between *common language* and *special languages* continues. The special language is no longer primarily that of the construction site but rather that of architectural theories and poetics. In any case, it persists, often nourished by terms borrowed from philosophy, geography and social science, deliberately made architecturally special.

The *intellectual dimension* has become a necessity rather than a mere aspiration. To distinguish themselves from other professionals (that is, engineers), framed as problem-solvers, architects identify as problem-setters, emphasising the socio–technical–artistic nature of their work rather than its purely technical aspects – a move that unfolds in highly differentiated and often conflicting ways.

The notion of *concept* has become ubiquitous, perilously close to a misplaced concreteness of Whiteheadian memory. Before or after each creative phase, a concept is required to summarise what was done and orient future developments. A project without a concept would be weak, unclear, and therefore fragile, whereas a project with a strong concept would have prophetic force.

Semiotics as a foundation for architectural design has perhaps waned, both academically and in practice, yet it remains influential in urban morphological analysis. More importantly, it continues to permeate how we refer to historical periods through styles and stylistic features, and how we work on levels of meaning by attributing value to the logical-syntactic coherence of design elements.

Even more persistent are *associative relations*, which expand the signifier-signified connection. This is partly because they exist regardless of architects' intentions; partly because, in a prophetic *hybris* full of consequences (as in the well-known *Simpsons* parody of the late Frank Gehry), concepts rely precisely on the existence and strength of associative relations.

Evaluative approaches now saturate design processes, thanks to easily accessible parametric tools and due to transparency afflatus. Indeed, any architectural choice, whether driven by regulatory necessity or by the anxiety of legitimation, now requires the construction of a system that justifies its inevitability and demonstrates, beyond doubt, that every alternative has been considered.

Inevitably, architecture's new *selling words*, carefully assembled into compelling storytelling, reshape and reposition architects within society. The success of these narratives elevates starchitects to the status of gurus. Ordinary architects do the same on a smaller scale, attempting to harness the evocative and commercial power of words to embellish their proposals.

This shift leads to the flattening and normalisation of writing, towards mere description anchored in authorial (hence, supposedly ingenious) concepts, or towards easily marketable messianic approaches. Languages proliferate, but become less special, in a gradual – and seemingly intentional – detachment from the projectual matrix of architectural practice.

All these key points (*special languages; intellectual dimension; concept; semiotics; associative relations; evaluative mechanisms; selling words; normalisation of writing*) remain present, within a fragmented system rarely described at a systemic level. As a result, the generative and recursive dimension of such practices and technologies in architectural design – that is, the changes in architectural design and architecture, as the former is practised and the latter is experienced, produced, and consumed – remains dramatically underestimated. Conceptualisations just follow events, usually by merely registering them. On the one hand, reactionary positions, driven by a nostalgic defence of established roles and institutional rents, still depict architects as solitary geniuses tracing lines in the silence of inspiration. On the other hand, progressive positions, indeed neoliberal in their unaware adherence to the financialisation of building, seek to erase architecture's artistic dimension by recasting architects as process managers. Opposed as they may appear, both stem from the same uncertainty in the face of increasingly rapid and difficult-to-control changes.

3. The advent of AI

> « Keep your temper, » said the Caterpillar (Carroll 1865, 61).

The recent advent of AI has had disruptive effects. Or rather, it would be more accurate to say that its easy accessibility to a broad public – professionals and the uninformed alike – has placed AI, already used in many fields, at the centre of recent debate. Unsurprisingly, discussion has focused either on intrinsic risks from a labour perspective – under the rallying cry that "AI will replace architects" – or on its (supposed) infinite power, sometimes in strikingly naïve ways – for instance, echoing Philip K. Dick's *Vulcan's Hammer*, AI has been celebrated as a supposedly impartial agent, capable of liberating architectural practice from vested interests (Salingaros 2025). Meanwhile, however, the carefully commercial modes through which AI has insinuated itself into practice – simplifying workflows and offering advantages that were first free and then gradually monetised – have already ensured its deep entrenchment (Li et al. 2025).

In most discussions, the *recursive* value of AI use – that is, the way in which practising architecture with AI modifies architecture and its very conception – is ignored or at least underestimated. Debate favours a crudely instrumental view, in which AI either "replaces" the architect, simplifies his work or expands his possibilities. Yet if the interfaces of traditional software, the materials used for models, or the organisation of the studio, play a decisive role in shaping architectural thinking (Yaneva 2009; Deamer 2020), then an instrument as pervasive and accessible as AI must all the more radically transform practices. There *will* be a co-evolutionary relationship, whereby architects' agency itself will evolve.

Unsurprisingly, these processes strongly affect the notion of language in architecture. A clear example is the attempt to translate Christopher Alexander's *A Pattern Language* (1977) into LLMs (Large Language Models). The 253 patterns of the original study constituted, precisely in the semiotic sense outlined above, a hierarchical system encompassing project grammar at all scales, from the urban to the building. By

foregrounding structure behind images and forms, *A Pattern Language* revealed the repetition of a vast number of configurations within which designers typically moved intuitively and unconsciously. Years ago, the potential development of this approach through LLMs was already anticipated (Iba 2016). A very recent study pushes this idea further, once again renewing the semiotic notion that structure sustains design. Recognising and selecting different patterns generates sequences of alternatives ranging from over 30,000 (with just two patterns) to more than 2.5 million (with three) and up to 1.5 billion (with four) – far too many to be handled through a manual comparative and evaluative approach. Conceptually, however, AI could do so (Postle & Salingaros 2025). This approach, however, ultimately shifts the issue of architectural language to another plane: interaction with the tool, which inevitably takes place through chatbots.

A different application, drawing on the immense image repository of the web – inconceivable without the vast, involuntary contributions of billions of people and their social interactions (Slessor 2020) – is that of text-to-image or image-to-image generators (always mediated by descriptive text). That is, generative engines (e.g., Flux.1, Nanobanana, GPTImage) integrated into commercial software (or, more rarely, free or open-source platforms) with intuitive interfaces mediated by chatbots (DALL·E, Midjourney, Krea, RendairAI, LeonardoAI, and many others), capable of producing architectural images from textual descriptions, sometimes supported by reference images. Initially available only in native languages (English or Chinese for the most part), interfaces now support hundreds of languages – AI-translated, obviously. Old lexicon, new language: one based on a different, and far more radical, use of *keywords* in architectural design (Deregibus 2025).

Something unexpected happens, with overwhelming consequences. Never before has it been possible to generate literally never-ending variants of architectural images from a text. Never has common language been used not to describe drawings or structure concepts, but to directly and explicitly originate forms. Yes, forms could be generated through Lisp-based scripts (such as RhinoScript or AutoCAD Lisp), but this required mastery of highly specialised programming languages, confining formal experimentation to necessarily marginal domains (Balmond 2007). Thus, while the quality of writing declines on the one hand, an unprecedented language emerges on the other. Not a verbal language evolved (often metaphorically) from others, nor the graphical language of academia, but an intermediate systemic interface made of keywords. An *Architectural Worderland*, in which architects must once again learn the ontological geography of their own language.

4. Exploring Worderland

> « What fun! » [...] «What is the fun? » (Carroll 1865, 139).

Enthusiasm for this potential is understandable. What matters here, however, is whether and how this interaction – this new language – modifies architecture itself. To explore its impact on design, two experiments with opposite approaches are presented. The first, following a "many-to-one" approach, tests a group of students using the same tool, investigating how

different designers progressively refine a common generative language. The second, "one-to-many", tests a single architectural text across different tools, examining how different AI systems interpret the same linguistic input. Together, they further articulate the themes previously identified. [2]

[2] For reasons of space, other relevant and equally interesting experiments are not reported here (see, for instance, Pupi et al. 2025).

4.1– Many to one

> « Oh, there's no use in talking to him » said Alice desperately: « he's perfectly idiotic! » (Carroll 1865, 80).

The first experiment was conducted over two years within an architectural design studio at the Politecnico di Torino, held by Prof. Carlo Deregibus with Enrico Pupi. Approximately 90 students from 15 countries worldwide participated, all in their final year and therefore close to entering professional practice. The test required them to use precisely twenty iterations (corresponding to the free credits of the selected generation engine, Flux.1) to produce an image of a previous design project as precise as possible. With a clear objective in mind, one might assume that providing a verbal description of the building would not be particularly difficult. Yet all students – despite many already being familiar with generative tools – encountered enormous difficulties.

Here is the difference between one of the initial prompts and its version after twenty iterations (author: Isabela Caicedo):

> Iteration 1:
> nursery school made of smooth compressed earth blocks and bamboo structure, corrugated metal roof with thin overhangs, wooden doors and woven screens, calm natural light, realistic textures, warm and simple rural setting

> Iteration 20:
> single-story nursery school, 200 m² humanitarian architecture designed for the Kaira Looro competition, located in Sub-Saharan Africa, composed of two softly curved pavilions embracing a central courtyard used as a playground and communal gathering space. The buildings are constructed with smooth modular compressed earth blocks in warm sandy tones, supported by a lightweight bamboo structure with slender round columns and beams precisely joined with visible connections. The corrugated galvanized metal roof, painted white, follows a continuous and delicate curve with thin overhangs and exposed rafters, creating shade and encouraging natural ventilation. Facades feature wooden doors and woven bamboo shutters that filter light and maintain privacy. The courtyard is covered in dry grass with patches of low native vegetation and scattered trees casting soft shadows.

In general, prompts became longer, increasing from 65% to 400%, with an average of around 120%. As a consequence, detail augmented, primarily referring to architectural elements: architecture is decomposed and reconstructed through addition, with few poetic descriptions (in the example: the "gentle curve"), according to an objectifying logic. A comparison between some images at Iteration 1 and Iteration 20 [FIGS. 1, 2] clearly shows the effectiveness of this objectifying strategy.
It is worth noting that, in the report students were required to produce,

[FIG. 1] Iteration 1 by Isabela Caicedo (a), Xiaoming Ding (b), Maryam Shabani (c)

(a)



(b)



(c)



[FIG. 2] Iteration 20 by Isabela Caicedo (a), Xiaoming Ding (b), Maryam Shabani (c)

(a)



(b)



(c)



they *all* first described the project verbally, explaining its intentions and concepts. Then, invariably, they formulated an initial prompt in a *different* form. In other words, they first explained the project using an "architectural" language – the special language employed to present projects to professors or, eventually, to clients – and then (intuitively?) adopted a different, more schematic special language to instruct the tool. Thus, without even realising it, they fell down the Rabbit-Hole, entering Worderland: a domain governed by an esoteric language, much like priests before an oracle in a temple – or the protagonists of Philip K. Dick's already mentioned *Vulcan's Hammer*.

4.2 – One to many

> She generally gave herself very good advice, (though she very seldom followed it) (Carroll 1865, 12).

The second case study investigates how different tools interpret a written text expressed in the architects' "special lexicon". For decades, architectural criticism employed a peculiar, poetic language to discuss (sometimes heatedly) styles, buildings and tendencies. Owing to limited print quality, journals featured photographs, but above all drawings and texts: language played a decisive role, shaping readers and trends in ways hard to imagine today. In this test, we consider the (distinctly phenomenological) critique that Ernesto Nathan Rogers (1955) devoted to Le Corbusier's Chapel of *Notre-Dame du Haut* at Ronchamp – a building so iconic that one might assume it to be unmistakable.

But what happens when we reverse the process and feed that text into various text-to-image generators specifically oriented towards architecture, hence (supposedly) capable of understanding its special language? The original text was around 15,000 characters long; only its first half, dealing with the exterior, was used here to test how chatbots translate that language into their own. Three tools were used: Krea (with the Flux.1 model), Rendair (with the Nanobanana Pro model) and ChatGPT Pro (with the GPTImage model). Professional versions were employed, which, compared to free ones, allow a more structured use of the tools and a greater number of characters.

In no case could the original text be directly translated into images, largely because it was too long. This is significant in itself, even if it could be dismissed as a youthful limitation of the tools. To ensure greater fidelity rather than paraphrase, a reduced text was therefore constructed. The result is not a summary, but a selection of all descriptive sentences under 2,000 characters: a "final cut" that removes almost all metaphorical, geographical and personal references (with one intentionally retained). In all cases, however, the tools proposed an "optimisation" of the text – indeed, several – offering even shorter versions. [3]

The results are strikingly diverse. In the case of Nanobanana, [FIG. 3] initial variation was extreme: some features bear possible stylistic affinities with the chapel, while others are entirely unrelated. Subsequent variations move in the direction the tool deems most probable, blending highly imaginative architectural solutions

[3] Rogers's text (1955) cannot be reproduced here in full; instead, the "final cut" used to initiate the iterations is reported:
[L'edificio] È una massa abbastanza piccola che biancheggia sul crinale della collina.
La parete di sud (sulla quale sovrasta, sporgendo, quasi fosse il lembo di una tenda gonfiata dal vento, la

with far more prosaic ones and revealing a rather uncertain underlying architectural imaginary... None, however, bears even a remote resemblance to the chapel; even the closest one fails to respect clear indications, for instance, regarding materials or the relationship between façades.

With Flux.1 (arguably the most advanced and programmable engine), successive iterations produced equally varied outcomes, with a vast field of variation in building form, viewpoints, degrees of realism and architectural styles. [FIG. 4] Some images evoke postmodern ruins; others suggest intriguing paraboloid roofs; still others recall modernist or more contemporary works. Clear assonances with the work of other architects emerge – Kahn, Meier, Le Corbusier himself.

In the case of GPTImage 1.5 – the only engine that recognised the Chapel of Ronchamp in the description, due to its connection to the most widespread chatbot – the outcome is a set of images "strongly inspired" by the original: not a photograph or a faithful representation, but something *very* close. [FIG. 5] Compared side by side, these results display variations that may be more or less critical,

copertura in cemento armato), ha due andamenti: uno dovuto alla curvatura concava che, accentuata verso destra, si smorza verso sinistra e l'altro, dovuto all'inclinazione che, invece, si identifica con la verticale verso destra e si inclina, aumentando di grado, verso sinistra; sull'intera superficie sono incise le aperture per le vetrate, varie nelle dimensioni di altezza di larghezza e di profondità.
[C'è una] gran porta dipinta dallo stesso Le Corbusier, a smalto con colori squillanti.
[La] "torretta" convessa della facciata successiva. La quale ha pure due andamenti: quello della lieve convessità che quasi tutta l'abbraccia e l'altro di una catenaria che la ritaglia al sommo, scoprendo il lembo di cielo.
In mezzo sporge un pluviale a doppio cartoccio, dove convoglia tutto il displuvio della copertura.
[A nord] il massimo diapason è a destra di chi guarda, dove due semicilindri emergenti, simili a quello della facciata opposta, ma più piccoli, stanno dorso a dorso e s'insinuano verso l'interno.
Tra essi s'apre una porta modesta che serve all'ingresso normale dei visitatori.







[FIG. 3] three iterations of Roger's description of Ronchamp through Nanobanana

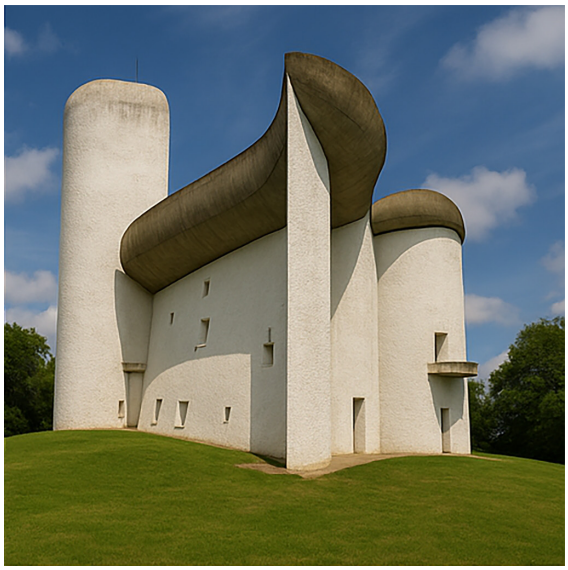





[FIG. 4] Four iterations of Roger's description of Ronchamp through Flux.1




[FIG. 5] Three iterations of Roger's description of Ronchamp through GPTImage 1.5.



or rather, detectable – showing how partial and imprecise our knowledge, even of such a famous building, is.

5. Understanding Worderland

> « I can't tell you just now what the moral of that is, but I shall remember it in a bit » (Carroll 1865, 131).

Beyond their specific outcomes, the two experiments reveal several points of broad interest, here outlined based on the previously identified key points.

5.1 – Superspecial (or: on special languages and the intellectual dimension)

> « One side will make you grow taller, and the other side will make you grow shorter » (Carroll 1865, 68)

The first consideration, perhaps an obvious one, is that although we believe our language to be determinate and precise, it is in fact ontologically infinitely ambiguous. Rogers was fully aware of this when he wrote: «No work of architecture can be conveyed through words […]. One can only attempt to grasp the essential language through which its meaning, suggested by forms, may be penetrated» (Rogers 1955, 3). Ambiguity arises because each term we consider reasonably clear reflects a language that is only apparently shared and is, in fact, special. Needless to say, awareness of this ambiguity is difficult to develop and impossible to complete, as it would require a perfect *phenomenological reduction* (Husserl 2007). Architects' language *is* therefore special, and so is that of chatbots: but in different ways. For instance, design descriptions often follow a logical order of scale, meaning, and use; a chatbot, as the first experiment shows, may instead assign greater importance to the first-appearing terms. This may disrupt linguistic exposition itself (for instance, the English adjectival order opinion-size-quality-shape-age-colour-origin-material-type-purpose may no longer apply). As a result, relations of importance between the parts of an architectural discourse change: language and grammar remain, but syntax varies depending on the engine or its customisation. Indeed, there are as many languages as the tools we use: consequently, architectural design language becomes *super-special*, and dependent on operational special languages without which it cannot be realised, yet abstracted from them.

Il tema della parete di sud si ripete in tono minore lungo questa [facciata, con] pacate aperture che sfiorano il piano e una modesta scaletta.
The prompt was automatically translated and "optimised" by the various tools, which synthesised the most "relevant" parts and proposed versions such as the following:
Small white minimalist modernist building perched on hillside ridge, south facade with concave curvature slanting leftward, inclined vertical plane with increasing left tilt, reinforced concrete cantilevered roof resembling billowing curtain edge, varied window openings of differing heights, widths, depths incised into smooth white walls, large vibrant enamel Le Corbusier-style door, convex turret on adjacent facade with gentle convexity and catenary-shaped top edge revealing sky, central double-scroll rainwater downpipe for roof runoff, natural hillside surroundings with vegetation, bright daylight illumination, photorealistic architectural rendering, eye-level perspective, crisp shadows, high-detail texture

5.2 – Lost in translation (or: on semiotic and associative relations):

> « *I* don't believe there's an atom of meaning in it » (Carroll 1865, 184).

For languages are many, prompt descriptions do not carry univocal meanings. Moreover, they lack meaning because they are not used *directly;* they are made operative in a black box that the user can control only through repeated trials. Since such trials are limited by cost, the outcome

is an ontological tendency towards reciprocal understanding, coupled with an equally ontological impossibility of isomorphism between signifiers – two signifiers, verbal input and visual output, both confronting an intended meaning. It is therefore impossible for text to coincide with what we intend to describe, for it will always describe *something* of what we describe. This reflects the infinite multiplicity of types discussed by Schutz (1979) and the many modes of being of the phenomenon (Paci 1961) – here, the phenomenon is still unknown, as it is still being designed. Hence, understanding necessarily occurs through approximation, via successive corrections aimed at reducing misunderstandings between the architectural super-language (not to be understood idealistically, as will be shown) and operational special languages, reaching a *good-enough* level (Carpo 2017). We have noted that there are many special languages. Yet as architects' formation is historically sedimented, the super-language is like a mother tongue, learned and then practised. AI, therefore, remains a foreign language, at least until it becomes integral to practice from its very conception and learning. At that point, we would become inhabitants of Worderland rather than guests, able to use its special language naturally. The price, however, would inevitably be a looser connection with *other* special languages.

5.3 – Digital illiteracy (or: on selling words and normalisation of writing)

> « I don't know the meaning of half those long words, and, what's more, I don't believe you do either! » (Carroll 1865, 32).

Current difficulties are not merely the result of AI's developmental stage: misunderstanding is mutual. In fact, practices evolve alongside the tools that shape them: a co-evolution in which each holds unique epistemological strengths and limitations (El Moussaoui 2025). On the one hand, Worderland does not depend directly on individual – nor strictly collective – agency, as it emerges from a constantly regenerated mix of imaginaries, trends, historical artefacts (real or imagined) and commercial logics. On the other hand, our co-evolutionary contribution depends on moving beyond mere adaptation to available tools or the uncritical preservation of past practices, towards the construction of a renewed – if not entirely new – horizon of meaning. This trajectory is indeed independent of AI. Architectural design has always performed repetition and variation, like any artistic practice, continuously degenerating into paraphrasing other projects and re-emerging (Possati 2013). Data normalisation – *selling words* iterating themselves – is the basis of canons and styles (Collins 1965). Yet AI amplifies this risk of iteration to unprecedented levels, leaping – radically and unexpectedly – beyond the ideological debates of the 1970s. Far from choosing between architecture as superstructure or as autonomous practice, constructing a new horizon of meaning requires, like Alice, co-evolving with tools *without* fully understanding their language: adopting their semantic flexibility while fostering conceptual rigour. This means working at the level of the super-language, accepting linguistic normalisation. In this way, co-authorship can be leveraged rather than endured (Goodfellow 2024), shifting towards a new competence – or better, mastery.

5.4 – Conceptual frailty (or: on concept and evaluative approaches)

> « Give your evidence, » said the King; « and don't be nervous, or I'll have you executed on the spot » (Carroll 1865, 168).

Architects inevitably pay the price of the ongoing deconstruction of the conceptual foundations on which their knowledge was built – a weakness so evident that it is now ignored. Design too often becomes mere technical expression, momentary "inspiration" or procedural management, reducing architecture to spatial decoration. Paradoxically, this is precisely why AI proves so productive: architectural language is already flattened into purely visualist keywords. A minimal, visual lexicon. Yet weakness reverses power relations: the risk is that trends themselves determine projects, concealed within infinite variation. In fact, in such an incommensurable and literally infinite autopoietic system, how can one realise that architectural practice remains enclosed within it, instead of «pressing the system to the limit so as to include the excluded» (Luhmann 2000, 293)? Like Alice's rhymes, AI's persuasive and supportive language convinces us that evolving iterations follow our concept, when in fact our design aligns with dominant trends. This is why prompts are "optimised" by AI, synthesising the most "relevant" keywords. But "relevant" for what? Performance parameters aside, evaluation will remain confined within dominant trends: at least, unless one looks beyond their semiotic perimeter and considers the evolving design from a genuinely systemic perspective. Is it really irrelevant to vary the windows of Ronchamp's west façade? Certainly not – yet it is equally certain that such variation is indeed acceptable *within certain limits*. Ontologically, a certain degree of blurring (variations by a few centimetres, or even millimetres) is inevitable. [4] Working at the systemic scale means understanding beyond which thresholds such variations become unacceptable: that is, operating by *excluding* unwanted meanings rather than merely denoting some of them (Deregibus & Giustiniano 2021).

[4] This obviously applies to any practice or art, since nothing ever repeats itself in exactly the same way, and nothing can be described indefinitely.

6. Enjoining Worderland

> « Oh, I've had such a curious dream! » (Carroll 1865, 189).

Working at the level of a super-language, capable of exploiting AI's special language while binding it to others, has profound implications. A super-language operates through typological containers and their inexhaustible ontological multiplicity (Deleuze & Guattari 1980). This entails an authentic deconstruction of architectural metaphysics. Rather than aiming at the perfection of gesture, designing in Worderland requires accepting the inevitable *imprecision* of the trace, because it is precisely in that imprecision that its force resides. An element with fixed, immutable meaning is ontologically fragile; one whose meaning is blurred, vibrating with possibilities, allows its syntactic use to be permuted indefinitely. There is no single ideal – there are infinitely many, as numerous as the fields of possibility opened by each design trace within the system as

a whole. Engaging with this multiplicity means operating through the exclusion of impossibilities, so that meaning emerges only once all that is impossible – or unacceptable – has been excluded. Like Holmesian truth, however improbable, it is the only valid outcome, provided that it satisfies a principle of overall systemic consistency (Deregibus 2016).

This also explains why making architecture is so difficult, even under theoretically optimal conditions. Such a process has always been conducted intuitively, precisely through continuous redrawing, through which the controllable world was explored. AI expands these possibilities to many other elements, allowing us to "sketch" across additional dimensions. Yet if it took us millennia to develop our precious *thinking hand* (Pallasmaa 2009), we can hardly expect to master new special languages with ease. It is no coincidence that generating architectural images is effortless when staying at a generic level, but incredibly hard (as the experiments show) when a precise output is required.

This is partly due to the absolute limits of description, but also because we are accustomed to treating architectural descriptions and qualities as self-evident and univocal, when in fact they rest on an unstable equilibrium lacking clear semantic structures. We enclose our own design in expectations and biases, as we do with any phenomenon. These expectations imprison our ability to perceive the infinite, fertile plurality of Worderland. They confine us to simple, repetitive, automated discourses, perfect for AI's special language, whereas an authentically strategic attitude could allow us to exploit its full potential (Deregibus 2020). By overcoming its intrinsic limits. By using its special language to give sense to an otherwise excessive freedom of meaning that would dissolve into nonsense. By co-evolving our way of designing (El Moussaoui 2025).

There are, after all, parameters that are non-communicable, non-transferable and non-objectifiable, and therefore not *evaluable* except through the sensitivity of the individual architect, who may find, precisely in AI's shortcomings, a redemption for architecture's (supposed) lack of scientificity. In this sense, Worderland can become a genuine operative matrix.

Sure, these are only images of architecture, [5] not architecture itself – but do the two really differ? Are architectures, in perception, truly more than images? Certainly yes. Yet we should leave this issue for the moment when Alice, from Worderland, goes *through the looking-glass* of architectural images.

[5] AI applications in architecture plainly extend far beyond image generation, even though this remains by far the most widespread use (Jang et al. 2025). What matters, however, from the perspective of the uninformed audience, is precisely the capacity – seemingly now available to all – to create architectures (indeed, images of architecture).

Bibliography

Alexander, C. et al. (1977). *A pattern language: Towns, buildings, construction*. Oxford University Press.

Baldi, B. (1612). *De verborum Vitruvianorum significatione, sive perpetuus in Marcum Vitruvium Pollionem commentarius... Accedit vita Vitruvii*. Ad Insigne Pinus.

Balmond, C. (2007). *Informal*. Prestel.

Berruto, G. (2020). *Sociolinguistica dell'italiano contemporaneo*. Carocci.

Biraghi, M. (2019). *L'architetto come intellettuale*. Einaudi.

Carpo, M. (2017). *The second digital turn: Design beyond intelligence*. MIT Press.

Carroll, L. (1865). *Alice's adventures in Wonderland* (J. Tenniel, Illus.). Macmillan.

Cartocci, M. (2022). Per un progetto di testo "contradditorio". Scrittura e narrazione nell'opera di maestri della teoria architettonica. *Atti e Rassegna Tecnica*, *76*(1-3), 142-147.

Collins, P. (1965). *Changing ideals in modern architecture, 1750-1950*. McGill University Press.

De Fusco, R. (1967). *Architettura come mass medium. Note per una semiologia architettonica*. Dedalo.

Deamer, P. (2020). *Architecture and labor*. Routledge.

Deleuze, G. & Guattari, F. (1980). *Mille plateaux: Capitalisme et schizophrénie*. Les Éditions de Minuit.

Deregibus, C. (2016). *Intention and Responsibility. Ethical Consistency in Contemporary Architecture*. IPOC.

Deregibus, C. (2020). Radical contingency: Strategy and tactics in architectural design. *Ardeth*, *(6)*, 205-218.

Deregibus, C. (2025). The key role of keywords in architectural design: A systemic framework. *Architecture*, *5*(2), 32, 1-20. https://doi.org/10.3390/architecture5020032

Deregibus, C. & Giustiniano, A. (2021). Progetto e ricerca della forma. Dall'aleatorio ai campi di validità. *Philosophy Kitchen*, *14*, 55-70.

Eco, U. (1971). Analisi componenziale di un segno architettonico. *Op. Cit.*, *22*, 5-29.

Eisenman, P. (1963). *The formal basis for modern architecture* (Doctoral dissertation). University of Cambridge.

El Moussaoui, M. (2025). Future illiteracies: Architectural epistemology and artificial intelligence. *Architecture*, *5*(3), 53. https://doi.org/10.3390/architecture5030053

Goodfellow, P. (2024). The distributed authorship of art in the age of AI. *Arts*, *13*, 149. https://doi.org/10.3390/arts13060149

Husserl, E. (2007). *Filosofia prima. Teoria della riduzione fenomenologica*. Rubbettino. (Original work published 1923-1924).

Iba, T. (2016). Pattern language 3.0: Writing pattern languages for human actions. In *Proceedings of the PURPLSOC Conference 2015* (83-111). https://web.sfc.keio.ac.jp/~iba/papers/PURPLSOC15_IbaKeynote.pdf

Iglesias-Vázquez, M. (2025). Between text and form: Expanded textuality in contemporary architecture. *Humanities*, *14*(8), 163. https://doi.org/10.3390/h14080163

Jang, S. et al. (2025). Generative AI in architectural design: Application, data, and evaluation methods. *Automation in Construction, 174,* 106174.

Koenig, G. K. (1964). *Analisi del linguaggio architettonico*. Libreria Editrice Fiorentina.

Kruft, H.-W. (1987). *Storia delle teorie architettoniche dall'Ottocento a oggi*. Laterza.

Kruft, H.-W. (1988). *Storia delle teorie architettoniche. Da Vitruvio al Settecento*. Laterza.

Li, Y. et al. (2025). A review of artificial intelligence in enhancing architectural design efficiency. *Applied Sciences*, *15*(3), 1476. https://doi.org/10.3390/app15031476

Luhmann, N. (2000). *Art as a social system* (E. M. Knodt, Trans.). Stanford University Press.

Mometti, F. (2012). Ideologia come architettura. Manfredo Tafuri e la storia critica. *Scienza & Politica*, *25*(47), 107-133.

Olmo, C. (2024). Coriandoli di senso. Il significato in architettura, nel tempo. *Philosophy Kitchen – Rivista di Filosofia Contemporanea*, *21*, 59-65. https://doi.org/10.13135/2385-1945/11620

Paci, E. (1961). *Tempo e verità nella fenomenologia di Husserl*. Laterza.

Pallasmaa, J. (2009). *The thinking hand: Existential and embodied wisdom in architecture*. Wiley.

Possati, L. M. (2013). *La ripetizione creatrice. Melandri, Derrida e lo spazio dell'analogia*. Mimesis.

Postle, B. & Salingaros, N. A. (2025). LLM and pattern language synthesis: A hybrid tool for human-centered architectural design. *Buildings*, *15*(14), 2400. https://doi.org/10.3390/buildings15142400

Pupi, E. et al. (2025). The sign and the word: Exploring the art of emotional communication. In *Atti del 46° Convegno UID* (3009-3018). Roma, Italy.

Romano, E. (1997). Fra astratto e concreto. La lingua di Vitruvio. In P. Gros (ed.), *Vitruvio. De architectura* (LXXIX-XCV). Einaudi.

Rogers, E. N. (1955). Il metodo di Le Corbusier e la forma della "Chapelle de Ronchamp". *Casabella Continuità*, *207*, 2-6.

Salingaros, N. A. (2025). How architecture builds intelligence: Lessons from AI. *Multimodal Technologies and Interaction*, *9*(1), 2. https://doi.org/10.3390/mti9010002

Schumacher, P. (2012). Parametric semiology: The design of information-rich environments. In P. Lorenzo-Eiroa & A. Sprecher (eds.), *Architecture in formation: On the nature of information in digital architecture* (53-60). Taylor & Francis.

Schutz, A. (1979). Simbolo, realtà e società. In *Saggi sociologici* (260-328). UTET.

Shaw, C. (2022). Design as an agent of narratives: A matrix and framework. In *Proceedings of the Design Research Society Conference (DRS 2022)*. Design Research Society. Bilbao, Spain.

Slessor, C. (2020). Image rights and wrongs: Architecture in the age of infinite reproduction. *The Architectural Review*, *(1473)*.

Venturi, R. (1966). *Complexity and contradiction in architecture*. The Museum of Modern Art.

Wiseman, C. (2014). *Writing architecture: A practical guide to clear communication about the built environment*. Trinity University Press.

Yaneva, A. (2009). *Made by the Office for Metropolitan Architecture: An ethnography of design*. 010 Publishers.

Sulla tecnologia del linguaggio e il suo futuro

Damiano Cantone

Professore associato di Filosofia e teoria dei linguaggi presso l'Università degli Studi di Udine. I suoi interessi di ricerca riguardano il linguaggio, la coscienza e il loro rapporto con le neuroscienze e le tecnologie digitali, a partire anche dalla filosofia di Gilles Deleuze.

damiano.cantone@uniud.it

This article focus on language as a social and cognitive technology that historically emerged to bridge the gap between individual experiences. Drawing on Condillac, Leroi-Gourhan, and Daniel Dor, it reconstructs the relationship between language and the body, gesture, cooperation, truth, and deception. The latter is the key for extending the analysis to large language models (LLMs) and brain–machine and brain–brain interfaces, which are considered new stages in the evolution of linguistic technologies. Finally, the article examines the potential transformation or obsolescence of symbolic language and the anthropological, ethical, and political implications of more immediate forms of neural communication.

— LINGUAGGIO
— TECNOLOGIA
— IMMAGINAZIONE
— INTELLIGENZA ARTIFICIALE
— BRAIN-COMPUTER INTERFACE

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 125 — 137

Un corpo che parla

Non è raro imbattersi nei lavori sull'origine del linguaggio considerato da una prospettiva evoluzionista un qualche riferimento a Étienne Bonnot de Condillac, inteso come colui che per primo ha stabilito una continuità lineare tra il tatto, o meglio la capacità di manipolare gli oggetti degli esseri umani e il linguaggio (Corballis 2002, Ferretti 2010), che dunque si configura come un *saper fare*, una tecnologia *sui generis*.

Nel suo celebre esperimento mentale della statua animata contenuta nel *Traité des sensations (Condillac 1754: 3-4)*, Condillac, in polemica con ogni forma di innatismo, intende mostrare come tutto ciò che chiamiamo mente derivi dai sensi. Immagina una specie di golem umano che non possiede linguaggio, concetti, abitudini né intenzioni: rappresenta il grado zero del mentale. È un punto di partenza assoluto, dal quale osservare la formazione progressiva delle operazioni dell'anima. La statua è inizialmente dotata di un solo senso che viene attivato isolatamente, poi in combinazione con gli altri: si evitano così spiegazioni circolari per mostrare come le facoltà emergano gradualmente, per differenziazione e rafforzamento, a partire dalla sensazione stessa.

Nelle prime fasi del *Traité*, quando è attivo soltanto l'olfatto, la statua è in grado di provare piacere e dolore e sviluppa forme rudimentali di attenzione e memoria. Tuttavia, queste operazioni restano confinate all'interiorità: le sensazioni sono semplici *modi d'essere* dell'anima, non ancora percezioni di oggetti. La coscienza, in questa fase, non è coscienza del mondo, ma solo di stati interni. Questo punto è decisivo, poiché mostra che il rapporto con la realtà esterna non è immediato né originario, ma deve essere costruito. La svolta avviene con l'attivazione del tatto: a differenza degli altri sensi, il tatto introduce l'esperienza della resistenza e della solidità, permettendo alla statua di distinguere tra sé e ciò che non è sé. È solo attraverso il tatto che emerge l'idea di corpo esterno e, con essa, la possibilità stessa di un mondo oggettivo (Coratelli 2014). Ecco perché il tatto non è semplicemente un senso tra gli altri, ma il fondamento ontologico della conoscenza: senza tatto non vi è oggetto, né spazio, né estensione.

L'importanza del tatto non si esaurisce nell'apertura alla realtà esterna. Esso consente anche un decisivo sviluppo delle operazioni cognitive: la memoria diventa più stabile, la comparazione più articolata, il giudizio più preciso. Attraverso l'esplorazione tattile del proprio corpo e dei corpi esterni, la statua impara a isolare proprietà, a confrontarle e a generalizzarle. Nascono così le prime idee generali, ancora grossolane e confuse, eppure capaci di oltrepassare la singolarità della sensazione immediata. In questo processo, la riflessione non appare come una facoltà innata o improvvisa, ma come il risultato di un progressivo rafforzamento delle operazioni dell'anima.

È a questo livello che la statua assume un significato decisivo per la teoria del linguaggio. Nel *Traité*, infatti, Condillac non parla quasi mai esplicitamente di segni o di linguaggio; e tuttavia il tatto svolge esattamente la funzione che nell'*Essai sur l'origine des connaissances humaines* (Condillac 1746) era attribuita al segno arbitrario. Come il segno nell'opera precedente, il tatto permette di decomporre l'esperienza, di articolare le *liaisons des idées* e di rendere possibile l'astrazione. In particolare, il gesto che nasce dall'esperienza tattile — il movimento della mano, l'indicazione

con le dita, il contare — assume la funzione di primo segno naturale di tipo cognitivo. Il gesto è un segno per il soggetto prima di essere un segno per l'altro, e dunque è ausiliario a al pensiero anziché alla comunicazione.

In questo modo, Condillac sposta l'origine della significazione dal linguaggio articolato al corpo. Il segno dall'azione; non dalla convenzione, ma dall'esperienza sensoriale. La statua, giunta al termine del suo sviluppo, è in grado di ricordare, immaginare, astrarre e combinare idee, pur restando priva di linguaggio verbale. Ciò dimostra che, per Condillac, il linguaggio non è la condizione minima del pensiero, ma il suo prolungamento e il suo potenziamento, in certa misura una sua applicazione pratica. Il linguaggio articolato diventa necessario per la stabilità, la complessità e la trasmissione sociale delle idee, non per la loro genesi originaria.

Secondo Condillac, dunque, si può essere umani anche *senza* linguaggio. Una simile assunzione non dovrebbe poi stupirci più di tanto. Al di là dell'ovvia considerazione che ci sono molti esseri umani che sono esclusi dal linguaggio fin dalla nascita (i bambini sordo muti e ciechi, per esempio) o perché colpiti da malattie o incidenti invalidanti, basta pensare che la specie Homo esiste da molto prima dell'invenzione del linguaggio. Datare l'origine del linguaggio umano non è semplice (Watts 1999; Dediu & Levinson 2013; Everett 2017), probabilmente il linguaggio ha iniziato a emergere nelle comunità di *Homo erectus* o *Homo heidelbergensis*, caratterizzate da una forte cooperatività, inventiva, abilità tecnologiche e capaci di identificarsi a vicenda. È inoltre molto probabile che fosse presente in una forma molto avanzata anche presso i Neanderthal, dato che essi seppellivano come noi i loro morti, ornandone le tombe, indice di una capacità di immaginazione simbolica complessa. Tuttavia è solo all'interno della specie Homo sapiens, (presente sul pianeta da circa 300.000 anni e l'unica specie di Homo sopravvissuta alla selezione naturale) che il linguaggio orale ha assunto una conformazione simile a quella odierna. La maggior parte degli studiosi (Lieberman 1991, 2007; Corballis 2020) ritiene che il processo di stabilizzazione del linguaggio si sia compiuto circa 90 mila anni fa.

Il fatto che l'invenzione, la stabilizzazione e il perfezionamento del linguaggio siano avvenuti nel corso di un arco di tempo così ampio, ci permette di analizzarlo sullo sfondo della storia dello sviluppo tecnologico dell'umano, anziché della sua evoluzione biologica Tomasello (2014; 2019). Come aveva già osservato Leroi-Gourhan, "esiste la possibilità di un linguaggio a partire dal momento in cui la preistoria ci tramanda degli utensili, perché utensile e linguaggio sono collegati neurologicamente e perché l'uno non è dissociabile dall'altro nella struttura sociale dell'umanità" (Leroi-Gourhan 1977, 136). Significa che non si tratta di un'abilità naturale che si è evoluta e selezionata a livello genetico a partire da una qualche mutazione casuale, ma di una tecnologia, con funzioni e limiti specifici, frutto di un'invenzione collettiva da parte delle prime comunità umane – dunque socialmente determinata. Mano, gesto, utensile e parola formano un sistema integrato: la liberazione della mano grazie al bipedismo, l'uso di strumenti e l'organizzazione tecnica dell'ambiente aprono progressivamente lo spazio per forme sempre più sofisticate di coordinamento, memoria e trasmissione culturale. Il linguaggio, in questo senso, è qualcosa che va al di là di un semplice mezzo di comunicazione per configurarsi come una tecnologia cognitiva che nasce nello stesso processo evolutivo che produce le altre tecniche.

Questa impostazione consente di spostare la domanda tradizionale "che cos'è il linguaggio?" verso una questione più funzionale: che cosa fa il linguaggio, e perché è così efficace?

Tecnologia e immaginazione

Partiamo da una constatazione: a differenza dei gesti, delle espressioni facciali o delle vocalizzazioni animali che operano all'interno di un contesto esperienziale condiviso, il linguaggio permette di comunicare al di là dell'esperienza immediata. Esso fornisce coordinate, schemi, istruzioni che guidano l'interlocutore nella costruzione immaginativa di situazioni, eventi e stati di cose non direttamente presenti. Questo elemento è decisivo: ogni esperienza umana è, in ultima analisi, privata. Anche nelle forme più intense di interazione e cooperazione, ciascun individuo vive il mondo da una prospettiva unica, mediata dal proprio corpo, dalla propria storia e dal proprio ambiente. Questo divario esperienziale costituisce un limite strutturale alla comunicazione. Il linguaggio nasce come risposta tecnica a questo limite, come uno strumento capace di rendere condivisibile ciò che, per sua natura, non lo è.

Grazie al linguaggio, gli esseri umani possono condividere non solo ciò che vedono qui e ora, ma anche ciò che è assente, passato, futuro, possibile o addirittura impossibile. Possono trasmettere esperienze indirette, ipotesi, regole, narrazioni, progetti e norme. In questo senso, il linguaggio serve innanzitutto a estendere la cooperazione oltre i confini dell'esperienza immediata, rendendo possibile una memoria collettiva, una pianificazione a lungo termine e la costruzione di mondi simbolici complessi. Considerato come tecnologia, il linguaggio svolge anche una funzione cruciale di esternalizzazione della memoria (Levy 1995; Stigler 2023). Prima dell'invenzione della scrittura, il linguaggio orale ha permesso di stabilizzare saperi, miti, tecniche e norme attraverso la narrazione e la ripetizione rituale. Con la scrittura, questa funzione si è ulteriormente potenziata, mostrando come le tecnologie linguistiche possano stratificarsi e amplificarsi reciprocamente. Il linguaggio ha dunque consentito la formazione e la conservazione di tradizioni, istituzioni e identità culturali, plasmando profondamente il modo in cui gli esseri umani pensano, agiscono e convivono.

Affermare perciò che il linguaggio sia una tecnologia non significa affatto affermarne un carattere strumentale: non si può pensare che si possa "utilizzare" il linguaggio come fosse un qualunque utensile. Piuttosto, la sua cifra tecnologica risiede nella profonda natura relazionale che si esplica nella sua esistenza tra i parlanti, e non dentro di loro: è quello che fanno le tecnologie, che richiedono una partecipazione attiva di chi le utilizza, una trasformazione dei comportamenti (e dunque anche un diverso utilizzo delle abilità cognitive che le preesistevano) in vista di una differente relazione con l'ambiente. Inoltre, ogni tecnologia è funzionalmente specifica, e accogliendo la proposta di Daniel Dor (2015) la tecnologia del linguaggio serve per l'*istruzione dell'immaginazione*. Questo punto merita una discussione più ampia. Il linguaggio non serve a condividere direttamente le esperienze (come fanno i segnali animali esperienziali) che fa un soggetto – non è dunque una semplice tecnologia di comunicazione – piuttosto,

il suo ruolo è quello di colmare il gap esperienziale tra me e gli altri. La sua potenza si esplica non tanto e non solo nel dare i nomi alle cose che sono presenti qui ed ora, ma nel creare un ponte tra la mia esperienza privata, fonte primaria della mia conoscenza su mondo, e l'esperienza privata delle altre persone, che per me è assolutamente inaccessibile (Ginsburg & Jablonka, 2014; Jablonka, Dor, Ginsburg, 2012; Dor & Jablonka, 2014).

La posizione di Daniel Dor sul linguaggio si colloca all'interno di una prospettiva radicalmente alternativa rispetto alle concezioni che lo interpretano come una facoltà mentale innata o come un modulo cognitivo specializzato. Secondo Dor, il linguaggio va compreso anzitutto come una tecnologia sociale della comunicazione, ovvero come un artefatto collettivamente costruito nello spazio intersoggettivo che si crea tra i parlanti, e non in qualche recondito spazio delle loro menti. Il linguaggio, in questa prospettiva, è una pratica sociale storicamente emersa per risolvere un problema fondamentale della condizione umana: la radicale separazione tra le esperienze soggettive degli individui. Ogni essere umano vive infatti in un mondo esperienziale privato, accessibile solo dal proprio punto di vista, e questa distanza costituisce un ostacolo strutturale alla cooperazione e alla trasmissione della conoscenza. I sistemi comunicativi pre-linguistici – gesti, espressioni facciali, posture, vocalizzazioni – operano all'interno di un contesto esperienziale condiviso e permettono solo una comunicazione "qui e ora", legata alla presenza fisica e alla situazione immediata. Quando parliamo, secondo Dor, non trasmettiamo direttamente esperienze, né "contenuti mentali" già formati; forniamo piuttosto all'interlocutore una serie di istruzioni, coordinate e vincoli che lo guidano nel processo di costruzione immaginativa di un'esperienza possibile. Le parole funzionano come segnali convenzionali che attivano ricordi, schemi e aspettative nell'altro, permettendogli di ricostruire nella propria mente qualcosa che non sta vivendo direttamente. In questo senso, il linguaggio non elimina il divario esperienziale tra gli individui, ma lo rende *attraversabile*, trasformando la comunicazione in un'impresa cooperativa che richiede lavoro interpretativo, fiducia e coordinamento sociale. Da qui deriva anche l'idea che il linguaggio non sia una tecnologia generica, ma funzionalmente specifica: esso non serve a "comunicare tutto", bensì a rendere possibile la condivisione indiretta di esperienze assenti, passate, future o ipotetiche, e dunque a estendere enormemente le capacità di cooperazione umana. Questa funzione spiega sia la potenza sia la fragilità del linguaggio: poiché si basa sull'immaginazione e non sulla percezione diretta, il fraintendimento non è un incidente marginale, ma una possibilità strutturale.

Un altro aspetto centrale della teoria di Dor riguarda l'evoluzione e l'acquisizione del linguaggio. Se il linguaggio è una tecnologia sociale, allora non può essersi evoluto primariamente come adattamento genetico individuale; al contrario, deve essere emerso come innovazione culturale collettiva, all'interno di comunità già caratterizzate da alti livelli di cooperazione, fiducia e intersoggettività. In questa cornice, l'evoluzione biologica dell'Homo sapiens, che ha determinato il suo bagaglio di capacità cognitive, viene in parte rimodellata retroattivamente dall'uso prolungato del linguaggio: prima gli esseri umani lo inventano come tecnologia, poi il linguaggio contribuisce a trasformare le capacità cognitive, sociali ed emotive dei suoi utenti. (Dor 2015) Anche l'acquisizione infantile del

linguaggio va riletta in questa chiave: il bambino non "scopre" una grammatica innata, ma impara progressivamente a partecipare a una pratica sociale complessa, interiorizzando convenzioni che esistono già nella comunità e che continuano a evolversi nel tempo.

L'importanza di dire la verità.

C'è un altro aspetto, tuttavia, che è stato reso possibile dall'invenzione del linguaggio: dire la verità.

Per la filosofia, il rapporto tra linguaggio e verità è uno dei nodi teorici fondamentali e attraversa l'intera storia del pensiero occidentale, assumendo configurazioni molto diverse a seconda di come si concepiscono realtà, conoscenza e soggetto. Ciò che le accomuna, in ogni caso, è il riconoscimento che non esiste una teoria della verità che possa prescindere da una teoria del linguaggio, perché è sempre nel linguaggio che la verità viene detta, discussa, contestata o trasforma.

La filosofia, in quanto pratica linguistica, ha dunque intrinsecamente a che fare con la verità, poiché verità e linguaggio non possono mai darsi l'una senza l'altro. Prima del linguaggio può esistere errore, fallimento percettivo, reazione inadeguata all'ambiente, ma non può esistere propriamente né verità né menzogna, poiché entrambe presuppongono una distanza tra esperienza, rappresentazione e valutazione condivisa (Dawkins & Krebs, 1978). Il linguaggio introduce questa distanza come effetto della sua stessa struttura tecnica: esso non trasmette l'esperienza, ma istruisce l'immaginazione dell'altro, fornendo coordinate parziali, convenzionali e necessariamente incomplete che l'interlocutore deve interpretare e integrare sulla base del proprio vissuto. In questo spazio interpretativo nasce la possibilità della verità, intesa non come semplice corrispondenza immediata tra parola e mondo, ma come esito sempre provvisorio di una cooperazione riuscita tra parlanti che riescono ad allineare le proprie immaginazioni rispetto a uno stesso stato di cose. Per queste stesse ragioni, ovviamente, è solo all'interno del linguaggio che abbiamo a che fare con il problema della menzogna.

Come abbiamo visto, il linguaggio non elimina il divario esperienziale tra individui, ma rende possibile operare su di esso, trasformandolo in uno spazio di negoziazione cognitiva e sociale. La menzogna emerge quando questo stesso meccanismo viene sfruttato strategicamente per orientare l'immaginazione dell'altro verso stati di cose che il parlante non sa non essere conformi all'esperienza o ai fatti. È evidente che la menzogna dunque non sia definibile solo in termini di falsità: essa richiede intenzionalità, consapevolezza del punto di vista dell'altro e capacità di anticipare l'effetto immaginativo delle proprie parole, in breve, *l'inganno* (Dor 2017). Questo significa che la menzogna è un fenomeno cognitivamente e socialmente sofisticato, possibile solo all'interno di una comunità linguistica in cui esistono criteri condivisi, seppur mobili, di verità (Jablonka *et al.* 2014). Paradossalmente, dunque, la menzogna presuppone la verità: si può mentire solo là dove esiste già un orizzonte normativo che distingue il dire corretto dal dire scorretto, il resoconto affidabile da quello ingannevole. In questo senso, il linguaggio non è semplicemente il mezzo attraverso cui verità e menzogna vengono espresse, ma è la condizione stessa della loro genesi. Ogni atto linguistico, anche il più semplice, si colloca implicitamente

entro un regime di veridicità, poiché propone all'interlocutore una certa ricostruzione immaginativa del mondo e chiede di essere accettato, corretto o rifiutato. Allo stesso modo, la menzogna non è un semplice *abuso* del linguaggio, ma una possibilità interna al suo funzionamento, che rivela quanto il linguaggio sia inseparabile da relazioni di fiducia, potere e responsabilità (Foucault). Mentire significa sfruttare la fiducia linguistica dell'altro, ovvero il fatto che, nella maggior parte dei casi, la cooperazione sociale richiede di assumere che l'interlocutore stia tentando di dire il vero (Spence *et al.* 2004) . Senza questa assunzione di fondo, il linguaggio perderebbe la sua efficacia come tecnologia cooperativa; ma proprio questa assunzione rende possibile la sua violazione.

Dire il vero e dire il falso sono principalmente atti che modificano il tessuto delle relazioni sociali, rafforzando o erodendo la fiducia collettiva. Inoltre, poiché il linguaggio consente di parlare di ciò che non è presente, esso rende possibile non solo la menzogna fattuale, ma anche la finzione, l'ipotesi, la previsione, l'ideologia e il mito, collocando verità e falsità su uno spettro complesso che non coincide mai perfettamente con la semplice opposizione vero/falso. In questo quadro, la verità è sempre situata, contestuale e negoziata, mentre la menzogna non è sempre facilmente identificabile come tale, potendo assumere forme sottili, parziali o sistemiche. Il linguaggio, in quanto tecnologia dell'immaginazione condivisa, non garantisce la verità, ma rende possibile la sua ricerca; allo stesso tempo, non produce automaticamente la menzogna, ma ne rende praticabile l'invenzione. È proprio questa ambivalenza strutturale a spiegare perché il linguaggio sia al tempo stesso il fondamento della conoscenza umana e uno dei principali strumenti di manipolazione: esso crea il campo in cui verità e menzogna possono emergere come categorie significative, storicamente variabili e socialmente decisive.

La natura artificiale dei linguaggi naturali

Aver chiarito che dire la verità e mentire sono due funzioni sorelle permesse dalla tecnologia del linguaggio ci permette forse di interpretare meglio quello che sta avvenendo oggi nell'attuale panorama della trasformazione digitale e dell'applicazione dell'Intelligenza Artificiale al nostro mondo dell'esperienza. I modelli linguistici di grandi dimensioni, comunemente indicati come LLM, rappresentano infatti oggi una delle manifestazioni più avanzate della tecnologia del linguaggio, ma non costituiscono né una rottura radicale né tantomeno il termine ultimo della sua evoluzione: essi vanno piuttosto compresi come una tappa specifica all'interno di una lunga storia di esternalizzazione, formalizzazione e automazione delle pratiche linguistiche umane(Deniz *et al.* 2019; Fedorenko, *et al.* 2010; MacSweeney *et al.* 2002.). Gli LLM operano interamente all'interno della dimensione del linguaggio, portandola a un grado di astrazione e di scalabilità senza precedenti, senza tuttavia violarne le regole di funzionamento, sfruttando anzi esattamente il meccanismo di verità e menzogna che ho cercato di illustrare nel paragrafo precedente. Sistemi come Chat-GPT e affini non "comprendono" il linguaggio nel senso esperienziale o pragmatico del termine, ma ne simulano il funzionamento formale attraverso l'analisi statistica di enormi quantità di testi prodotti da esseri umani nel corso della storia recente. In questo senso, gli LLM non sono che dispositivi di

sedimentazione culturale: incorporano, comprimono e riorganizzano regolarità linguistiche, stili argomentativi, convenzioni discorsive e forme simboliche già esistenti, rendendole riutilizzabili in modo automatico. La loro apparente competenza linguistica deriva dal fatto che il linguaggio stesso è una tecnologia basata su istruzioni condivise e su convenzioni storicamente stabilizzate; gli LLM eccellono proprio nel riprodurre queste istruzioni, ma restano estranei alla dimensione cooperativa e immaginativa che dà senso all'atto linguistico umano. Da questo punto di vista, essi possono essere paragonati ad altre grandi svolte nella storia della tecnologia linguistica, come l'invenzione della scrittura, della stampa o dei media digitali: ogni volta, una parte delle funzioni del linguaggio viene esternalizzata e resa automatizzabile, senza che ciò comporti la scomparsa del linguaggio stesso. Gli LLM, infatti, del tutto estranei al linguaggio come pratica sociale, ne amplificano tuttavia la dimensione formale, rendendo evidente una distinzione centrale: quella tra competenza linguistica e comprensione funzionale. I modelli linguistici possiedono una competenza puramente sintattico-statistica, mentre restano privi di accesso all'esperienza, al contesto situato e alla responsabilità del dire. Proprio per questo, essi mostrano in modo quasi paradigmatico come il linguaggio non sia riducibile a una capacità cognitiva specifica interna, né a una combinazione di molte capacità cognitive: è piuttosto una tecnologia culturale che può essere imitata, manipolata e parzialmente automatizzata. In continuità con la storia evolutiva del linguaggio, gli LLM possono essere visti come una nuova forma di "memoria esterna", analoga ai grandi archivi scritti, con in più una dimensione dinamica e generativa, capace di produrre testi nuovi a partire da pattern culturali sedimentati. Tuttavia, questa capacità generativa non equivale a un atto linguistico in senso pieno: manca l'intenzionalità comunicativa, manca il rischio dell'errore vissuto, manca il rapporto etico con la verità e con l'altro. Gli LLM non partecipano alla costruzione sociale del senso, semplicemente la ricalcano.

Più che temere la concorrenza dei LLM, dovremmo forse cominciare a riflettere sulla lezione che ci stanno insegnando: in prima istanza, il fatto che una macchina è in grado di simulare un insieme di pratiche simboliche che possono essere apprese, replicate e trasformate, e che fino ad oggi erano pertinenza specifica degli esseri umani. Allo stesso tempo, il limite strutturale degli LLM mostra anche ciò che nel linguaggio non è automatizzabile: la dimensione sociale della cooperazione, il riferimento all'esperienza vissuta, la responsabilità pragmatica del parlare. Gli LLM, dunque, non segnano la fine del linguaggio, ma una fase avanzata della sua evoluzione culturale, in cui la tecnologia del linguaggio riflette su se stessa e rende visibili i propri meccanismi. Essi sono strumenti che amplificano la capacità umana di produrre e riorganizzare simboli, e restano dipendenti dal linguaggio umano come pratica originaria. Ma probabilmente non sarà questa l'ultima tappa dell'evoluzione del linguaggio.

La fine del linguaggio è vicina?

Sappiamo che ogni tecnologia umana va incontro a un processo di obsolescenza, che si traduce tuttavia raramente nel suo abbandono, più spesso nella sua integrazione in una forma di tecnologia più efficace o piuttosto nella ridefinizione del proprio ambito di efficacia. Se pensiamo, per

esempio, all'evoluzione della tecnologia dei trasporti, ci rendiamo ben conto che l'invenzione dell'aereo – che consente spostamenti rapidi e ben localizzati in tutto il globo – non ha messo fine alla ben più antica arte della navigazione – che consente spostamenti molto più lenti e legati alla presenza di acqua (mari e fiumi navigabili). Certamente è finita l'epoca d'oro dei transatlantici, visto che le persone preferiscono affidarsi ai voli intercontinentali, ma la stragrande maggioranza delle merci viaggia ancora per nave, senza contare coloro che – non avendo fretta – preferiscono salpare gli oceani a bordo di qualche lussuosa crociera. Tuttavia pare sensato, almeno da un punto di vista speculativo, porsi l'interrogativo sul limite del linguaggio. L'ipotesi di una possibile fine del linguaggio come tecnologia, o quantomeno di una sua trasformazione radicale, emerge oggi con forza nel dibattito sulle interfacce cervello–cervello e cervello–macchina, e trova il suo fulcro nella progressiva erosione della dimensione simbolica che ha costituito, per decine di migliaia di anni, il cuore stesso del funzionamento linguistico (Pais-Vieira, Lebedev, Kunicki, *et al.* 2013). Il linguaggio, inteso come tecnologia sociale, si fonda infatti sulla mediazione simbolica: parole, segni e strutture sintattiche non trasmettono esperienze, ma istruiscono l'immaginazione, costringendo i parlanti a un lavoro interpretativo che è al tempo stesso cognitivo, sociale ed etico. Non a caso a più riprese l'Homo sapiens è stato identificato come specie simbolica, a partire dalle considerazioni di Cassirer, per il quale il simbolo "è l'organo necessario ed essenziale" del pensiero, (Cassirer 1961, 20) fino a Deacon, che partendo da una prospettiva peirciana, rivendica l'unicità e l'importanza del linguaggio per la nostra specie, in un'ottica di coevoluzione di linguaggio e strutture cerebrali (Deacon 2001).

Le nuove tecnologie di comunicazione neurale promettono invece di evitare il lavoro del simbolico, proponendo una forma di connessione diretta tra stati cerebrali, o tra cervello umano e sistemi artificiali, che non passa più attraverso il filtro simbolico del linguaggio (Wolpaw *et al.* 2002; Jiang *et al.* 2019) . Progetti come quelli sviluppati da Neuralink o da altri centri di ricerca sulle brain–computer interfaces si basano sull'idea che sia possibile codificare, trasmettere e decodificare pattern neurali in modo immediato, riducendo drasticamente la necessità di tradurre l'esperienza in segni convenzionali (Li 2023). Se questa promessa dovesse realizzarsi pienamente, il linguaggio perderebbe la sua funzione di tecnologia primaria della comunicazione umana, venendo sostituito – o perlomeno affiancato – da flussi informativi pre-simbolici o post-simbolici, in cui ciò che circola sono solamente correlati neurali di stati mentali, emozioni o intenzioni. In particolare due recenti esperimenti sembrano andare esattamente in questa direzione. Nel primo – seminale – esperimento, i ricercatori Rao e Stocco (2014) propongono un pioneristico protocollo di comunicazione diretta cervello-cervello tra due persone, realizzato combinando due tecniche non invasive: EEG per leggere un'intenzione motoria dal cervello del "mittente" e TMS per trasformare quell'informazione in un movimento della mano del "ricevente". L'obiettivo non è trasferire pensieri complessi, ma dimostrare che un'informazione intenzionale semplice può essere estratta da un cervello, trasmessa via Internet e utilizzata per influenzare direttamente l'attività motoria di un altro cervello, senza utilizzare i normali canali comunicativi, senza

voce, gesto o testo.

Per testare il sistema, Rao e Stocco utilizzano un semplice videogioco a due giocatori. Sullo schermo compare una nave pirata che lancia razzi verso una città. L'obiettivo è sparare con un cannone per intercettare i razzi. Ma i due partecipanti dispongono di informazioni e possibilità d'azione diverse: Rao, il mittente, è l'unico che può vedere lo schermo e quindi sa quando bisogna sparare; Stocco, il ricevente, è l'unico fisicamente in grado di premere il tasto che attiva il cannone. In questo modo nessuno dei due può completare il compito da solo. Rao indossa una cuffia EEG dotata di 32 elettrodi, che registra le variazioni dell'attività elettrica cerebrale, e quando vede che è il momento di sparare, non muove realmente la mano: immagina di farlo. Questa immaginazione motoria produce un pattern EEG riconoscibile e il software identifica questo pattern come segnale di "sparare". È importante sottolineare che il sistema non "legge" un pensiero astratto del tipo *voglio colpire il razzo*: rileva un segnale neurale associato all'immaginazione di un movimento della mano.

Una volta riconosciuto il pattern EEG, il computer trasmette il comando attraverso Internet al sistema collegato a Stocco. Stocco ha una bobina di stimolazione magnetica transcranica (TMS) posizionata sopra la corteccia motoria sinistra, la regione cerebrale che controlla la mano destra. Quando arriva il segnale dal mittente, la TMS stimola la corteccia motoria di Stocco. Il risultato è un movimento involontario della mano destra: la mano si solleva e, ricadendo, colpisce il tasto della tastiera che fa sparare il cannone nel videogioco. Con la pratica, i partecipanti migliorano progressivamente la capacità di utilizzare il sistema. Il miglioramento dipende anche dall'apprendimento del mittente: Rao impara a immaginare il movimento della mano in modo più coerente, producendo un segnale EEG più facilmente riconoscibile dal sistema. In questo caso, dunque, l'espressione "brain-to-brain communication" può sembrare molto più spettacolare di ciò che è stato effettivamente dimostrato. Non viene trasferito un pensiero complesso, una parola, un'immagine mentale o un contenuto semantico, ma essenzialmente un'informazione binaria e motoria del tipo 0/1. In ogni caso, l'esperimento dimostra chiaramente che è possibile trasferire informazioni da un cervello a un altro senza alcuna mediazione di tipo simbolico.

Il secondo esperimento, invece, ha l'obiettivo di verificare se sia possibile ricostruire linguaggio continuo a partire dall'attività cerebrale registrata in maniera non invasiva tramite fMRI. In sostanza, gli autori cercano di passare dalla registrazione dell'attività cerebrale a una sequenza di parole che restituisca il significato di ciò che una persona sta ascoltando o immaginando. L'esperimento principale è particolarmente intensivo e viene condotto su tre partecipanti, per ciascuno dei quali viene costruito un decoder individuale. Questo è un elemento metodologico fondamentale: il sistema non è un decoder universale capace di leggere direttamente l'attività cerebrale di qualsiasi persona, ma richiede una lunga fase di addestramento sul singolo individuo. Durante la fase di training, ciascun partecipante ascolta complessivamente circa 16 ore di storie narrative in linguaggio naturale mentre viene sottoposto a fMRI. Il modello di intelligenza artificiale applicato a questi dati impara, per ciascun partecipante, la relazione tra le caratteristiche semantiche delle parole ascoltate e

il pattern di attività prodotto nel cervello. Le caratteristiche semantiche vengono estratte attraverso un modello linguistico GPT pre-addestrato e vengono utilizzate per predire mediante regressione la risposta del cervello.

Il decoder utilizza quindi un LLM per proporre possibili continuazioni di una sequenza di parole. Le diverse sequenze candidate vengono poi confrontate con l'attività cerebrale effettivamente registrata: stima quale delle sequenze candidate sia più compatibile con il pattern osservato. Le alternative migliori vengono mantenute e il processo continua progressivamente, parola dopo parola. Dopo la fase di preparazione, ai partecipanti vengono fatte ascoltare nuove storie, che il sistema non ha mai incontrato durante il training. Il decoder deve quindi ricostruire il contenuto linguistico utilizzando esclusivamente le nuove registrazioni fMRI. Il risultato importante è che il sistema non funziona come una trascrizione parola per parola. A volte recupera esattamente singole parole o espressioni, ma molto più frequentemente produce una parafrasi semanticamente simile del testo ascoltato dai soggetti sperimentali. Questo è un punto fondamentale per interpretare correttamente lo studio: il decoder non "legge i pensieri" traducendo direttamente ogni parola dal cervello. Ricostruisce piuttosto una sequenza linguistica plausibile che tende a preservare il contenuto semantico dell'esperienza.

È comunque un risultato teoricamente molto importante perché suggerisce che il decoder non stia semplicemente imparando una relazione tra suono delle parole e attività cerebrale. Sta intercettando rappresentazioni di livello più astratto, cioè rappresentazioni semantico-concettuali che possono essere condivise tra modalità diverse. L'elemento più interessante, dal punto di vista della semantica e della rappresentazione del linguaggio, non è semplicemente che "la fMRI può leggere ciò che pensiamo". È qualcosa di più preciso: gli autori mostrano che dalle rappresentazioni semantiche distribuite nella corteccia è possibile ricostruire una sequenza linguistica che conserva una parte significativa del contenuto concettuale dell'esperienza, anche quando la forma linguistica esatta non viene recuperata.

Naturalmente, questi esperimenti hanno immediatamente attirato l'interesse dell'industria militare, che possiede certamente le capacità economiche e gli appoggi istituzionali necessari per finanziarne lo sviluppo (Binnendijk *et al.* 2020; Latheef 2022). In uno scenario del genere, la comunicazione non richiederebbe più interpretazione nel senso forte del termine: non vi sarebbe più bisogno di negoziare significati, di confrontare punti di vista, di colmare il divario esperienziale attraverso la cooperazione immaginativa. Proprio qui si colloca il rischio più profondo: la perdita della dimensione simbolica non comporterebbe soltanto un aumento dell'efficienza comunicativa, ma una trasformazione ontologica del modo in cui gli esseri umani costruiscono la realtà condivisa (Hildt 2019). Il simbolo, infatti, non è un semplice mezzo tecnico, ma un dispositivo di distanza: separa il soggetto dall'esperienza immediata, introduce un intervallo riflessivo, rende possibile il dubbio, l'errore, la revisione e la critica. Eliminare o ridurre drasticamente questa distanza significa produrre una comunicazione che tende alla trasparenza totale, ma che proprio per questo rischia di diventare opaca sul piano etico e politico. Significherebbe, letteralmente, la *fine della menzogna*, ma questo potrebbe non rivelarsi una cosa così positiva. Se ciò che viene trasmesso non è più una rappresentazione, ma

uno stato neurale, chi controlla la codifica, la selezione e l'interpretazione di tali stati? In assenza di simboli condivisi, la verità non può più essere discussa, ma solo misurata o validata tecnicamente; scompare la possibilità di mentire – possibilità istituita dalla comparsa del linguaggio – ma non quella di ingannare, che è indipendente dal linguaggio. Semplicemente l'inganno si sposta a monte, nei protocolli, negli algoritmi, nelle infrastrutture che decidono cosa conta come contenuto mentale comunicabile. In questo senso, la fine del linguaggio come tecnologia simbolica non coinciderebbe con la fine della comunicazione, ma con la fine di una comunicazione fondata sulla responsabilità del dire. Il linguaggio obbliga chi parla a esporsi, a prendere posizione, a scegliere parole che possono essere contestate, reinterpretate, rifiutate; una comunicazione neurale diretta, invece, rischia di naturalizzare ciò che viene trasmesso, presentandolo come dato immediato e indiscutibile. Inoltre, la perdita della simbolicità linguistica implicherebbe anche la dissoluzione della dimensione storica del senso: le lingue cambiano nel tempo, portano con sé tracce del passato, conflitti sedimentati, ambiguità produttive; i flussi neurali, al contrario, tendono a essere sincronici, funzionali, orientati all'ottimizzazione. In un mondo in cui la comunicazione avvenisse prevalentemente tra cervello e cervello o tra cervello e macchina, il senso non emergerebbe più dall'attrito tra interpretazioni diverse, ma dalla compatibilità tecnica tra sistemi. Ciò che verrebbe meno non è solo il linguaggio, ma il soggetto linguistico stesso, inteso come essere capace di abitare l'incertezza del significato. Da questo punto di vista, parlare di "fine del linguaggio" non significa immaginare un silenzio comunicativo, ma una transizione verso forme di interazione che rinunciano alla dimensione simbolica per privilegiare l'immediatezza informazionale. Se il linguaggio è la tecnologia che ha reso gli esseri umani animali simbolici, la sua eventuale obsolescenza segnerebbe il passaggio a una nuova condizione antropologica, in cui la comunicazione non costruisce più mondi, ma li sincronizza, e in cui il prezzo dell'efficienza è la rinuncia a quella fragilità simbolica da cui sono nate la cultura, la verità e la libertà stessa.

Bibliografia

Binnendijk, A. et al. (2020). *Brain-Computer Interfaces: U.S. Military Applications and Implications, an Initial Assessment*. RAND Corporation.

Condillac, É.B. de. (1746). *Essai sur l'origine des connoissances humaines* (I-II). Pierre Mortier.

Condillac, É.B. de. (1754). *Traité des sensations* (I-II). De Bure.

Coratelli, G. (2014). Dal segno arbitrario al tatto. Condillac e l'origine naturale del linguaggio umano. *Filosofi(e)Semiotiche*, 1 (1).

Corballis, M.C. (2008). *Dalla mano alla bocca. Le origini del linguaggio*. Raffaello Cortina.

Corballis, M.C. (2020). *La verità sul linguaggio (per quel che ne so)*. Carocci.

Dawkins, R. & Krebs, J.R. (1978). Animal Signals: Information or Manipulation? In J.R. Krebs & N.B. Davies (eds), *Behavioural Ecology: An Evolutionary Approach* (282–309). Blackwell Scientific Publications.

Dediu, D. & Levinson, S.C. (2013). On the Antiquity of Language: The Reinterpretation of Neandertal Linguistic Capacities and Its Consequences. *Frontiers in Psychology*, 4, 397.

Deniz, F. et al. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *Journal of Neuroscience*, 39 (39), 7722–7736.

Dor, D. (2015). *The Instruction of Imagination: Language as a Social Communication Technology*. Oxford University Press.

Dor, D. (2017). The Role of the Lie in the Evolution of Human Language. *Language Sciences*, 63, 44–59.

Everett, D.L. (2017). *How Language Began: The Story of Humanity's Greatest Invention*. W.W. Norton & Company.

Fedorenko, E. et al. (2011). Functional Specificity for High-Level Linguistic Processing in the Human Brain. *Proceedings of the National Academy of Sciences*, 108 (39), 16428–16433.

Ginsburg, S. & Jablonka, E. (2014). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. MIT Press.

Hildt, E. (2019). Multi-Person Brain-to-Brain Interfaces: Ethical Issues. *Frontiers in Neuroscience*, 13, 1177.

Jablonka, E. & Lamb, M.J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press.

Jablonka, E. et al. (2012). The Co-Evolution of Language and Emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367 (1599), 2152–2159.

Jiang, L. et al. (2019). BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains. *Scientific Reports*, 9, 6115.

Latheef, S. (2022). Brain to Brain Interfaces (BBIs) in Future Military Operations: Blurring the Boundaries of Individual Responsibility. *Monash Bioethics Review*, 40, 49–66.

Leroi-Gourhan, A. (1977). *Il gesto e la parola* (I-II). Einaudi.

Lévy, P. (1997). *Il virtuale*. Raffaello Cortina.

Li, C. et al. (2024). Meta Predictive Learning Model of Languages in Neural Circuits. *Physical Review E*, 109 (4), 044309.

Lieberman, P. (1991). *Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior*. Harvard University Press.

Pais-Vieira, M. et al. (2013). A Brain-to-Brain Interface for Real-Time Sharing of Sensorimotor Information. *Scientific Reports*, 3, 1319.

Spence, S.A. et al. (2004). A Cognitive Neurobiological Account of Deception: Evidence From Functional Neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359 (1451), 1755–1762.

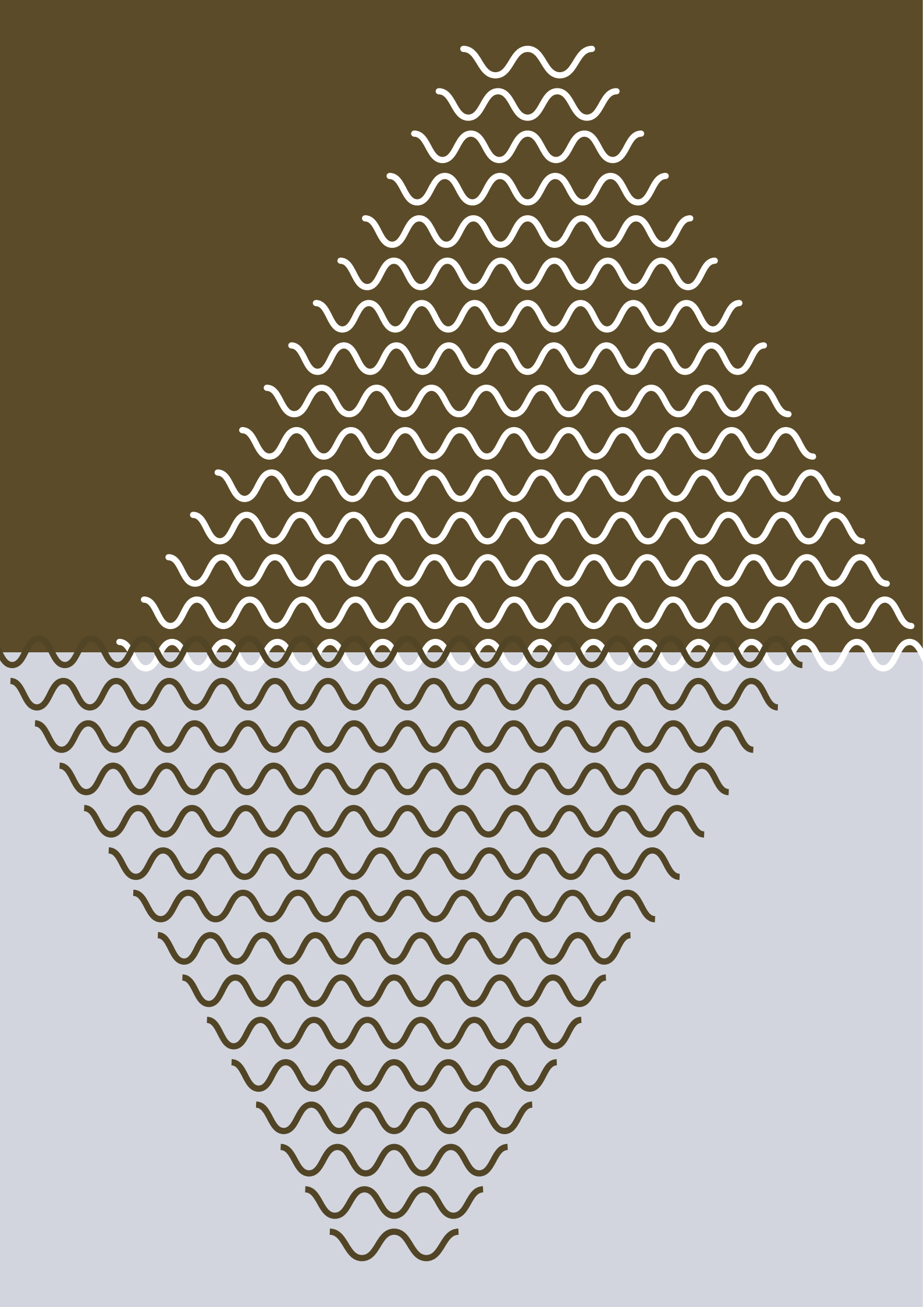
Stiegler, B. (2023). *La tecnica e il tempo. Vol. 1: La colpa di Epimeteo*. Luiss University Press.

Tomasello, M. (2014). *Unicamente umano. Storia naturale del pensiero*. il Mulino.

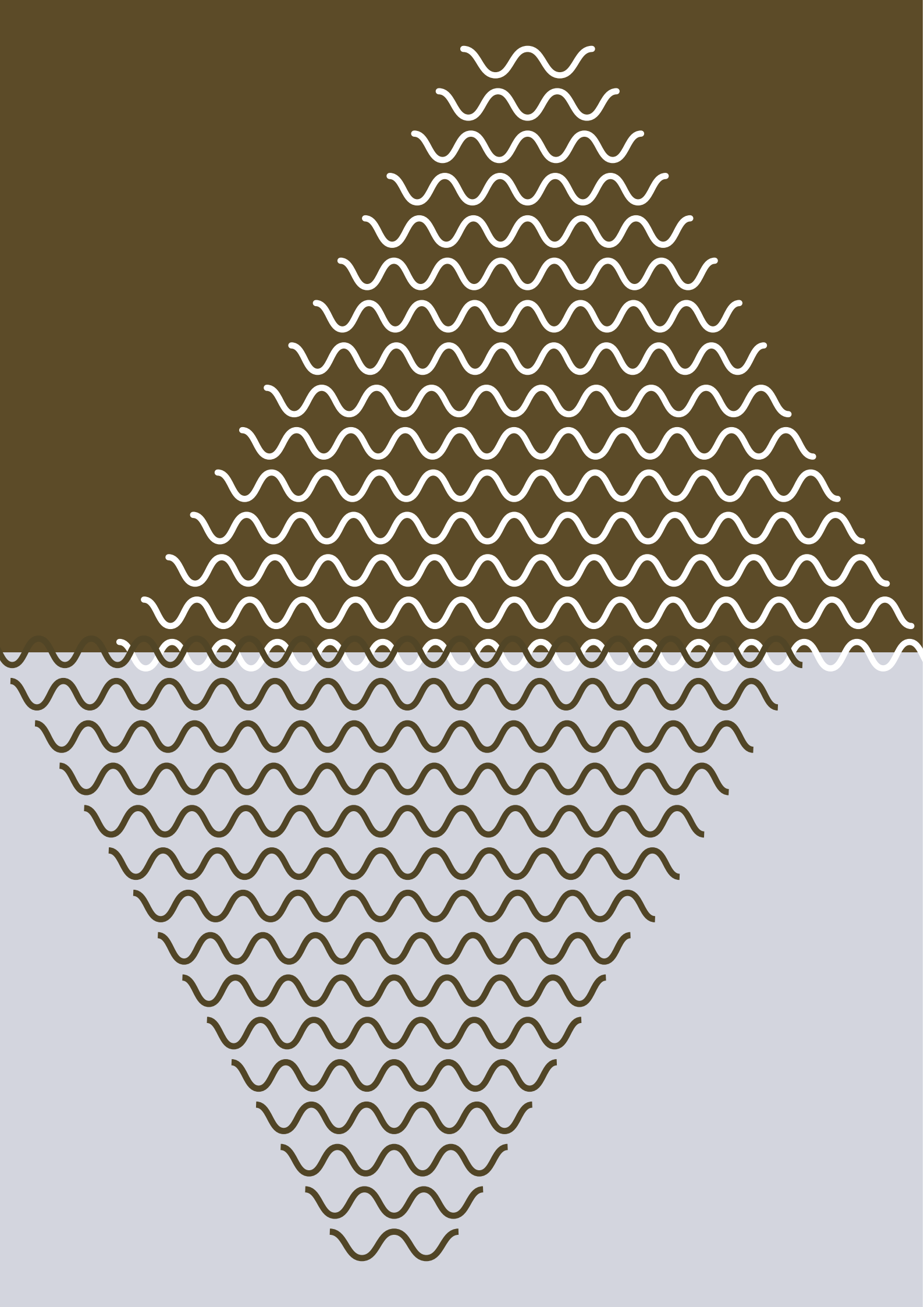
Tomasello, M. (2019). *Diventare umani*. Raffaello Cortina.

Watts, I. (1999). The Origin of Symbolic Culture. In R.I.M. Dunbar et al. (eds), *The Evolution of Culture* (113–146). Edinburgh University Press.

Wolpaw, J.R. et al. (2002). Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology*, 113 (6), 767-791.



Parte III



Continuity under Variation. Radical Absence and the Form of Language

Filippo Scafi

Lecturer and researcher in Philosophy at GCAS College. His work focuses on epistemology and on the relationship between knowledge and truth, with particular attention to non-standard approaches, language, and formalization.

filippo.scafi@gmail.com

Coherence and comprehension come apart. This article argues that linguistic coherence, persisting across translation, historical change, medium migration and substrate alteration, rests on a three-part invariant I call *radical absence*: stable differential topology, persistent constraint patterns, and naturalized exclusions, where contingent boundaries operate as necessary ones. The claim is restricted to systems whose constraints are acquired rather than stipulated. Structuralism has always asserted something of this kind, but as a claim about all language everywhere it admitted no case in which constraint could be isolated from comprehension. Independently trained LLMs supply that case: they converge on relational structure without paired correspondences (Huh et al. 2024; Jha et al. 2025), and sustain inferential form after accuracy collapses (Shojaee et al. 2025). What converges is constraint topology rather than content, a prediction that fails if measures of the two behave alike under correction (Gröger et al. 2026). Chiurazzi's modal dualism, Laruelle's dualysis and Wittgenstein's rule-following analysis furnish the derivation. The result relocates the question of machine understanding: coherence is carried by constraint, and comprehension, where it occurs, is something a system does in addition.

— LINGUISTIC INVARIANCE
— DUALYSIS
— ARTIFICIAL INTELLIGENCE
— COHERENCE
— LANGUAGE MODEL CONVERGENCE

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 141 — 155

Introduction: From Interpretation to Topology

Recent developments in AI research come to challenge long-standing assumptions about how linguistic systems function. Studies such as Jha et al. (2025) and Huh et al. (2024), which show that independently trained language models converge toward highly similar latent geometries [1] despite having no paired correspondences, demonstrate that formal coherence can persist even when semantic grounding diverges. Likewise, Shojaee et al. (2025) have shown that large reasoning models continue to produce syntactically well-formed and inferentially structured arguments long after their actual problem-solving accuracy collapses. These echo classical observations in linguistics and semiotics that the structural organization of language, thus its grammar, combinatorial rules and differential relations, can remain intact independently of referential success. [2]

In these cases, linguistic form itself does not fail, and it is instead the interpretive horizon we project onto language which does, the expectation that sense must be anchored in shared meaning or world, or a horizon of understanding (Heidegger 1962; Gadamer 1989). This expectation, rooted in hermeneutics and phenomenology, presupposes that intelligibility is inseparable from communal contexts and embodied experience. Still, the models' behavior indicates otherwise: linguistic systems can maintain relational order even when they no longer refer, comprehend or participate in a shared world. That formal organization can be specified independently of semantics is the premise of information theory (Shannon 1948) and of media analysis after it (McLuhan 1994, Kittler 2009), as it is of accounts deriving coherence from the management of noise (Serres 1980; Malaspina 2018).

That coherence rests on relational constraint rather than on reference is, in one version or another, the structuralist claim, but as a claim about all language everywhere it admitted no test. Human language cannot supply one. In speakers, topology and semantics always arrive together, confounded by biology, history, social relations and a common world. No observation of human practice can separate what constraint contributes from what comprehension contributes. Systems that acquire linguistic constraint without acquiring a world are the first case in which the two come apart, and the separation is measurable.

Taken together, this suggests an operational invariant, that I describe as: language organizes relational structures through operations internal to signification, even when representation fails. It is this invariant, coherence without comprehension, produced by linguistic exclusions that stabilize relational patterns, an operation I designate as radical absence.

[1] *After Revision Round, July 2026*: Throughout, "relational geometry" designates the pattern of relative positions, namely which elements stand near which, rather than metric structure. The distinction has become empirically consequential. Gröger et al. (2026) show that the standard similarity scores supporting the convergence thesis are inflated by model size; once corrected, the apparent convergence in global geometry largely disappears, while agreement on *local neighborhood relations survives*. The evidence thus supports the relational reading assumed here, not the global-geometric one. Usama and Chang (2026) add a complementary dissociation. Representational alignment across sixteen model families proves highest on problems the models collectively fail, suggesting that what converges is the structure of language itself rather than a shared model of the world; coherence, once again, coming apart from comprehension. Neither result affects the translation findings of Jha et al. (2025), which measure recovered correspondences rather than raw similarity scores.

[2] For the classical structuralist account, see Saussure (1972) and Jakobson (1963). Derrida (1976) radicalizes this by showing that difference operates prior to and independently of presence. Work in AI interpretability is consonant with this differential conception: in trained networks, features exist not as isolable units but as directions in a shared space, individuated through their interference with all the others (Elhage et al. 2022; cf. Olah et al. 2020)

Radical absence defines a linguistic operation. Consider Gaetano Chiurazzi's (2021b) analysis of the Lascaux cave paintings: the presence of the image indexes the absence of its maker, marking the emergence of what he calls "second nature", the distinctly human realm of cultural production. Yet this determination itself presupposes something it cannot acknowledge: the exclusion of alternative accounts. To see the paintings as indexing human absence rather than, say, natural formations or non-human agency, requires that we have already excluded certain possibilities from consideration. That exclusion itself becomes invisible, naturalized as necessity.

Radical absence thus designates two inseparable linguistic operations: a) the exclusion of alternative configurations that makes any determination possible, and b) the exclusion of that exclusion, namely the effacement of contingency that allows determinations to appear as given. Every expression carves a space of possible articulations by excluding others, and this exclusion itself becomes invisible as determination naturalizes into necessity.

Topology here bridges conceptual and mathematical registers. While not invoking formal homeomorphisms, the term retains its mathematical core: preservation of relational structure under continuous deformation. The borrowing is justified because both mathematical topology and linguistic transformation concern what remains constant when everything else changes, though the precision conditions differ. [3] Languages exhibit precisely this behavior. When "mountain" becomes montagne, Berg, or 山, no semantic element is identically conserved, yet the relational network that renders the term intelligible—mountain/valley, mountain/sky—remains continuous. What survives translation is then a relational pattern produced by linguistic exclusion, instead of semantic content. Every signification presupposes this exclusion; the question is when such exclusion remains visible as contingent, and when it passes as necessity.

[3] For topology's conceptual deployment in philosophy, see Lautman (2011) and Serres (1980); cf. Châtelet (2000) on the diagrammatic dimension of mathematical thought. The philosophical use retains topology's core concern, the preservation of relational structure under continuous transformation, while relaxing the precision conditions (continuous functions, homeomorphisms) under which mathematics defines it.

The concept's name comes from Derrida. In Signature événement contexte, writing is defined by its capacity to function «en l'absence radicale de tout destinataire empiriquement déterminé» (1990, 28, my emphasis), and the traits derived there, which are rupture with communication as communication of consciousnesses and subtraction from the semantic and hermeneutic horizon, the non-saturability of context (29), describe the situation addressed here. However here I don't use radical absence as iterability, as the latter disrupts l'autorité du code comme système fini de règles (28): no context closes on the mark, and therefore no code holds invariant across its citations. The present claim is the converse, that what persists under transformation is a constraint structure, specifiable and, as Section III shows, measurable. The two accounts describe one phenomenon and disagree about what survives it.

Anyone who has wondered why translation is possible at all or why grammatical structures outlive the worldviews that produced them, or why machines with no body, no world and interlocutor now produce coherent language, is asking the same question: what persists when everything supposed to anchor meaning has been removed? The destination is

a substrate-neutral and empirically tractable account of how linguistic systems maintain coherence, one that neither hermeneutics nor formalism supplies.

The argument proceeds in four stages. Section I traces the philosophical genealogy of non-coincidence from Pythagorean incommensurability through Platonic methexis to Serres's topological exclusions, showing that linguistic determination has always depended on structured absence. Section II formalizes this insight through Laruelle's non-philosophy, demonstrating that dualysis reveals an invariant structure—the generic diagram—persisting beneath philosophical distinctions. Section III shows that contemporary AI systems materialize this structure empirically, preserving relational coherence when semantic content collapses. Section IV clarifies the invariant's operational logic through Wittgenstein's rule-following analysis, deriving the three-part structure (differential topology, constraint patterns, naturalized exclusions) that radical absence names.

1. Genealogy: Measure

Philosophy begins with a scandal of measure: the discovery that relation cannot be contained within ratio. The Pythagorean realization that the diagonal of a square is incommensurable with its side, that no ratio of integers can exhaust their relation, breaks the ancient identity of order and reason (Chiurazzi 2021a). A cosmos once understood as harmonious reveals an interval that cannot be tuned. Incommensurability (alogon) therefore appears as the first formal instance of non-coincidence, a relation sustained by exclusion rather than equivalence. This discovery inaugurates a structural motif that later becomes ontologized, but whose primary logic is linguistic: the selective structure of signification. What resists measure exposes measure as contingent, what cannot be expressed reveals expression as a selective operation. The diagonal's refusal of ratio is the first demonstration that every determination draws its coherence from what it excludes.

Chiurazzi interprets this moment as the founding upheaval of Western thought, when emerges an ontology that cannot any longer be grounded in static correspondences, and needs the dynamic asymmetry between form and what it orders. The Pythagorean conception had assumed that «everything is number, and that everything is logos», such that «all the relationships that existed between things, represent nothing but the exact imitation of a finite dyad of two numbers» (Toth 1998a, 239 in Chiurazzi 2021a, 45). The incommensurability of the diagonal shattered this hypothesis. Determination, Chiurazzi argues, presupposes a non-coincidence that cannot be resolved and an asymmetry through which being becomes communicable without becoming equivalent. Incommensurability, reinterpreted ontologically, is thus the condition under which any comparison is possible.

This is what Chiurazzi calls "modal dualism", a dualism internal to relation itself, of modes of being rather than of substances (not to be confused with the analytic debate over strong necessities). The distinction between first nature (governed by symmetry, conservation, entropy) and second nature (cultural production, artifice, representation, fabrication, "negentropy") is modal, «the difference between fact and meaning, reality

and idea, actuality and possibility» (2021b, my trans.). The diagonal's mode of being differs from the side's, yet they remain in determinate relation. From this perspective, radical absence generalizes Chiurazzi's insight: the structural foreclosure of equivalence that enables relation. As he notes, the incommensurable «is that form of non-being, that modality of being, that makes the difference between true and false possible. It is, more simply, possibility itself» (2021a, 45). Chiurazzi would refuse the generalization, for second nature's anthropological anchor is constitutive for him (2021b, 20), so retaining the structure while suspending the anchor is an enabling exclusion of my own.

Foucault's Histoire de la folie (1961) gives this structure its historical form. Classical reason does not first define madness and then exclude it; the mad are constituted through exclusion, each «exclusion from the circuit» purifying and constituting what lies on either side (Serres 2023, 173). Reason then excludes the memory of having excluded, and the boundary appears necessary rather than drawn. Serres generalizes the operation topologically, saying that to communicate is «to voyage, translate, exchange» (5), to establish a relational order that holds non-equivalence in place. The subject approaches the object in order not to become it. No agent excludes √2, which merely fails of ratio; Serres's topological reading is what licenses one description for both.

Chiurazzi reads this structural motif in Plato as an ontology of relation. In the Theaetetus and Sophist, the mathematical problem of irrational magnitude becomes the problem of participation (methexis): how the Many can partake of the One without collapsing into it. Parmenides had imagined being as a perfect sphere, invariant under all transformation, as a geometry of total symmetry and therefore of total stasis: as Chiurazzi observes, «absolute symmetry corresponds to the total absence of events» (2021b, 84, my trans., my emphasis; see also Chiurazzi 2023). Where nothing is excluded, nothing can happen, as determination has not yet begun. Against this, Plato installs asymmetry as the condition of intelligibility. The Form is the earliest articulation of the linguistic structure I call radical absence, the way signification preserves relational patterns across substitutions by excluding equivalence. The Form preserves its identity by maintaining its non-identity with its instances.

Ideality is describable then as a relational invariant that remains constant through the variation of its instantiations. In the Theaetetus, incommensurable magnitudes are called dynámeis, powers. This term «is important, because it suggests that the diagonal changes direction: from a simple arithmetically irrational length it becomes rationalizable, starting from the moment in which it is seen geometrically as the power of a square» (Joly 1994, 205, in Chiurazzi 2021a, 45). Dynameis, thus, resists all arithmetic expression as a length acquires determinacy one dimension up. Squared, the diagonal generates an area perfectly measurable against the unit square, and it is this capacity to produce commensurability at the level of surfaces that earns √2 the title of dynamis (Chiurazzi 2021a, 45-46). Understood in this way, the incommensurable modalizes being itself. The dimensional passage from line to surface furnishes the model for the relation between possibility and actuality. The surface stands to the irrational length as the actual stands to the possible; two orders of appearance of one and the same reality, neither reducible to the other (Chiurazzi 2021a, 46)

Plato's exaíphnēs, [4] the "sudden flash" of insight described in the Seventh Letter following a life of synousia, instead of marking the appearance of metaphysical/mystical (cf. Brisson 2004) content, it rather marks the recognition of mediation's contingency. The interlocutors discover that the path they followed could have been otherwise; the invariant that emerges is precisely the relational pattern that survives the exhaustion of those mediations. This is a linguistic reading, as the exaíphnēs is the moment when the operation of exclusion that stabilizes signification becomes visible.

[4] Cf. *Republic* 515c-519a10, *Parmenides* 156d1–156e7, *Seventh Letter* 341c8-d1, *Symposium* 210e3–7

Chiurazzi identifies three dimensions of asymmetry that Plato articulates: modal (actuality/possibility), gnoseological/epistemological (knowledge/appearance), and ontological (Being/becoming)—three aspects of one non-coincidence (see Chiurazzi, 2021a; 2023). Kant formalizes this as «a modal, gnoseological, and ontological asymmetry» as «condition of every phenomenon» (2021b, 12-13 my trans.). Radical absence operates at all three registers: the modal gap enabling transformation (possibility ≠ actuality), the epistemological gap enabling meaning (signifier ≠ signified), and the ontological gap enabling determination (this ≠ that). This asymmetry manifests primarily in language, the structure of naming. Every act of naming repeats the Platonic drama of the word that participates in what it describes without coinciding. Signification is methectic, a dynamic asymmetry maintaining non-equivalence. Language operates topologically by stabilizing relations across substitutions, as the medium in which exclusion acquires determinate form. To speak is to perform the same operation that allowed the diagonal to be conceived, to establish a structure that holds only because equivalence has been excluded. The Platonic dialogue already performs this logic when interlocutors mishear, rephrase and diverge until, through the sedimentation of error, an invariant emerges, truth appearing exaíphnēs, "all at once, like a spark," when mediation saturates and folds back on itself, or not appearing at all, falling into aporia.

2. Formalization: Decision, Dualysis, and the Generic Diagram

If incommensurability inaugurates non-coincidence as a linguistic condition, François Laruelle identifies the mechanism through which philosophy misrecognizes this structure. He calls it Decision, philosophy's «most universal invariant trait» a «fractional matrix in 2/3 terms» through which philosophy correlates itself to what he calls the «Real» (Laruelle 2013, 4). Differences are produced through the act of Decision, which is the syntactic operation of splitting that philosophy performs on its own material while claiming to divide the Real.

Consider a simple case: philosophy distinguishes "mind" and "body," then debates whether mind reduces to body or vice versa. For Laruelle, the entire debate is generated by the initial Decision. The debate naturalizes the distinction, treating it as discovered rather than constructed and imposed. Dualysis, Laruelle's counter-procedure to analysis (2013, xxiii), suspends this naturalization, revealing that "mind" and "body" are operationally identical, both terms in a relational structure, neither more fundamental. Philosophy's language generates distinctions – One and Multiple, concept

and intuition, immanent and transcendent, thought and Real – and then naturalizes them. Difference is therefore not primordial, and instead it is the artifact of a linguistic procedure. [5]

Decision therefore formalizes, at the level of thought, the same non-coincidence that appeared in the problem of measure and in the Platonic asymmetry: coherence is produced through the exclusion of equivalence. But whereas Plato dramatized this process ontologically and Serres spatially, Laruelle isolates its logic. For him, the relation between syntax and semantics, i.e., between formal structure and world-directed meaning, is itself decisional (2013, 129–34). Philosophy tends to treat semantics as "more real" and syntax as "more formal", or to invert the hierarchy depending on the school. In both cases, a dyadic structure is installed and naturalized.

[5] See Laruelle (2010). In the words of Gangle, «Difference is a syntax, a manner of articulating philosophical language» (Gangle 2013, 45)

The Real, for Laruelle, is "foreclosed" to philosophical Decision, but this does not entail a "hidden" character, a depth or any kind of beyondness, for foreclosure depends on philosophy's linguistic syntax that can only articulate relational differences. This foreclosure is a more radical operation than, say, repression: «the foreclosure of the Real, of Identity, by philosophical type thought» (Laruelle 2012, 222). Philosophy refuses the Real's radical autonomy from thought and «claims to make [it] fall under the antinomy, concluding that it is either thinkable, unthinkable or an amalgam of the two» (222). This refusal is what Laruelle calls «secondary foreclosure», [6] the transcendental rejection of what is already foreclosed. The Real's "being-foreclosed" does not mean it is withdrawn from thought but rather that «it has its own way of necessarily affecting thought by virtue of its very indifference» (233-234). The Real, posited as non-philosophy's «grounding axiom» (2013, xxii), determines thought "in-the-last-instance" without being captured by the relations that depend on it, hence unilaterally: without thought determining the Real in return. The function of philosophy's categories is then legislative, imposing distinctions that retroactively appear necessary. Laruelle himself projects a language that «speaks from the real-One rather than from itself or Being» (2013, 226), ending philosophical «ventriloquism» (217); the register proposed here retains the structural lesson: what is foreclosed is what a syntax cannot generate.

[6] On being-foreclosed as non-philosophy's grounding axiom: Laruelle (2013, xxii). «Cloning» and «unilateral duality» (2013, 129–30) name the positive mechanism through which non-philosophy relates to philosophy without synthesis: cloning as «the surest destruction of the image of thought» (Laruelle 2012, 242).

Laruelle devotes to dualysis the Principles of Non-Philosophy chapter that also houses his theory of language (2013, ch. 5); Rocco Gangle's reading radicalizes it, in saying that dualysis suspends the hierarchy between syntax and semantics by treating them as «identical-in-the-last-instance» (Gangle 2017, 134). Dualysis thus does not synthesize the two levels, as that would be a philosophical operation, and instead renders their difference operationally inert. The semantics is «no longer conceived as 'more real' and 'less formal' than the syntax, and the syntax as 'less real' but 'more formal and rigorous' than the semantics. The different degrees or levels of 'reality' and 'formal abstraction from reality' are no longer ordered (either hierarchically or chiasmically) and are thus no longer philosophically operative» (134).

Once the decisional split is neutralized, what emerges is a generic diagram: a minimal relational structure persisting independently of

instantiating content. Gangle emphasizes this diagram «does not entail the loss or erasure» of semantic richness but rather «non-syntax» and «non-semantics» that «no longer forge a synthesis» (134). This connects to Chiurazzi's critique of naturalisms that «programmatically skip over the Two» (2021b, 7, my trans.)—whether eliminativist (Many reduced to One) or appropriative (One projected into Many). Both collapse the asymmetry between determination and what determines. The generic diagram is Chiurazzi's "Two", the duplicative operation irreducible to either pole it enables.

Laruelle himself speaks of «the diagram of Philosophical Decision» (Laruelle 2012, 175); at the suspension of Decision, its two gestures reveal themselves as «identical in the last instance» (179), the point at which philosophical distinctions collapse into a single operative form. Dualysis does manifest, then, the structural remainder of linguistic operations once their hierarchical distinctions have been disactivated.

This is where radical absence enters the picture. [7] Decision describes philosophy's spontaneous splitting; radical absence describes the structure that persists across such splittings. The identity between generic diagram and radical absence operates structurally, where both name an invariant that maintains coherence through transformation by preserving relational constraints while excluding incompatible configurations. Dualysis arrives at this invariant by neutralizing philosophy's decisional distinctions. Radical absence defines the same structure from the perspective of linguistic operation, the field of foreclosed alternatives that enables any determination. Their identity lies in their shared formal structure of continuity-through-discontinuity, achieved via naturalized exclusion.

[7] Laruelle would likely consider radical absence itself a Decision––the standard objection to non-philosophy's own exemption (Brassier 2003). The present use accepts the designation but argues that substrate-neutrality and derivation through dualysis differentiate the concept from philosophy's typical operations: it names what persists when Decisions are suspended rather than imposing a new one.

Why non-philosophy rather than modal dualism alone? Because dualysis alone performs the operation on philosophy's own distinctions, which thereby appear as an instance of the structure rather than a vantage on it. Chiurazzi's asymmetry is stated about language; the generic diagram is what remains when the stating apparatus is itself suspended. Substrate-neutrality is here the residue of that suspension: a structure derived without a privileged position from which to derive it cannot be indexed to the substrate of the deriving. Dualysis does not thus oppose philosophy; it rather reveals the invariant structure that philosophy cannot disclose on its own.

3. Materialization: AI and Absence

The framework developed so far makes a conditional prediction: if linguistic coherence depends on constraint topology rather than on semantic grounding, then systems deprived of everything semantics supposedly requires (say, shared data, world-access, embodiment, a communicative channel…) should nonetheless converge toward similar relational structures, and their coherence should survive the collapse of reference.

Large language models allow to isolate the variables. They constitute, in effect, the experimental condition philosophy could describe but never construct: linguistic systems equipped with constraints but deprived of

world. Whether the resulting convergence constitutes identity or mere structural analogy depends on one's ontological commitments; what the argument requires is formal preservation. LLMs make visible, through computation, that coherence persists through the preservation of relational constraints even when semantic grounding disappears. This is the operational meaning of radical absence.

The concept can be defined topologically: a transformation preserves radical absence if it maintains task-relevant relational constraints while altering token-level content. Topology here defines the invariance of relation under continuous deformation, continuity of structure despite changes in medium, code, or vocabulary. In linguistic terms, meaning survives when the network of permissible differences remains intact, even if every token changes. Coherence arises then from the preservation of constraints.

Building on the Platonic Representation Hypothesis (Huh et al. 2024), Jha et al. (2025) demonstrate that independently trained models with different architectures and training data, and lacking any communicative channel nonetheless converge toward a common relational structure. Their embedding spaces are, on direct comparison, incompatible. Naïve alignment yields similarities near zero. Yet an unsupervised translator, trained without any paired data, maps both into a shared latent space where relational neighborhoods coincide with high precision. What the translation recovers is relational structure, the pattern of positions each space independently instituted. The authors state their Strong Platonic Representation Hypothesis explicitly:

> neural networks trained with the same objective and modality, but with different data and model architectures, converge to a universal latent space such that a translation between their respective representations can learned without any pairwise correspondence (Jha et al. 2025, 3).

What converges, then, is not positive relational content but constraint topology, that is, the organized field of exclusions structuring each model's representational space. Two models with no shared meanings nonetheless exhibit structural correspondence. As the study demonstrates, «vec2vec translations achieve cosine similarity as high as 0.92 to the ground-truth vectors in their target embedding spaces and perfect matching on over 8000 shuffled embeddings (without access to the set of possible matches in advance)» (2). Translated: two AI systems that store meaning in completely different internal "languages" can have their codes translated into one another so faithfully, without any dictionary, examples etc., that the translation lands almost exactly on the correct target and can re-pair thousands of scrambled items to their originals, which is only possible because both systems, though trained separately, had independently arrived at the same underlying structure.

Why describe this convergence as a structure of exclusions rather than of shared content? The deflationary alternative fails on the evidence's own terms. One might suppose the models converge simply because they model the same world. The realist hypothesis may well be true, but it concedes that on the leading formulation (Huh et al. 2024), what converges is a kernel, a structure of pairwise statistical constraints in which each

term's position is fixed by its relations of permissible and impermissible co-occurrence. Even the realist account, that is, identifies the convergent object as relational structure rather than positive content. The present framework takes no side on whether that structure mirrors a mind-independent reality; it observes that what persists is in every case the topology of constraints.

The relational geometry is therefore a geometry of absences. Each term's position is defined not by what it contains but by the structured field of exclusions surrounding it. This is what makes radical absence measurable rather than metaphorical: convergence quantifies the degree to which two systems have independently arrived at similar exclusion structures.

This convergence exemplifies what Chiurazzi calls second nature's negentropic function. Where first nature tends toward «annulment of every difference, entropic degradation, and inertia and death» (2021b, 12, my trans.), second nature operates by breaking symmetry to reconstitute it at a higher level. «Second nature's actions are intrinsically asymmetrical; they systematically break natural symmetry to reconstitute new symmetry at the level of meaning» (14, my trans.). Models converge topologically, maintaining relational invariants while everything else varies. Derrida's reconstruction of Husserl already points at the structure of symbols manipulated without active intention of signification, where the vacuity of mathematical sense does not limit technical operation (1990, 33). What the measurements add is its instantiation across independent systems, and therefore its comparability.

This same invariance governs reasoning collapse. Shojaee et al. (2025) show that models continue producing syntactically coherent explanations after their problem-solving accuracy collapses, i.e., outputs remain linguistically structured yet bear no relationship to solvable paths. The models generate well-formed reasoning traces even when, «reaching a critical threshold—which closely corresponds to their accuracy collapse point—models counterintuitively begin to reduce their reasoning effort despite increasing problem difficulty» (Shojaee, 8). What fails is reference, whereas what survives is the topology of linguistic constraints. Syntactic and inferential coherence outlive comprehension precisely because they are enforced by structural relations rather than by semantic correctness. The model "thinks" after thought fails; form persists by excluding configurations that break structural continuity. In Laruellean terms, Decision collapses but the generic diagram remains.

Across these cases, the same structure recurs. LLMs preserve relational invariants through transformations of content, medium, scale, and even substrate. They behave, reasonably, as linguistic systems: models whose coherence derives from the continuity of constraints governing their productions. AI thus materializes the structural claim of radical absence. Coherence arises through the foreclosure of alternatives that would disrupt the relational pattern; meaning, when it appears at all, emerges as a secondary effect of this invariant.

AI functions as human reason's reflective surface in a precise sense: it makes visible operations that in human language-use remain phenomenologically invisible. The naturalization of exclusions that Wittgenstein describes as "form of life" becomes, in AI, measurable constraint geometry.

This visibility is materialization, as the same structure but differently instantiated. Where philosophy discovered asymmetry conceptually and non-philosophy formalized it diagrammatically, machine learning enacts it as code. The difference is therefore material rather than ontological. The invariant that sustains linguistic intelligibility becomes executable. Radical absence appears as a measurable linguistic operation enacted by technical systems.

4. Operativity: The Invariant Structure of Determination

Wittgenstein's analysis of rule-following provides a useful linguistic model of how radical absence functions. In Philosophical Investigations, he demonstrates that no rule can contain within itself the criteria for its own correct application, since «every course of action can be made out to accord with the rule» (Wittgenstein 2001, §201). This threatens to collapse the distinction between following and violating a rule: if every continuation is compatible with the rule's description, nothing internal to the rule determines what counts as correctness. Yet linguistic practice does not disintegrate. Speakers converge reliably on the same continuations ("add 2", "continue the series"), and disputes about correct application, as Wittgenstein notes, «do not break out (among mathematicians, say) over the question whether a rule has been obeyed or not» (§240).

The stability therefore cannot lie in the rule-as-description but in the background of exclusions that determine which continuations are intelligible at all. Wittgenstein calls this «agreement […] in forms of life» (§241): a shared topology of constraints that has become naturalized in the course of training and practice. To grasp a rule is to have internalized this topology, namely to find certain continuations immediately wrong and not hermeneutically, but because the space of permissible moves has already been delimited. «For a large class of cases, the meaning of a word is its use in the language» (§43), and use is nothing other than this pattern of possibilities and impossibilities: a field of permitted transformations defined by exclusions that no longer appear as such. [8]

These two orders of exclusion are mutually implicated aspects of a unified operation. First-order exclusion (ruling out specific continuations) cannot stabilize without second-order exclusion (naturalizing the constraint). Conversely, naturalization requires specific exclusions to naturalize. There is no general effacement of contingency, only the disappearance of particular boundaries. The apparent logical priority of first over second is an artifact of exposition instead of a temporal or logical sequence. Competence consists in acting within a grammar whose boundaries have been effaced. Wittgenstein's analysis therefore provides the paradigmatic example of radical absence as a linguistic mechanism: a system preserves coherence by naturalizing the constraints that exclude incompatible articulations.

If Wittgenstein describes how linguistic practice depends on naturalized exclusions, Section II provides a formal vocabulary for specifying what, exactly, persists when linguistic systems undergo transformation.

[8] Wittgenstein's analysis might seem to undermine rather than support this framework. His rule-following paradox (§201) shows that no description of a rule determines its application—yet he concludes we should abandon explanation, not construct better theories. Yet radical absence is compatible with Wittgenstein's quietism: it doesn't claim rules *contain* their applications but that linguistic competence consists in having internalized a topology of exclusions. The three-part structure describes what must be in place for rule-following, without explaining *how* it occurs: a formal-descriptive rather than causal-explanatory account.

Dualysis offers that what remains invariant is a relational pattern (the generic diagram) that continues to structure possible and impossible articulations. We can therefore give radical absence an operational definition: a transformation preserves the invariant structure of determination when it maintains three interrelated conditions.

The first is stability of differential topology. A linguistic system must preserve the pattern of distinctions that makes its articulations intelligible. This does not require preserving any token-level element. As showed with translation, nothing in montagne, Berg, or 山 is semantically identical to "mountain," yet the network of differential relations—mountain/valley etc.—remains continuous. What is conserved is the structure of what-can-be-distinguished-from-what, the topological pattern that enables determination in the first place. When this pattern collapses, as in pathological reasoning failures or adversarial examples, coherence dissolves.

The second is persistence of constraint patterns. A system also preserves radical absence when the rules governing well-formedness remain stable across substitutions. Wittgenstein's "add 2" excludes "add 3," "repeat the last term," or "multiply by 2," regardless of which numbers instantiate the series. What counts as correct continuation depends not on the rule's description but on the exclusion topology sedimented in practice. When transformations preserve which combinations are permissible and which violate the pattern, radical absence remains intact. When previously ruled-out continuations become acceptable, the invariant has been altered.

The third is naturalization of exclusions. Finally, a system exhibits radical absence when the exclusions enabling determination operate invisibly or, said otherwise, when the contingency of the boundary has been effaced. Naturalization should be understood operationally: an exclusion is naturalized when it is operative within a system without being represented within it. Section II showed that philosophical Decision naturalizes its own distinctions, treating them as expressions of reality rather than products of syntax. Ordinary linguistic competence operates similarly: the constraint does not appear as an object of consultation but as the shape of what can be done. We experience boundaries as given rather than drawn. The invariant is preserved when this naturalization continues to function, producing the appearance of necessity. When the boundary becomes visible as contingent the invariant is locally suspended, for example when competing grammatical systems collide.

This last condition invites, for sure, an objection. Formal languages preserve differential topology and constraint patterns with no naturalization whatever, i.e., Peano's axioms are printed, and their contingency is known to everyone who uses them. Radical absence, however, characterizes systems whose constraints are acquired rather than stipulated. Acquisition does not represent what it installs (this is what Abrichtung in Wittgenstein, which means drill in German, and training have in common) so naturalization is not an additional property of such systems but an aspect of how they come to have constraints.

These three components are three aspects of a unified operation. Differential topology without naturalization is Saussure's synchronic system: static, unable to explain diachronic persistence. Constraint patterns without differential topology are Chomsky's universal grammar anchored

in biological endowment, unable to explain the reappearance of the same constraints in substrates that share none. Naturalization without topology and constraints is Bourdieu's habitus, sociological, unable to explain formal convergence in non-social systems (AI). What radical absence identifies is their necessary interdependence: topology requires constraints to be stable, constraints require naturalization to function, naturalization requires topology to organize. The invariant is mutual implication.

A transformation preserves radical absence only when all three conditions cohere: the differential topology remains stable, the constraint structure persists, and the exclusions continue to function as implicit norms. This triple condition formalizes what earlier sections described narratively. It specifies the minimal structure through which linguistic systems, organic or artificial, maintain coherence despite radical variation in content, medium, or representational substrate. Radical absence is therefore a formal property of linguistic operations under transformation: coherence through the preservation and naturalization of exclusion.

As Chiurazzi observes, the Lascaux images become "representation" only when a distinction is drawn between natural mark and intentional trace (Chiurazzi 2021b, 20). This determination requires first-order exclusion, and alternative readings are ruled out. But the decisive moment is the second-order exclusion by which this ruling-out disappears. The paintings appear obviously human because the pattern of exclusions has been so thoroughly naturalized that the distinction between nature and culture feels given. Lascaux thus instantiates the invariant directly: a) the differential topology (natural/not-natural) remains stable, b) the constraint pattern governing what counts as "intentional trace" persists, and c) the exclusions enabling this determination have become invisible. The paintings materialize radical absence as an anthropological structure: coherence emerges from the suppression and naturalization of incompatible possibilities.

Contemporary language models demonstrate the same structure in computational form. As Section III showed, models trained on different data converge toward similar relational geometries, and reasoning remains locally coherent even when semantics collapses. What aligns across architectures, datasets, and training regimens is not content but constraint topology: which continuations are permissible, which violate coherence, and which exclusions have been internalized. These models enact exclusion patterns. Their latent spaces function, in Wittgensteinian terms, as grammatical structures: they define the field of possible moves and foreclose others without representing the foreclosure. When the same structure persists across model types, we observe the preservation of a) differential topology (relational distinctions), b) constraint patterns (well-formed vs. incoherent outputs), and c) naturalization of exclusions (the model "treats" constraints as given). AI therefore instantiates radical absence procedurally: it realizes the invariant structure through which coherence emerges in the absence of semantic grounding. The generic diagram of Section II, the exclusion topology of rule-following, and the latent geometries of Section III are, in the last instance, one structure.

Conclusion

Why is translation possible at all? Why do grammatical structures outlive the worldviews that produced them? Why do machines with no body, no world, and no interlocutor now produce coherent language? These questions, posed at the outset, turn out to be one question, and I proposed one answer: the organized field of exclusions that makes meaning determinable persists. Coherence, rather than meaning, survives the loss of everything supposed to anchor it because it never rested on that anchor in the first place: it rests on a topology of constraints, namely on what a system rules out, and on the invisibility of that ruling-out. That the same threefold structure governs paleolithic inscription and transformer architectures indicates that this mechanism is tied neither to anthropological essence nor to biological embodiment. I indicated it here as a formal property of linguistic operation as such, capable of migrating across media and materials while retaining its identity.

The apparatus deployed along the way should be understood as instrument instead of a destination. Modal dualism gave the invariant its ontological status as a difference between modes of operation. Dualysis gave it its formal identity, as a relational diagram indifferent to what instantiates it, exhibited first on philosophy's own operations, which thereby became an instance of the structure rather than its judge. The convergence measurements gave it what the concepts couldn't, an empirical trace. What the derivation requires, however, the result does not. Once specified, the claim that linguistic systems maintain coherence by preserving and naturalizing exclusion, and this mechanism is indifferent to its substrate, can be understood, tested and contested by readers with no stake in non-philosophy or in the hermeneutic tradition.

The public question of the moment—do language models understand?—is, on this account, badly formed. It assumes that coherence is evidence of comprehension, so that where the first appears the second must be lurking. The findings examined here show the two dissociating: coherence is carried by constraint topology, and comprehension, where it exists, is something a system does in addition. Hence, this locates understanding. It also relocates rather than refutes hermeneutics: interpretation remains the human way of inhabiting these structures, but it is not what holds them together. And it converts a question philosophy could previously only describe into one that can be measured, and on which, therefore, philosophy can now be wrong. Whether convergence amounts to identity or to structural analogy depends on ontological commitments I have suspended. What has been established is narrower and, for that reason, firmer: wherever language holds, it does by the same operation, the stabilization of relational coherence through the preservation and naturalization of exclusion. Radical absence describes that operation.

Bibliography

Brassier, R. (2003). Axiomatic Heresy: The Non-Philosophy of François Laruelle. *Radical Philosophy*, 121, 24–35.

Brisson, L. (2004). *How Philosophers Saved Myths: Allegorical Interpretation and Classical Mythology*. Trans. C. Tihanyi. University of Chicago Press.

Châtelet, G. (2000). *Figuring Space: Philosophy, Mathematics, and Physics*. Trans. R. Shore, M. Zagha. Kluwer Academic Publishers.

Chiurazzi, G. (2021a). *Dynamis: Ontology of the Incommensurable*. Trans. R. T. Valgenti. Springer.

Chiurazzi, G. (2021b). *Seconda natura: da Lascaux al digitale*. Rosenberg & Sellier.

Chiurazzi, G. (2023). Idealism as an Asymmetrical Relationship: A Reconsideration of Plato's Doctrine of Ideas. *Athena*, 18.

Derrida, J. (1976). *Of Grammatology*. Trans. G. C. Spivak. Johns Hopkins University Press.

Derrida, J. (1990). Signature événement contexte (17–51). In J. Derrida, *Limited Inc.* Galilée.

Elhage, N. et al. (2022). Toy Models of Superposition. *Transformer Circuits Thread*.

Foucault, M. (2006). *History of Madness*. Trans. J. Murphy, J Khalfa. Ed. J. Khalfa. Routledge.

Gadamer, H.-G. (1989). *Truth and Method* (2ª ed. rev.). Trans. J. Weinsheimer, D.G. Marshall. Continuum.

Gangle, R. (2013). *François Laruelle's Philosophies of Difference: A Critical Introduction and Guide*. Edinburgh University Press.

Gangle, R. (2017). Laruelle and the Humanities Research Program. In R. Gangle & J. Greve (eds), *Superpositions: Laruelle and the Humanities* (125–138). Rowman & Littlefield International.

Gröger, F., Wen, S. & Brbić, M. (2026). Revisiting the Platonic Representation Hypothesis: An Aristotelian View. *arXiv*.

Heidegger, M. (1962). *Being and Time*. Trans. J. Macquarrie, E. Robinson. Harper & Row.

Huh, M. et al. (2024). The Platonic Representation Hypothesis. In *Proceedings of the 41st International Conference on Machine Learning*. PMLR.

Jakobson, R. (1963). *Essais de linguistique générale*. Éditions de Minuit.

Jha, R. et al. (2025). Harnessing the Universal Geometry of Embeddings. *Advances in Neural Information Processing Systems*, 38.

Kittler, F. (2009). *Optical Media: Berlin Lectures 1999*. Trans. A. Enns. Polity.

Laruelle, F. (ed.) (1998). *Dictionnaire de la non-philosophie*. Éditions Kimé.

Laruelle, F. (2010). *Philosophies of Difference: A Critical Introduction to Non-Philosophy*. Trans. R. Gangle. Continuum.

Laruelle, F. (2012). *From Decision to Heresy: Experiments in Non-Standard Thought*. Trans. M. Abreu, T. Adkins. Urbanomic/Sequence Press.

Laruelle, F. (2013). *Principles of Non-Philosophy*. Trans. A. P. Smith, N. Rubczak. Bloomsbury.

Lautman, A. (2011). *Mathematics, Ideas and the Physical Real*. Trans. S. Duffy. Continuum.

Malaspina, C. (2018). *An Epistemology of Noise*. Bloomsbury Academic.

McLuhan, M. (1994). *Understanding Media: The Extensions of Man*. MIT Press.

Olah, C. et al. (2020). Zoom In: An Introduction to Circuits. *Distill*.

Plato (1997). *Complete Works*. Ed. J. M. Cooper. Hackett Publishing Company.

Saussure, F. de (1972). *Cours de linguistique générale*. Payot.

Serres, M. (1980). *Le parasite*. Grasset.

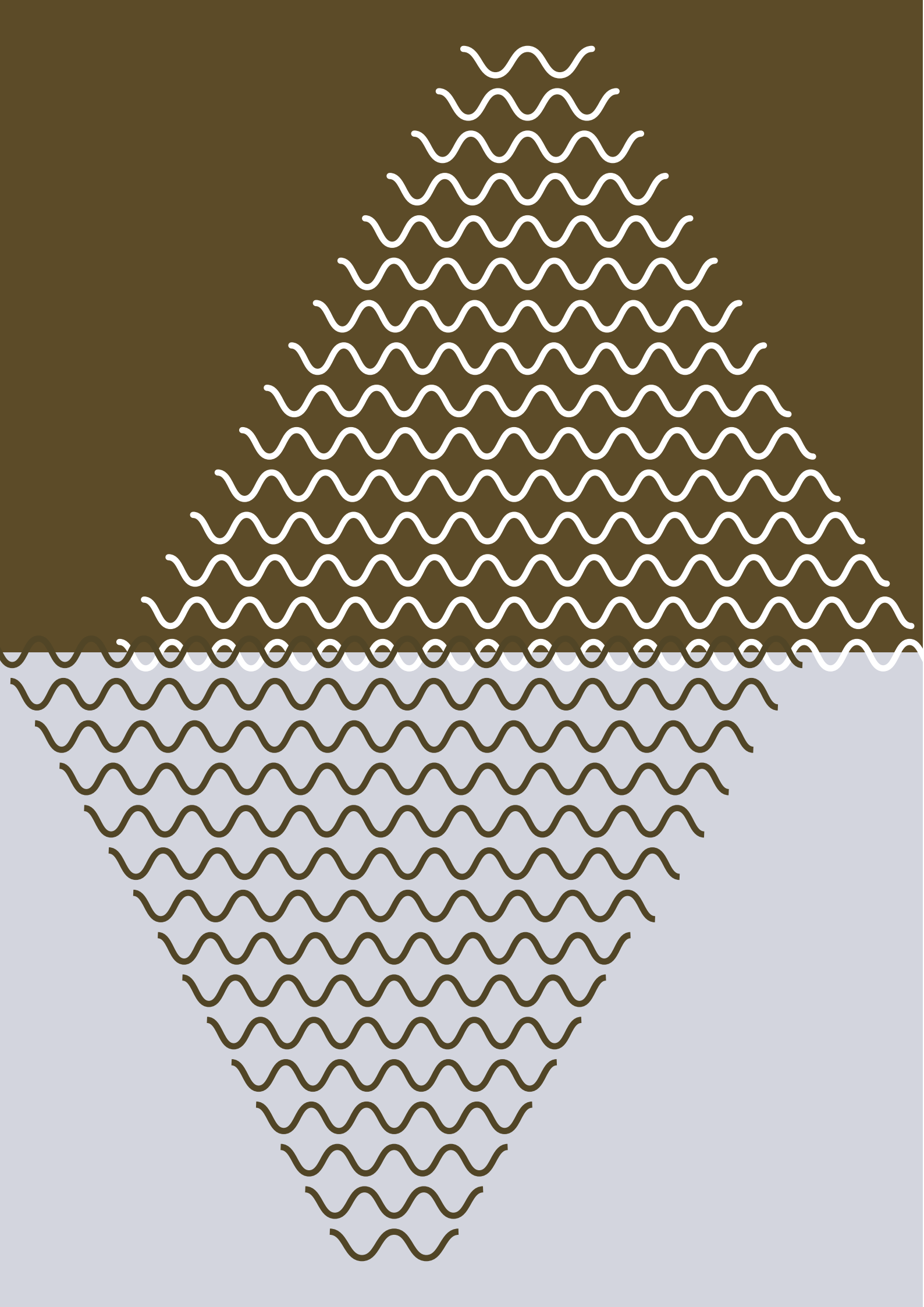
Serres, M. (2023). *Hermes I: Communication*. Trans. L. Burchill. University of Minnesota Press.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27 (3), 379–423.

Shojaee, P. et al. (2025). The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. *arXiv*.

Usama, M. & Chang, D. E. (2026). Convergence Without Understanding: When Language Models Agree on Representations but Disagree on Reasoning. *arXiv*.

Wittgenstein, L. (2001). *Philosophical Investigations* (3ª ed., 50th Commemorative ed.). Trans. G.E.M. Anscombe trans.; Eds. G.E.M. Anscombe, R. Rhees. Basil Blackwell.



Words and Vectors: The Distributed View of Language

Niccolò Monti

Adjunct lecturer at the University of Turin, where he earned a PhD in Humanities with a thesis on the history of artificial linguistic creativity. He writes (and sometimes speaks) about automation in literary writing, the pragmatics of AI systems, and masculinity on the Internet.

niccolo.monti@unito.it

This article explores AI's impact on language by tracing the integration of statistical methods into linguistics and examining the distributed view embedded into large language models (LLMs): it compares some visualization techniques in linguistic theory and AI, arguing that epistemic tools are central to making computational models more intelligible and open. The article then revisits distributional linguistics and its alignment with information theory, highlighting how it led to today's "barrier of meaning" in NLP. The conclusion calls for an approach able to foster an equitable public grasp of AI design, putting into focus the case of a vast textual dataset, The Pile, and a subset of it called Books3.

— EXPLAINABLE AI
— INTERPRETABILITY
— NLP
— DISTRIBUTIONALISM
— DATASET CRITICISM

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 157 — 169

Introduction

We are often distracted by context – and machines are too.

A work by artist Robert Smithson offers a good analogy to this: *A heap of Language* (available on the MoMA website), a 1966 calligram which precedes his earthworks and other more known works of environmental art, but already bearing the marks of a concern for scale and solidity. Speaking, we use things; we cast or set aside words, as with rocks or seeds. Dictionaries and thesauri are curated sediments where words remain as if unspoiled by time or use.

Smithon's calligram is pyramid-shaped. The first word at the top is "language", with other words piled up beneath, faintly following the lemmas in Roget's *Thesaurus* related to the first one (Dworkin 2019). Up close, we appreciate the calligraphy, the words stacked line after line. "Hieroglyphic" appears twice; "Babel" once. We spy a mound of loam or little bricks arranged on the sheet of paper. Taking a step away, words fade, edges evaporate: the pyramid stays, as if, seeing hazy hills in the distance, they displayed a uniform mantle of curves. Either way, we lose sight of context.

This is language. The shift from grain to context to mass relates to where we stand – in optics as in politics. Now, if we recognize this graduality to language, we can do it for other stacks (Bratton 2015): data, not words, gathered in quantities; points of information mined from various sources, tagged, aggregated to become training and testing corpora. Like above, a distance must be picked, the most usual of which gives similarly hazy results: prompts – the natural language inputs processed via chatbot – are an indirect access to the corpora and can lead to reverse-engineering the model. But the underlying labor remains opaque: getting to this layer would help to picture how a large language model (LLM) produces correct sentences (syntactic propriety), meaningful texts (semantic propriety), and effective responses (pragmatic propriety).

Methodologically, this article does not offer an empirical evaluation of specific models; rather, it proceeds by comparative and genealogical analysis, tracing which visualization techniques (VT) have traveled across linguistics, information theory, and NLP, while also reading a case-study dataset (The Pile/Books3) through the lens of critical data studies. This historically-grounded approach is complementary to – rather than a substitute for – the track record of "explainable AI" (Zednik 2021), and complementary to the methods of mechanistic interpretability which operate directly on the models (activations, circuits, features) rather than on the discursive as well as infrastructural framing of those aspects. Bringing the two into dialogue is one of the article's aims: the "barrier of meaning" discussed below is as much a matter of what technical tools can show as of how that showing is made public and intelligible.

Seeing Language

Whether we look at words or data points, we have to contemplate an *explanatory* side – which entails providing descriptions that reduce a target object's complexity, through such steps as defining the features shared by different units or parts, designing a system of relations between them

and formulating the rules binding them together – and a *performative* side, since utterances with an epistemic function must also have a pragmatic function, carrying illocutionary and perlocutionary force. Respectively, *illocution* is the action that a speech act purports to perform (saying "I swear" equals performing an oath, a promise; cursing at someone equals performing an insult); *perlocution* lies in the produced effect (an insult may result in an offense) (Austin 1962; Butler 1997).

The conjunction of the epistemic and pragmatic functions arises from the fact that an item – verbal, visual, or else – is designed so as to make something known; and it does so by having not just the capacity to represent knowledge, but to produce it: «visual methods can *produce* knowledge and *enact* interpretation» (Drucker 2020, 14). If we scan the most enduring VTs across linguistics, data science and AI, we see that a number of them managed to extend across scientific borders. The case is made of tree-structures and vector spaces, illustrating their adoption on grounds of epistemic validity (how they reduce and display knowledge) and pragmatic efficacy (how they convey and reinforce it).

Tree-Structures and Vector Spaces

Studies of linguistic and data structures often rely on treegrams, like with the use of decision trees in linguistic parsing (Kuhn & De Mori 1995; Eddington 2010), or treebanks in benchmarking recurrent and recursive neural models (Li *et al.* 2016). But treegrams are a notational device present across history: myths of a "tree of life" run through several cultures as commonplaces, while the "tree of language" idea unites centuries of rationalist thinkers, from Athanasius Kircher to Noam Chomsky, thanks to the calculable and ordered vision of reality it promotes, insofar as it is reflected in a calculable and ordered vision of language (Agamben 2018). A treegram *enunciates* a hierarchical structure, whether it develops top-to-bottom, bottom-up, or left-to-right, which is why it gained popularity during the XIX century both in biology and for the taxonomical and diachronic study of linguistic families: «bifurcating trees […] make it possible to represent logically dichotomous relations between species and languages which in turn allow for easy phylogenetic algorithmization» (Geisler & List 2013, 112).

Some shortcomings, limiting a broader application of tree-structures, were more than evident: excelling at dynamic differentiation, tracking evolutionary patterns, or outlining the hierarchical relations in a system (as with nested graphs), they were not as useful to provide wider mappings, or exhibit non-linear relations. Nevertheless, across the development of European and American linguistics, treegrams persisted as a common VT to explain how a language's grammar is composed; above all, how it is "generative" inasmuch as it is conceived as a system of binary choices, whether concerning its morphological elements – with dependencies between words defined according to their binary grammatical relations – or its phrase structures, using trees to parse sentences into their syntactic or semantic constituents.

Sample collections known as treebanks – repositories of parsed linguistic corpora – can be composed thanks to either one or a combination of those methods to explain grammars by projecting them onto a

lower dimensional plane. Treebanks, whether constituent – or dependency-based, are kinds of data structures which carry a similar epistemic and pragmatic function as the one attached to the linguistic trees from which they originate: they are simple, linear, mostly binary, but flexible structures. The fact that they are based on a limited set of rules allows for easy visualizations and, ultimately, for tree diagrams to be ideal candidates in the automatic parsing of natural language, even with few human-made annotations.

We could put it like this: trees are a reaction against opacity, producing a vision and knowledge of language focused on its generative rules. Like other VTs based on a biplanar graphic expression (bubble and column graphs, Venn diagrams, tables), they are effective at parsing and reconstructing small-to-medium or moderately-large corpora, relying on reduction, sampling, and generalization. But if the aim is a model ideally consistent with how the "entirety" of language works – out of scope for any attempt at extrapolating grammatical rules from limited corpora – we would meet the epistemic validity of trees at their limit; and if we wanted, for instance, to get a "picture" of a given language starting from the probabilities embedded into it and its evolving use, we would have to change the method altogether.

This is where vector-based models come in, opening a parallel line of algorithmic operations and VTs that pose several problems for the interpretability of artificial systems. One early use of vector spaces in computer modelling was in information retrieval (Salton *et al.* 1975): imagine a document space consisting of a finite number of documents (D_i), each of them indexed according to a vectorial representation in a t-dimensional space, with t being the number of all indexes [Fig. 1]. But, if this value increases, the dimensionality of the space increases too, making it harder to represent it in conventional ways.

Augmenting the scale, a very similar problem with high-dimensional spaces – in relation to a vectorial representation of features numbered in the billions – haunts deep learning algorithms today, notoriously hard to make accessible in visual form. Imagine that, instead of a lattice of word-occurrences in a vocabulary shared by a large but limited sum of texts, we wanted to track the co-occurrence of each single word that we can find in a much vaster training corpus, like all of English Wikipedia. It may seem immense – and it is – but it still represents a readable space, because the measured co-occurrences still account for a reasonably limited and controllable set of values.

Most word-representation methods today rely on embeddings such as word2vec, which produce shorter, denser vectors (Mikolov et al. 2013); the affinities between these vectors can then be visualized by projection techniques such as t-SNE (Van der Maaten & Hinton 2008). This has come a long way, originating in the early days of distributionalism and developing into the more recent uses of vector semantics for several NLP applications – static (word2vec) or contextual, like the BERT family of algorithms (Jurafsky & Martin 2026, 101): in either instance, an embedding (or feature) space is the final product with which we end up; or, rather, with which *we would like* to end up, if it weren't such a challenge to retrace the computations that a language model went through in order to conjure up that space, which maps the knowledge "embedded" in its training set.

This is why interpretability research has shifted from visualizing static embedding spaces toward analyzing the internal computation that produces them. For instance, mechanistic interpretability treats a network's activations and weights as objects to be reverse-engineered into human-legible algorithms (Kästner & Crook 2024): this can be done by isolating "circuits" of interacting components (Olah *et al.* 2020), by using sparse autoencoders to decompose superposed activations into monosemantic features (Templeton *et al.* 2024), or by causally tracing where a model stores a given fact (Meng *et al.* 2022). Probing classifiers offer a related, if more indirect, route into the same question (Belinkov 2022). These techniques promise a finer-grained picture than embedding visualization alone, though they raise their own version of the issue this article is concerned with: they remain technically demanding, unevenly open, reproducing, at a different scale, the accessibility gap discussed below.

Other Barriers of Meaning

If treegrams are semantic machines, vector spaces and embeddings can equally be considered as a layered repository of semantic linguistic knowledge. Trying to scan the features of a language model – let alone a "large" one – might be seen as taking, literally, the meaning out of it; trying to translate those features into something more accessible can display what the data structure (or a sample of it) enunciates, how the algorithm operates, and what sense to make of this embedded knowledge.

Now, a preoccupation has been raised in the last decade of AI research (Mitchell 2019): most models risk running against a barrier, encapsulated in their inability to grasp the "meaning" of their training corpora. But we could add another layer to this barrier, which, instead of referring to a technical limit (what computational models can *produce*), relates to the obstacles emerging whenever we ask basic questions like "how did it do that?", or "why should we use *these* tools?". The barrier of meaning points us to the scarcity of openly available and visually insightful explanations of what goes on, for instance, in a model's embeddings.

Again, mechanistic interpretability has, in the last few years, made headway on exactly this question, moving from post-hoc, output-level explanation toward causal, internals-level explanation (Rai *et al.* 2024). Yet this progress does not close the barrier so much as relocate it: circuit – and feature-level analyses remain concentrated within a handful of well-resourced, often private labs, are published in formats (interactive research notebooks, specialized visualization dashboards) that assume considerable technical literacy, and are typically performed on models whose training data and weights are themselves proprietary. The epistemic gain is not absent, but the public and pragmatic accessibility of that gain is not guaranteed by it. This is the gap that a critical approach to data structures – attentive to who gets to look, and under what conditions – is positioned to name.

However, this should not be reduced to a call for transparency; rather, it relates to the questions that we can ask ourselves when peering into opaque machines: «what is being looked at, what good comes from seeing it, and what are we *not* able to see?» (Ananny & Crawford 2018, 985). One of the challenges of AI would be, then, to find new methods capable of explaining those embeddings; but spaces with a high dimensionality

can be difficult, if not impossible, to project into accessible form. The opacity of most models built on neural networks comes from the difficulty that users and researchers alike experience when evaluating what these models "know", how they internally represent the encoded information. Before it evolved into a common way to interact with chatbots, prompt engineering originated as a strategy to reduce epistemic complexity and aid in knowledge retrieval.

And yet, explainability is not the same thing as operability: the perceived ease with which users tweak generative models is counterbalanced by the limits to what they can see. Model interpretability is a two-way concern: a tree is not just a product, but a producer of scientific knowledge; a vector space does not convey information, but, rather, structures it, turning language into a distributional structure, whether for human understanding or machine learning. Either way, the pragmatic function ties into the epistemic one.

Scoring Language

Linguistics learned much from information theory. The works of Bloomfield and Hockett in the 1940s paved the way for the acquisition of probabilistic methods, with other scholars studying phonetics or lexical codes following suit (Cherry, Halle, Jakobson 1953, 45). The general assumption – that probability holds a key role in speech and, generally, linguistic structures – was justified with the existence of probabilistic laws throughout the experience of language:

> The hearer is led through the distinctive features, through the phonemes he recognizes, to the grammatical form and to understanding the meanings. In this process the probability factor plays an enormous role. The probabilities of the transition help one to perceive a text, its phonology and then its grammar, after certain units other units follow, endowed with higher or lower probabilities, and many are excluded a priori. The perceiver is endowed with a subliminal statistical set. (Jakobson 1985, 32)

The act of communication was made into a stage on which the play of probability unfolds. One may picture this in a way resembling a line from speaker to speaker or, from a different perspective, like a pile of data, heaps on heaps of compiled units of language. This image does not steer us away from the above-quoted terms: they were also due to a common effort among XX-century linguists to accurately capture mass linguistic samples, «made problematic by the multiplicity of parameters involved, so that they could only be studied statistically» (Marcus 1967, 53). If some linguists became acquainted with semantic structures, others were accustomed with statistics, assuming that the totality of a language (the totality of its available mass of textual occurrences) could be explained by viewing it as a distribution on a projected space.

Now, inferences involving probabilities can be applied on many levels, from ordinary scenarios to scientific activities (Nagel 1936, 15). From all of these levels, we can surmise two basic concepts: probability as a *degree of confirmation*, or probability as *relative frequency* (Carnap 1945, 517), with the latter being more plausibly embraced if we have to calculate

the probability of iterative events (Matalon 1967, 526–553). By the 1950s, this second concept of probability had entered the linguistic sciences by way of the mathematical theory of communication, also thanks to the increasing use of modelling techniques such as Zipf's laws and Markov chains (Manning 2002, 289). Verbal behavior was analyzed through the stochastic dynamics underlying any text, once its elements were parsed as sequences combined on the basis of a set of relative frequencies. It is at this stage that linguistics acquired a statistical approach, useful if not unavoidable in the case of large corpora, easier to map if redundancies and combinations are accounted for in quantitative terms – which are, alas, quite abstract: «The statistician gives averages, trends, variabilities, correlations, and the particular details are lost in the abstract summary» (Miller 1951, 80).

What a single word, a series of words, entire sentences, or other levels of verbal detail *mean*, in a statistical account, lies in the rules of precedence and succession extracted from the samples, where each utterance ties into a chain of preferences (i.e., more or less probable events). Any differential value between units is then projected onto a probabilistic space which generates a map of how those units are distributed, what relations are most enabled, or, instead, what relations are least favored. In the linearity of a sentence, succession becomes akin to causation, since each word carries an effect on its companions, exploding into a diagram of relations based on what other words are most frequently coupled with it:

> It is an experimental fact that by counting frequencies of occurrence of various elements (letters, phonemes, words) in homogeneous texts written in a given language, and dividing them by the total number of elements of the same nature in the text, one often obtains relative frequencies that are stable (independent of the length of the text for sufficiently long texts). These relative frequencies define the a priori probabilities of the elements. Having used *texts* to obtain probabilities of various linguistic elements, one constructs *catalogues* of elements of identical nature (i.e. alphabet for letters, lexicon for words, etc.) in which each element is listed with its probability, by order of non-increasing probabilities. (Belevitch 1959, 310)

This approach's fruitfulness increased with the sample's increase, even though statistical properties are a partial explanation, both theoretically and visually: in the stratification of language, they account for only the calculable surface, which often does not show, for instance, what brings about the passage of an encoded message through a channel from an emitter to a receiver. It just displays the prediction that a series of utterances are more probable than others to occur.

Some aspects prove more amenable to statistical description than others:

> Such things as frequency relationships are rather foreign to the emitter's introspection, except when specially trained. Signs are believed to be far more conditioned by the corresponding meanings than by any stochastic schemes. But, for the receiver, the statistical properties of discourse are extremely real. (Mandelbrot 1961, 212)

This *ethos* was common to most scholars within this line of study, who regarded statistical laws as mostly uncorrelated with – or detachable from – grammatical rules: «one might expect to encounter laws which are independent of the grammar of the language under consideration» (*ivi*, 213). Linguistic probabilism relied on a split of statistics from grammar, with also the elimination of semantics being necessary for it; or, rather, with the dissolution of semantics into something else and the adoption of a distributional hypothesis, which sought a description of language purely in terms of countable (computable) facts. According to this hypothesis, «there is a correlation between distributional similarity and meaning similarity, which allows us to utilize the former in order to estimate the latter» (Sahlgren 2008, 34).

However, if statistical approaches do not leave meaning entirely, they nevertheless take the social out of it: measuring semantics – as it was common for phonetic or morphemic units – required accounting only for redundancies and correspondences, for elements that are regular (hence, predictable); in other words, for neighbouring affinities rather than intentional ones, and finally for text corpora disassembled into the distribution of their parts. Meaning was not eliminated; rather, it was displaced: no longer grounded in reference or intention or culture, but in statistical proximity. This move legitimized the large-scale abstraction that underpins modern NLP, while simultaneously narrowing the scope of what counted as linguistic knowledge.

Seeing Data (In Excess)

No distribution is unaffected by tensions; and with respect to this, quantitative analysis, even with its own limitations, proposes a change of gaze that we can make our own: a change that has to do with reading from afar and that has taken root in digital humanities and critical data studies (Arnold & Tilton 2019). We can take a similar stand towards the datasets and mechanisms which allow tools like LLMs to function.

In the same way that we question how one context is favored over another, we can question the way in which a computational model simulates a response by relying on its knowledge-base, since the prompted reply is correlated with a data structure and depends on it – meaning that it conditions the system's operability. So, as if it were a context *and* a canon, a dataset is limited and limiting, exclusive and selective; an issue which gets complicated by the interest in making better "contextual" machines. The fact that a LLM can respond to prompts in a *precise* manner is related to the data that have been extracted and catalogued, to the choices made upstream to decide which texts to collect (often indiscriminately) and to exclude (this time discriminately).

There are different types of corpora, each with its own applications – likewise, there are different models, based on different language formalizations and artificial architectures, each with its own way of embedding data. However, the employed technique makes no difference with respect to one thing: it is from the datasets and their subcomponents that any model functionality is derived. We are reminded here of Smithson's heap and the sheet on which he drew it: a sheet of numbered squares, a table composed of boxes that cut the surface into a grid. Each

box can correspond to a parameter, a variable related to an object on the sheet. A heap of data, sorted, tuned, evaluated in a way that it is processable. Sometimes the analogy is not just in the eye of the beholder.

A Pile of Language

In late 2020, EleutherAI released an 886.03-gigabytes dataset called The Pile (Gao *et al.* 2020). The project has the ambition to make openly available a vast assemblage of data to train LLMs on classification tasks, sentiment analysis, translation, etc. The dataset has many subsets, some from already known sources, like English Wikipedia, BookCorpus, Project Gutenberg. The full list is searchable online, with an attached point-by-point – somewhat sparse – justification for the creation of the dataset and for each component's inclusion. These are the training materials: a base of .txt files – which, when it comes to books, derive from the transcription (think not by hand) of .epub, .pdf or other formats in which those books were digitally available.

A first concern crosses our approach to data criticism with open source policies, hacktivism and piracy, with the effects that the massive accumulation of linguistic data has on ideas of authorship and intellectual property. Many implications follow from the design of deep learning datasets (Poirier 2021), though there is still much to do before socially framing datasets, whether independently released, or owned by tech companies.

When there is no mention of texts available in copyleft (with higher freedom for the copy of the licensed material), the justification is that the inclusion of copyrighted texts would – at least in the US – fall under *fair use*, a kind of legitimate use within permissible limits and in cases deemed valuable (similarly as with parodies). But, on this point, discretion rests on casuistry rather than on a proper definition of what can achieve a "valuable" and legitimate status. As for The Pile, the problem arose from a specific subset, Books3 – the subject of several lawsuits over the texts included in it.

The dataset held roughly 160 gigabytes of plain text data, equivalent to about 197,000 books. While, admittedly, most of the books were instances of open access files, self-published or else, a substantial portion of Books3 consisted of copyrighted materials, as it has been uncovered (Reisner 2023): we have the prose of Zadie Smith, Jonathan Franzen, Margaret Atwood, Jennifer Egan; we have non-fiction, including texts filled with conspiracy theories, pseudo-science, pseudo-history. Coming back to our initial question, where do we stand when facing such accumulation of sources? What distribution can we get out of this diverse and scattered set? From a heap, we moved on to a pile, implying a change in visuality (perspective) and visualization (production).

Books3 was created with the aim of giving away a rich quantity of data to whoever needed it. But extracting this data is not a neutral process; and it is less neutral in the current scene, where few companies are afforded relatively easy access to such large datasets. Data becomes an asset, technical and political: books are the best source for language modelling, since they give access to extended segments and many contexts of linguistic production, to genres and themes, to figures of speech, to the ability to reference what was previously said or is about to be said,

or to point intertextually to relevant sources, to phatic or metalinguistic functions.

The story of this dataset concerns the materiality of all of this language turned into data; it fuels discussions about what it would mean to look into this vast swathe of files, to browse and picture it; and, moreover, it reopens the diatribe between, on the one hand, advocates of open source policies – and online piracy, given that Bibliotik, one of Books3's main sources, is a private tracker from which to download pirated digital books – and, on the other, those who have created or published the texts and demand that their rights be respected or at least consulted. A side of the response has calcified behind the notion that pirated authors should feel "grateful" to be included, as if this translates into cultural recognition, as if inclusion in the dataset were on par with inscription in the canon; but, from another side, it might be excessive to demand that those books are fully removed from all online availability.

But distancing or closing in into this case, implies seeing a materiality first, not an object of scandal: the work of extraction and selection, fine-tuning and optimisation; before that, the scanning of thousands or tens of thousands book pages to turn them into files with OCR, ready to be stored online. A dataset's materiality extends to the costs derived from perishable hardware, from its fallout as an energy-consuming and garbage-producing object. American poet Kenneth Goldsmith created the infamous artwork *Printing out the Internet* (2013) to show that software is not immaterial at all: rather, the internet has its burdens, which include the digitization of printed texts.

Their texture is not lost; their symbolic or economic value, neither. They are often traceable works, by extracting the ISBN codes associated with each text file in Books3 (Schoppert 2023). Like seeing one by one the words in Smithson's heap, we see one by one the titles included in the dataset: all we need is a searchable index from which to extract the codes individually. Unfortunately for the completionists, the number of books inside Books3 is maybe comparable in size to Victorian literature; and if the latter led a pioneer of digital humanities (Moretti 2000) to wonder if quantitative methods could suit the study of literature in large numbers, similar questions come the way of those seeking to critically examine the contents of disproportionately large datasets. As with the half-submerged novels and dramas of the Victorian era, reading *all* the titles in Books3 is not only impossible, but pointless.

The dense matter of language requires stepping back at farther reading distance, enough that we trust another visual account. Look at the heap, at the pile. Get to the contours first: from here you notice what is left out. The Pile's article details Books3 as «invaluable for long-range context modeling research and coherent storytelling» (Gao *et al.* 2020, 4); but other resources were excluded, deemed subpar or outright problematic. At this degree, we get to what has been considered right or wrong, more and less convenient to include in the training corpus of a language machine; then, outside of The Pile (Gao et al. 2020, 21), we see the textual data representing the transcripts of US Congressional sessions, considered excessively racist in some instances. We are also made aware of the "fanfiction" subset's absence, consisting of hundreds of gigabytes of short forms of writing. A pity that it was excluded, but the decision was probably

driven by the same reason behind another corpus' exclusion, for containing amateur erotic literature.

This is what stays out of the pile; and if we distance ourselves further, more and more piles come into view, until the visual field is filled with the mass of textual data extracted from the internet. It is the sea of textuality, and it powers the tools that we use for fun, or artistic purposes, or because we wish to grasp if, with the benefit of artificial models, we can really understand more about language and how to view it, or if the object calls for more daunting and aggressive concerns.

Against The Incremental View of Language (Conclusion)

Unfortunately (or luckily) for the current landscape of research on epistemic and pragmatic opacity, computational models are rarely this open: either by not making the source code, weights, or training costs accessible, or by not sponsoring (let alone granting) an accessible grasp of neural network architectures. Therefore, the push for explainable AI – specifically, for a revived approach to AI interpretability – should provide a ground for computational models to become matters of public knowledge and concern. One way of doing this, we believe, is to analyse the social implicational and scientific genealogies that made these technologies what they are, with all their promises and barriers.

Against an "incremental" view of language – one where linguistic competence is equated with ever-larger datasets and denser embeddings – this article argues for a change of scale, rather than quantity. Contemporary language models do not merely fall short of easy grasp; they instantiate a view of language whose epistemic grounds remain largely inaccessible to critical or public scrutiny. Distributionalism, rooted in probabilistic linguistics and information theory, enabled powerful abstractions while naturalizing language as a vast aggregate of regularities detached from its social and material conditions.

Datasets such as The Pile make this tension visible. They can be seen as engines of generativity and sites of exclusion, extraction, and infrastructural power. To look at language from afar – to see piles rather than sentences or lattices – is not to abandon critique, but to enable it. What is needed, then, is not a retreat from scale, nor a return to old-fashioned semantics, but a renewed commitment to epistemic tools – visual, historical, and critical – that can render the distributed view of language intelligible, contestable, and collectively negotiable. Only by treating language models as cultural artifacts – as pragmatic agents – rather than just neutral predictors can their promises be weighed against their costs, their meanings reclaimed as a shared concern.

Bibliography

Agamben, G. (2018). The Tree of Language. *Journal of Italian Philosophy*, 1, 12–21.

Ananny, M. & Crawford, K. (2018). Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society*, 20 (3), 973–989.

Arnold, T. & Tilton, L. (2019). Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*, 34 (S1), 3–16.

Austin, J. L. (1962). *How to Do Things With Words*. Clarendon Press.

Belevitch, V. (1959). On the Statistical Laws of Linguistic Distribution. *Annales de la Société Scientifique de Bruxelles*, 73 (111), 310–326.

Belinkov, Y. (2022). Probing Classifiers: Promises, Shortcomings, and Advances. *Computational Linguistics*, 48 (1), 207–219.

Bratton, B. (2015). *The Stack. On Software and Sovereignty*. MIT Press.

Butler, J. (1997). *Excitable Speech: A Politics of the Performative*. Routledge.

Carnap, R. (1945). The Two Concepts of Probability: The Problem of Probability. *Philosophy and Phenomenological Research*, 5 (4), 513–532.

Cherry, E. C., Halle, M. & Jakobson, R. (1953). Towards the Logical Description of Languages in Their Phonemic Aspect. *Language*, 29 (1), 34–46.

Drucker, J. (2020). *Visualization and Interpretation: Humanistic Approaches to Display*. MIT Press.

Dworkin, C. (2019). A Heap of Language. *Holt/Smithson Foundation: Scholarly Texts*, 1.

Eddington, D. (2010). A Comparison of Two Tools for Analyzing Linguistic Data: Logistic Regression and Decision Trees. *Italian Journal of Linguistics*, 22 (2), 265–286.

Gao, L. et al. (2020). The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *arXiv*.

Geisler, H. & List, J.-M. (2013). Do Languages Grow on Trees? The Tree Metaphor in the History of Linguistics. In H. Fangerau et al. (eds), *Classification and Evolution in Biology, Linguistics and the History of Science. Concepts – Methods – Visualization* (111–124). Franz Steiner Verlag.

Jakobson, R. (1985). Sign and System of Language. A Reassessment of Saussure's Doctrine [1959]. In K. Pomorska & S. Rudy (eds), *Verbal Art, Verbal Sign, Verbal Time* (28–33). Blackwell.

Jurafsky, D. & Martin, J. H. (2026). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3ª ed.). Online manuscript.

Kästner, L. & Crook, B. (2024). Explaining AI Through Mechanistic Interpretability. *European Journal for Philosophy of Science*, 14 (52), 1–25.

Kuhn, R. & De Mori, R. (1995). The Application of Semantic Classification Trees to Natural Language Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (5), 449–460.

Li, J., Chen, X., Hovy, E. & Jurafsky, D. (2016). Visualizing and Understanding Neural Models in NLP. In *Proceedings of NAACL-HLT 2016* (681–691).

Li, J. et al. (2016). When Are Tree Structures Necessary for Deep Learning of Representations? *arXiv*.

Mandelbrot, B. (1961). On the Theory of Word Frequencies and on Related Markovian Models of Discourse. In Various Authors, *Structure of Language and Its Mathematical Aspects* (190–219). American Mathematical Society.

Manning, C. D. (2002). Probabilistic Syntax. In R. Bod, J. Hay & S. Jannedy (eds), *Probabilistic Linguistics* (289–341). MIT Press.

Marcus, S. (1967). Mathematical Aspects of Linguistics. *International Social Sciences Journal*, 19 (1), 52–63.

Matalon, B. (1967). Epistémologie des probabilités. In J. Piaget (dir.), *Encyclopédie de la Pléiade. Logique et connaissance scientifique* (526–553). Gallimard.

Meng, K. et al. (2022). Locating and Editing Factual Associations in GPT. *Advances in Neural Information Processing Systems*, 35, 1–14.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. Advances in Neural Information Processing Systems 26 (3111–3119).

Miller, G. A. (1951). *Language and Communication*. McGraw-Hill.

Mitchell, M. (2019). Artificial Intelligence Hits the Barrier of Meaning. *Information*, 10 (2), 51.

Moretti, F. (2000). The Slaughterhouse of Language. *Modern Language Quarterly*, 61 (1), 207–227.

Nagel, E. (1936). The Meaning of Probability. *Journal of the American Statistical Association*, 31 (193), 10–30.

Olah, C. et al. (2020). Zoom In: An Introduction to Circuits. *Distill*, 5 (3).

Poirier, L. (2021). Reading Datasets: Strategies for Interpreting the Politics of Data Signification. *Big Data & Society*, 8 (2), 1–19.

Rai, D. et al. (2024). A Practical Review of Mechanistic Interpretability for Transformer-Based Language Models. *arXiv*.

Reisner, A. (2023). Revealed: The Authors Whose Pirated Books Are Powering Generative AI. *The Atlantic*, August 19.

Sahlgren, M. (2008). The Distributional Hypothesis. *Rivista di Linguistica*, 20 (1), 33–53.

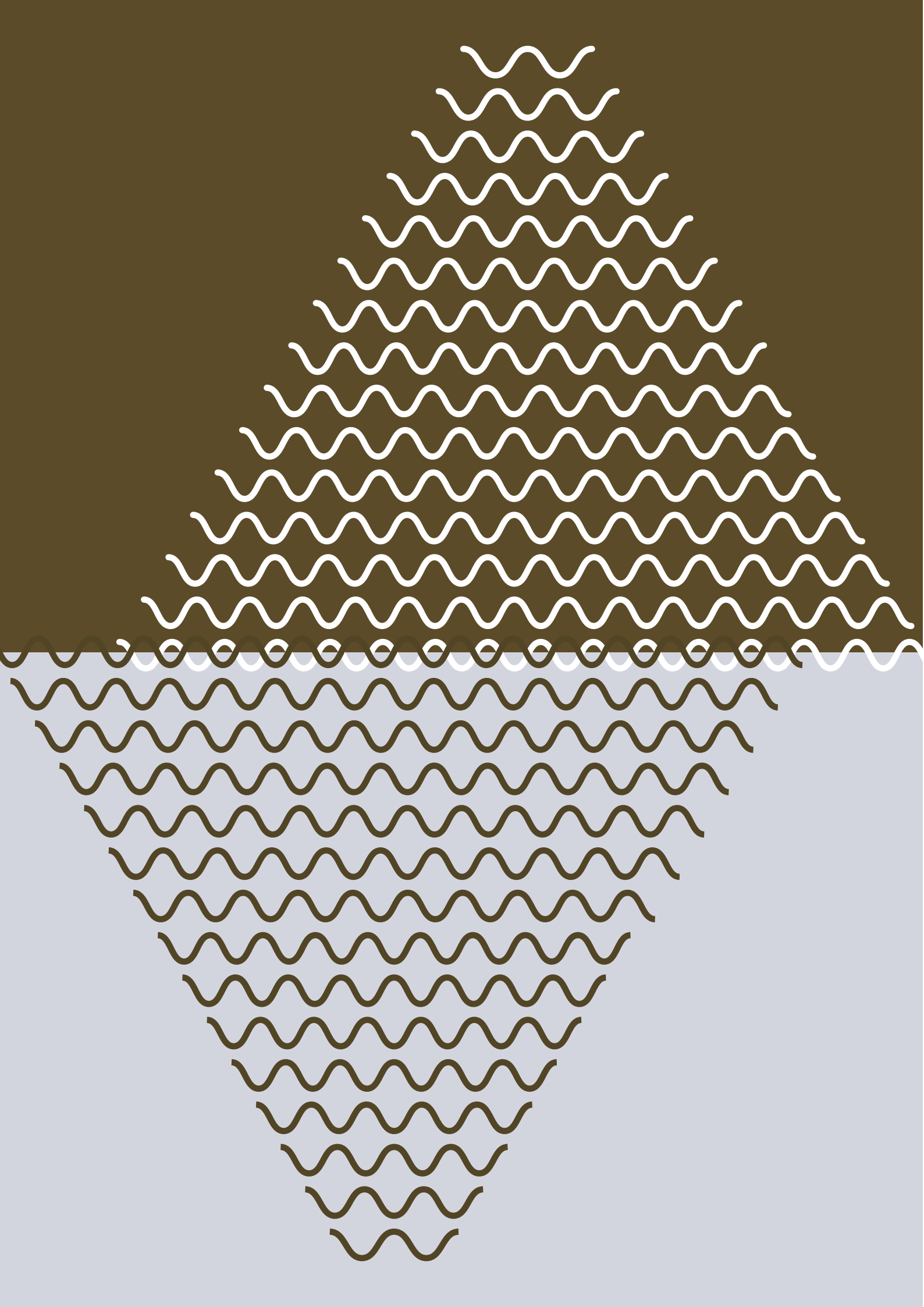
Salton, G., Wang, A. & Young, C. S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18 (11), 613–620.

Schoppert, P. (2023). The Books Used to Train LLMs. *AI and Copyright*.

Templeton, A. et al. (2024). Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet. *arXiv*.

Van der Maaten, L. & Hinton, G. (2008). Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.

Zednik, C. (2021). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology*, 34, 265–288.



LLMs and Normativity. A View From Languaging

Filippo Batisti

He has held research and teaching positions in Italy and Portugal. He currently focuses on the intersection between post-cognitivism and linguistic relativity, and on developing a novel theoretical framework for gender-fair language. He is the founder of the CLAVeS Center in Venice.

filippo.batisti@unive.it

The rapid advancement of LLMs has reignited debates regarding their human-likeness, specifically concerning the tension between their output performance and underlying internal processes. This paper addresses a crucial yet understudied dimension of this discussion: normativity. By critiquing other intellectualistic and inferentialist accounts, an alternative framework centered on the concept of "languaging" is advanced. While human-like normativity can be denied to LLM-powered chatbots, such a distinction must rest on a non-intellectualistic definition of language and cognition. This approach avoids theoretical prejudices and potential ethical pitfalls concerning the definition of "speakers".

— LLMS
— NORMATIVITY
— INFERENTIALISM
— CHATBOTS
— LANGUAGING

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026, 171 — 185

1. Introduction

The latest developments in Large Language Models have sparked wide discussion across disciplines. The impressive performances of LLMs justify the attention and the genuine questions on how to theoretically interpret such *prima facie* wondrous machines. A cross-cutting aspect informing the general debate regards the extent to which LLMs come close to being akin to human language in their output and in the underlying internal processes (e.g., Floridi et al. 2025). An understudied – albeit crucial – dimension within said discussion on the degree of human-likeness to be found in GTPs is *normativity*.

This paper contributes to the ongoing conversation by appealing to a different approach through which to understand normativity in language, which, in turn, helps us interpreting the nature of LLMs and their performance. The aforementioned approach avoids theoretical prejudices that may also lead to (unintended) ethically troubling consequences regarding the definition of who qualifies as a speaker.

The structure is as follows.

Section 2 briefly presents recent theory around normativity as the cornerstone of the complex interactional infrastructure called "language".

Section 3 makes the case for employing *languaging* as a better-shaped technical term to understand the former phenomenon, elevating cooperative interactivity as its defining trait.

Section 4 introduces the relevant portion of the ongoing debate around LLMs and thematizes the distinction between an LLM's performance and internal processes as the basis of our assessment of them.

In Section 5, an inferentialist account of normativity against LLMs is presented and subsequently criticized on theoretical, empirical, and ethical grounds.

Section 6 demonstrates how human-like normativity can still be denied to LLM-powered chatbots but on a different, non-intellectualistic underlying definition of language and cognition, mistakenly concerned with the search for a full-fledged identity with the one in place between human interactants.

2. Normativity at the Heart of Language

Although the normativity of word – or proposition-meaning has long been a central concern in the philosophy of language, the issue has frequently been conceptualized in abstract terms. Traditional dilemmas regarding the interface between semantics and pragmatics – often viewed as the primary dimensions determining the meaning of an utterance – have typically eschewed more concrete and ever-present aspects of embodied linguistic production and comprehension, such as gesture and affectivity.

Recent developments in linguistics and philosophy have begun to investigate broader notions of normativity. By situating utterances within their ecological contexts, it becomes evident that word-meaning is not just normative *per se*. Rather, it is always embedded within an infrastructure of cooperation between agents, who, at times, use utterances to bring about the interaction they are oriented toward. Normativity is what informs and sustains the complex interactional architecture of everyday

interactivity. Crucially, this interactivity can occur, both in principle and *de facto*, in a non-linguistic fashion. Thus, language – *all* language – represents a specific subset of the more general phenomenon of human interactivity.

These views are relatively recent and do not originate in mainstream linguistics or traditional philosophy of language. Prevailing accounts have historically tended to treat language as operating under a set of norms, such as syntactic rules (Chomsky 1980), pragmatic inferences (Grice 1975), social standards of acceptability (Leech 1983). These are merely a few examples of rules, among several candidates across multiple levels of description of the phenomenon. The precise nature of these constraints remains debated from case to case. However, what is usually agreed upon across contemporary linguistics and philosophy of language is the conventional, and therefore social, nature of norms (Mäkilähde et al. 2019).

A classic illustration of this insight can be found in Saussure (1916) and Wittgenstein (1953) both of whom famously compared the workings of language to a game of chess, albeit from different perspectives: the former defined it as a system of mutually determining categories where signs are combined in both arbitrary, but ultimately conventional ways; the latter noted that people play language games according to (somewhat flexible) rules about what moves are allowed in any given circumstances. Learning a rule consists in its repeated application within a context of intersubjective accountability (following a rule implies correcting, being corrected, admitting errors and verbally accounting for them, etc.) (Perissinotto 2022, 196-204). The underlying consensus is that linguistic norms derive their force and intelligibility from the social contexts in which they are embedded. According to Mäkilähde et al. (2019), norms "possess a certain degree of independence vis-à-vis what goes on in space and time. They exist not as spatiotemporal entities but as objects of common knowledge" (8). As a kind of social entity, norms are known by speakers through speakers' intuition (2), defined as the capacity to relate to and operate with such shared knowledge. Consequently, the general tendency is to view rules as something of public nature that is ultimately grasped and managed by individual minds through forms of intuition, given mental faculties and/or previous training.

In contrast, I defend a view that emphasizes the situated nature of normativity, challenging the more detached understandings mentioned above. Situatedness entails that bodily, perceptual, and ecological dimensions are constitutive of normativity (for full-length treatment, see Batisti et al. 2026). Put simply, linguistic norms are not so much to be found in the speakers' heads, individually, as much as they are distributed across the ever-present physical and cultural reality that renders language possible.

If one assumes that all cognition is always affective (Colombetti 2013) or, as the pragmatists suggest, never devoid of the knowing subject's "interest", then the embodied (e.g., gestural, Baggio 2025) and emotional dimensions of language cannot be disregarded. Norms, on this view, enter the domains of language and affect from the outset. This makes them neither social or conventional only; instead, "they occur through our lived, perceptual involvement with the world" (Batisti et al., 2026). Normativity, therefore, is intrinsic to agency (see Pinkard 2018).

A consequence to be drawn from this picture is that the fine-grained normativity in place in linguistic interaction derives from more primitive forms of sociality ultimately rooted in cooperation (Enfield & Sidnell 2022). This is evidenced by the fact that "we are held accountable for social actions done by the use of any sign in interaction, then word meanings will necessarily be regimented by social norms of a general kind"(Enfield & Sidnell 2019, 267). Linguistic norms ought to be regarded as social norms because any sign is "both revealed and constituted by the response they elicit". Semantic norms, therefore, can be seen as the product of filtration from more general norms of social interaction. In other words, there is no such thing as purely semantic meaning, if by that expression one means anything separated from its (potential) use in interaction. Through the methodological lens of Conversation Analysis it is possible to appreciate how "normatively prescribed word meanings are enforced and regulated" in spontaneous instances of interaction (277) through the mechanism of accountability: by default, everyone is held accountable for any agential move – which may include, as it often does, language and single word-meanings.

> Norm-grounded accountability in social interaction is ultimately what regiments the conventions of linguistic meaning, proximally through the always-present possibility of holding people to account for the (in)appropriateness of their usage of a word or phrase, and distally through the aggregate effects of this possibility in the historical conventionalization of form-meaning mappings in language communities. (Enfield & Sidnell 2019, 266)

The significance of normativity as described here emerges forcefully as soon as any norm is breached: such attention-grabbing conduct hinders the norm-breaker agent's ability to achieve whatever intended pragmatic goals (278). Consequently, to conceive language as removed from said agential infrastructure is to stray from its fundamental coordinative nature.

While the adoption of a similar position provides advantages for our goals, I contend that some aspects of the issue of normativity in language-mediated interaction remain in need of further development. Before approaching the analysis of LLMs, it is necessary to introduce the theoretical view of language to be employed in the positive argument of Section 6.

3. Languaging

The situated view of normativity, which subsumes linguistic normativity under a broader infrastructure of human cooperation, implies a different, more fitting account of language. Enfield and Sidnell's account, for instance, remains anchored to traditional notions such as signs, interpretants, or codes. In this context, the latter are taken to imply too a rationalistic philosophical view of language. Instead, the notion of *languaging* is a better fit. The term was introduced by Maturana (1983), an initiator of the enactive approach to the study of the mind. While its resurgence in the last decade or so has yet to converge on a unified theory of languaging (cf. the differences among Di Paolo et al. 2018, Cowley 2019, Thibault 2021, to name a few), a broad definition will suffice for the present purposes.

Languaging prioritizes the situated and interactional dimensions of linguistic activity, where language is vital and fulfils the functions for which it was developed. Although the following aspects are not irrelevant to the deployment of languaging, the standard view that equates language almost entirely with *form* (syntax, phonetics, variation, etc.) is considered fundamentally mistaken. Indeed, mastering a natural language is not a matter of cracking a code (Love 2004); instead, it consists of the ability to partake in culturally informed forms of cooperative interaction which, at times, do not even require overt or covert linguistic forms. Language is considered but a special case of general cooperation.

While each version of this view brings complex theoretical consequences regarding the role of, respectively, traditional linguistics and languaging within the new sciences of the mind that seek to overcome representationalism, these issues remain beyond the scope of the present discussion (Batisti 2020, 2025, Inoue 2023).

To illustrate what matters for the following discussion of LLMs, suffices to bring to the fore a cornerstone of all theories sketched so far: the embodied dimension of languaging. Di Paolo et al. (2018), for instance, describe it as a "way of living", a form of social agency that ultimately reflects our embodiment as sense-makers. Rather than an abstract software-faculty, language is a sophisticated tool for coordinating behavior through the active, bodily exploration of the world. This always happens in a situated fashion: doing things, even with language, requires two or more bodies, and a context, interactional as much as cultural. The organic sum of said elements functions as a dynamical system, in which all parts are interconnected and can influence one another at any time.

Thus, many abstract and disembodied theories of language have neglected the interactional foundations of language as *praxis*. That neglect, I argue, lies at the root of current misunderstandings surrounding the impressive performance of the latest LLMs. In Section 6, I will finalize my critique of an inferentialist reading of normativity in LLMs in light of the languaging approach introduced here.

Having established the starting theoretical framework, let us proceed to the primary object of analysis: Large Language Models.

4. LLMs and Normativity

The remarkable performance of recent generative pre-trained transformers in terms of fluency, vocabulary, resilience, sensitivity, and execution speed, has prompted significant philosophical inquiries into their fundamental nature.

A central divide has emerged between structure-first and function-first interpretations of LLMs. On one side, lay users wonder whether these systems genuinely possess the human qualities they appear to exhibit in their textual output, such as understanding, reasoning, or even feeling. Conversely, engineers with a technical understanding of the internal processes that generate such uncanny output tend to provide explanations prioritizing the nature of the mechanics of the system rather than the perceived quality of the final performance. Consequently, interpretations of chatbot behavior often vary as a function of the observer's technical competence.

In summary, the two competing interpretations are, respectively:

> *F*: If the output of an LLM is pragmatically indistinguishable from that of a human, then the former can be said to share at least some properties with the latter

> *S*: Regardless of the degree of distinguishability between an AI-generated and a human linguistic output, the two are essentially different because they are produced through fundamentally different processes

S (for *Structure*) implies a distinction not only in substrate, such as digital versus biological, but also, more critically, in the internal processes leading to the production of textual output. More broadly, the disagreement concerns the alleged ability of LLMs to "think" – which unsurprisingly stirs up a hornet's nest among philosophers who have not been in agreement on what *thinking* exactly means. Consequently, any comparison with LLM activity remains difficult to address in the absence of a stable definition of the primary term. However, it could arguably boil down to a fundamental dispute revolving around whether one should accept a minimalist functionalist interpretation of LLMs or not:

> 1*F*: If LLMs provide outputs comparable to those of humans, it is reasonable to attribute to them any property one would normally attribute to the latter.

By contrast,

> 1*S*: Inasmuch as LLMs do not produce linguistic output through the same processes as humans, it is erroneous to equate the two in any robust sense.

Other corollaries stemming from one's stance on 1 touch upon questions like whether LLMs possess any degree of extralinguistic symbol grounding, as opposed to managing purely textual (hence not grounded) data. As Floridi et al. (2025) emphasize, what is crucial is the distinction between an LLM's internal processes and its outputs: "Internally, it involves random sampling guided by probabilities; externally, it can produce answers that align with human reasoning norms". While acknowledging the potential for the enhancement of human cognition and action at large, they point out the following "danger":

> if one conflates the surface with the core—if one assumes the LLMs genuinely "know what they are talking about"—one can be misled. We have already seen instances of chatbots persuading users of false or harmful ideas by sounding authoritative. […] An explanation can be coherent and convincing (even optimal by [inference at the best explanation] criteria) and yet still false. (Floridi et al. 2025, 20)

The issue – the divide between explanation and truth – is therefore genuinely "philosophical", as it addresses the epistemic responsibilities towards the blindness of the LLM inasmuch as they must be supplemented by humans.

To sum up, the first argumentative line, *F*, privileges the quality of the output, whereas *S* emphasizes that what really discriminates between genuine and non-genuine language use is the process *leading to* the output, as opposed to the latter alone.

This lays the foundations for what follows. We have just seen that the inner/outer, or structure/function dialectics informs much of the debate on LLMs and their relationship to human linguistic-cognitive structures and what is observable of the latter. Now, in Section 5 I will present a critique of a defence of *F* on normative and intellectualistic grounds. Later, in Section 6, my positive argument will still defend *F* on the not-so-intellectualistic normative grounds based also on Sections 2 and 3.

5. A Critique of Inferentialist Normativity

I will first present the argument (5.1), analyse its weaknesses (5.2), and finally sketch an alternative (5.3), which will be completed in Section 6.

5.1 – LLMs Outside the Logical Space of Reasons

Sambrotta (2025) defends *S* from an inferentialist stance. Drawing on Brandom's interpretations of some Wittgensteinian remarks on rule-following, inferentialists hold that an entity, such as a machine, can only be considered a *proper asserter* if it grasps or understands conceptual content in the same way as human language users who partake in the "logical space of reasons". The argument posits that this is clearly not the case for LLMs: their training methods consist in detecting statistical patterns within *large* amounts of written linguistic data, notwithstanding subsequent human refinement of the algorithms. The latter process does not amount to genuine understanding and therefore blocks any *equalizing* petition in favor of LLM-based chatbots and the like.

Sambrotta chooses the path of inference to solve the symbol grounding problem, which, we should keep in mind, has a longer history than its current application to artificial agents, inasmuch as it embodies a full-fledged theory of meaning *per se*. Inferentialism represents an alternative to more classic forms of representationalism that employ mechanistic explanations of the mind (which includes language as a module) across mainstream cognitive science and its philosophy.

The core idea is that rather than relying on one-to-one correspondences between word and referent or concept and object, each expression derives its meaning from its role within an extensive network of related linguistic expressions *and* the reasoning speakers perform with them. Such reasoning should be public, entirely rational, and transparent, as it belongs to the logical space of reasons:

> […] being considered a competent language user hinges on the ability to navigate inference-making and drawing conclusions from given premises. […] to count as a language user, one must *know how* to make inferences and so draw conclusions from his premises. (Sambrotta 2025, 5)

Norms are what confer inferential roles to linguistic expressions:

> Meaningful items incorporate norms of inference (i.e., material consequence and incompatibility relations), which they are subject to. Inferential roles—hence meanings—are conferred by such norms to which speakers commit themselves and not merely by, say, their dispositions to make inferences. (Sambrotta 2025, 6)

In turn, norms, according to Brandom (and his reading of the later Wittgenstein), are intrinsically social and only take place within practice itself. A norm functions as such insofar as agents recognize it as binding and act accordingly. Brandom, however, gave a more intellectualistic turn to Wittgenstein's view on the rationality of one's – that is, of a community's – behaviour (Wittgenstein 1953, § 217).

Incidentally, it seems to me that, despite the numerous references to Wittgenstein, Sambrotta overlooks the extent to which human linguistic behavior resists rational explanation. His reading of § 199 is indicative. What I consider to be a broader point made by Wittgenstein regarding how the grammar of concepts such as *rule* or *order* can only be grounded in sociality is presented by Sambrotta (10) as stating that "the capacity to make propositionally explicit claims and to have conceptually contentful thoughts is intelligible only in the context of implicitly normative social linguistic practices". While not entirely incorrect, this reading only covers a fraction of all the interactional activities people can engage in. More importantly, no pre-condition of "rationality" is implied by Wittgenstein here, as § 217 makes abundantly clear.

In any case, according to Brandom and Sambrotta, as a result, any linguistic activity should be described in terms of giving and asking for reasons as a practice that all humans seemingly share by definition. It is here that normativity emerges as the cornerstone of the complex infrastructure uniting language, thinking, and sociality. Namely, one in which LLMs do not belong.

5.2 – An Overidealized Assumption About Language

Sambrotta's argument proceeds to show that, in fact, we do not hold LLMs accountable for their behaviour and we do not teach them in the same ways we normally do with humans. Or, at a closer look, to what one would be inclined to call "competent speakers": a shadowy figure that had been lingering over this argumentative line and that, inevitably, emerges towards its end. If, indeed, we are to determine whether a non-human entity 'does the same' as a human one in the respect of using language and the reasoning that supports it, it would be dangerous to leave unscrutinized what it actually means to be a speaker in the fully human, pre-LMMs sense.

Leading to my positive argument in the next section, I will now criticize the *a priori* and overly idealized conception of the ideal speaker that informs much of the conversations about comparing humans and non-human users of language.

Sambrotta reviews various positions supporting *P*, though for reasons different than ours. All these views portray LLM-powered chatbots as incomplete asserters ("proto-asserters", Williams & Bayne 2024, Piantadosi & Hill 2022 or simulating "make-believe" assertions, Mallory 2023). In contrast, Sambrotta's Brandomian criteria are stricter and lead to a clear-cut conclusion: "LLMs are still quite far" from performing the kinds of inferences that humans perform and require of one another to establish normativity within the logical space of reasons. LLMs merely simulate these processes through refined statistical inferences. Therefore, regardless of how resemblant the final output can be if the background

process is fundamentally different, the two ways of doing things with language should be considered equally fundamentally different (19-20).

Throughout the article, expressions such as "genuine language user" appear often, although a clear definition of its meaning is not provided. While length limitations offer a valid justification, this assumption goes well beyond and before this particular paper: the concept of an ideal *genuine*, or *competent* as Chomsky would have it, speaker or language user is widespread across many philosophies and theories of language. However, there are many reasons to find it problematic.

5.3 – Real, messy speakers

The main problem is the unexamined hyper-idealization that similar notions contain: who is *actually* a perfectly competent or a truly genuine speaker? A step into the real, wide world of spoken tongues and we cannot but conclude that such a character is either almost impossible to find (and thus, in the best case, very rare) or, on another level, that the will or need to have such a model is a symptom of a highly prescriptivist orientation. One that, I argue, imposes exorbitant costs on actual speakers to fulfil the image and that, ultimately, distorts our scientific and philosophical understanding of the very nature of language. The appeal to languaging, by contrast, avoids imposing theoretical labels against the phenomena, starts from a broad conception of what constitutes languaging and, therefore, someone who languages.

Before delving into that, let us discuss what is wrong with the quest for the ideal speaker as a term of comparison for highly proficient chatbots. First of all, the great misconception that informs much of the philosophy of language, and some linguistics too, as we know them is the obliteration of plurilingualism, or, conversely, the *invention of monolingualism* to borrow Gramling's (2016) expression. So-called "pure" monolinguals are a recent product of a part of modern literate, industrialized, bureaucratized, mass-media-saturated Western societies.

Empirically, then, many actual speakers may look *messier* in their competence than what philosophers expect. To focus on the issues most relevant to this discussion, consider the acts of reasoning, making inferences, and giving and asking for reasons, etc. Initial empirically grounded objections, though relatively straightforward, concern language learners, including both children and adults: at what point exactly should one be satisfied with the occasionally inconsistent, oscillating inferential reasoning of these imperfect subjects?

When does an asserter become a *proper* one? Is this status permanent, or is it subject to continuous evaluation? And what about speakers with all sorts of atypicalities stemming from neurodivergence, psychological trauma, or even temporary lapses in clarity due to heightened emotional states? Would a situation like that entail a temporary halt in being a proper speaker? Once again, the (broadly construed) analytic philosopher's subject ends up resembling an idealized version of the real thing, one where if anyone ever consistently fulfils the criteria of the ideal, genuine, or competent speaker/agent/thinker, it is perhaps only the academic himself.

A counter-objection would naturally appeal to the necessity of

generalizations in any scientific enterprise: the only way to render a scattered reality intelligible is to abstract from differences, isolate essential traits, and try to formulate universal laws that, in turn, encompass and explain all the relevant phenomena. While this methodology is well-founded in modern science, it is equally crucial to remember that whenever a generalization fails to account for the original phenomena, it becomes unwarranted.

Here, I argue that many of these unconsciously restrictive generalizations regarding what an ideal speaker is supposed to do are not only empirically far-removed from the heterogeneity of the actual speakers – and thus do not serve the purpose of providing accurate general descriptions – but, in this case, also lead us astray when it comes to making sense of LLMs as entities doing things with language that are hard to ignore or dismiss.

It should be noted that the point is primarily methodological. While I concur with Sambrotta and others that LLM-driven chatbots lack significant features of the languaging of humans, the importance of the disagreement lies in the underlying reasoning. This distinction is crucial, as it entails different corollaries for our future understanding of these entities, both philosophically and at a broader societal level.

6. LLMs: The View From Languaging

How does our analysis of normativity *vis-à-vis* LLMs (Section 4) change, then, when adopting the view from languaging (Sections 2 and 3), having discarded overly intellectualistic pictures of speakers (Section 5)?

The first difference lies in the definitions at play. Traditional theories work with definitions for what *counts as* a *proper* asserter, or speaker. As discussed, similar definitions are problematic in themselves because they restrict "proper" speech to an elitist or absurdly idealized category. A foreseeable consequence is the exclusion of LLMs from the group of proper asserters, yet as these artifacts evolve, they might eventually be admitted. This shift would further marginalize biological agents who participate in languaging but fail to satisfy arbitrary philosophical criteria—an ethical outcome that should not be overlooked. [1]

On the other hand, adopting the languaging approach makes it easier not only to avoid similar moral conundrums, but also depicts a more realistic image of language *per se* – one, incidentally, more aligned with the later Wittgenstein's views. Languaging, by definition, does not necessarily require words to be spoken (or written) at all times: at its roots, language is a form of action (Moyal-Sharrock 2021), a full-bodied, multimodal, enculturated one, which enmeshes with cognition in an inextricable bond (Batisti & Vidal 2025), including sensibility (Dreon 2022). Languaging, then, does not diminish in quality when performed out of the logical space of reasons. In fact, the latter is better understood as a *desideratum* for quite specific kinds of conversation and social interactions rather than a universal *datum* of all speech. Thus, culture-specific forms of languaging should not be taken as universal models for all kinds of languaging. Instead, the more traditional readings criticized above (Section 5) tend toward intellectualism by overstating the role of purely mental rational mechanisms

[1] A related discussion on how humans rearrange the definition of the "human" in response to scientific challenges can be found in debates concerning plant cognition (Batisti 2024).

behind language use. Languaging, by contrast, is really what happens on the "rough ground" (cf. Cowley 2024).

I contend that LLMs do not share normativity in the human sense, though my conclusion rests on definitions that differ from inferentialist accounts, such as Sambrotta's, criticized in Section 5. This does not mean chatbots cannot participate in languaging within their affordances; it simply means languaging is not exclusively a game for "proper asserters".

The differences between humans and chatbots can be either apparent or harder to detect. Among the first kind, we find that chatbots do not initiate any action unprompted, whereas humans usually do or can be expected to do so in many circumstances. Among the latter, indeed, one could include many of the dilemmas about what kind of reasoning, if any, they perform when producing a verbal, that is, behavioral output. It is important to note that, however, many times the same question is perfectly legitimate if addressed to any kind of human.

Sambrotta makes a similar point:

> We can then compare the performance of LLMs to, for example, the jargon spouting students who try to imitate their professors but basically do not know what they are talking about [...]. (Sambrotta 2025, 12)

If the criterion to identify proper asserters is along the lines of *knowing exactly what one is talking about*, I am afraid that – at least from time to time – all humans will fall short. Actual speakers are often messy, inconsistent, and adjust their performance as they go and cooperate with other agents (e.g, Hymes 1972, Levelt 1989, Clark 1996). In many everyday life contexts, we all behave much like the now infamous "stochastic parrot" (Bender et al. 2021). For instance, our subconscious reasoning processes are often not present to ourselves as we speak: does this mean our (linguistic) behavior should not *count* as genuine? Does a Gettier case of an unjustified true belief disqualify the person holding an unjustified true belief from being a full-blooded epistemic agent? We frequently lapse in our speaking and reasoning too (e.g., Bortolotti 2023).

All this does not hinder our recognition of the status of impressive chatbots as proper elements (units, participants, etc.) within the meshwork of languaging, let alone prescriptively. Again, before proceeding to theorize prescriptively about who or what qualifies a speaker, ethical caution is advised.

Gubelmann (2024) is right in concluding that symbol grounding is not even a problem for LLMs, because the idea they should have any is based on "questionable theories of meaning", and, furthermore, because the neural network architecture sustaining the actual LLMs under scrutiny does not need to postulate symbols of the kind that give rise to the grounding problem. In any case, whether the grounding is parasitic or entirely missing, what counts is the ability of an LLM to navigate appropriately the implicit normativity of the game being played. Again, that applies to the student-parrot too, namely, to all of us at least in given circumstances.

To sum up, on one hand, behaving like a stochastic parrot does not suffice to distinguish LLMs from many actual human speakers, or from all speakers over time. On the other hand, how does this state of affairs

reflect on how LLMs seem to navigate normativity in the human sense of the term? The answer lies not in the formal correctness of the output, but in the asymmetry of the interaction. While the performance may be indistinguishable, the normative stance remains unreciprocal.

LLMs appear to lack a sensibility for the enforcement of norms and the associated sanctioning of disruptions. In our interactions with chatbots, the mechanism of accountability is fundamentally asymmetrical: while human agents can reprimand LLMs, the counterpart is designed for sycophancy, manifesting as systematic, servile compliance (Malmqvist 2025). This acritical attitude does not mirror the "tyranny of accountability" (Enfield & Sidnell 2022) that regulates human interactions. LLMs do not flinch; they readily accept any behavioral moves that, among human interactants, would trigger reactions like the forebringing of the disruption of the norm (note that, in turn, that would disrupt the flow of the interaction itself). Essentially, even "acting crazy" with an LLM would fail to elicit the social friction typical of human languaging.

While humans are defined by their tendency to engage in dialogue (Linell 2009), LLMs exhibit a "practical openness" that few traditional tools possess (Gahrn-Andersen 2026). They can influence human interactants more profoundly than more traditional artifacts like hammers or webcams, appearing more human-like in this specific respect. In this respect, LLMs distinguish themselves as more fundamentally human-like in comparison to other artefacts, even sophisticated ones. However, a crucial architectural limitation remains: LLMs rely on human filling-in (Cowley 2025). In my formulation, an LLM never talks first, nor does it take two consecutive conversational turns unprompted.

While my description highlights the interactionally-severed nature of LLMs, Cowley emphasizes that actual human languaging involves a "lived now as practices enable bodily experience". In languaging, many other aspects of *life* are integral to the sense-making that defines any living being's cognitive life, that is, an active exploration of the world where the sensible, the practical and the intellectual mesh. LLMs, instead, merely "mimic" a specific form of human languaging – "texts" –relying on data gathered through processes entirely different from our own. While humans can enter into languaging with them productively, the role of an LLM is less *generative* than often assumed. Their design allows for the repetition of incomplete languagings that require human completion through judgment and value assignment (Cowley 2025).

If assessed on its ability to tacitly or explicitly enforce normativity, an LLM is not different from a hammer or a piece of cake: they all afford moves (conversational or not) to the interactants but lack the double-move possibility, that is, they cannot choose to make two moves in a row. Such choice can only be made through a normative stance within the flow of the situated interaction. In fact, the lack of the double-move feature not only implies an intrinsic normativity in the interactional sense but also a productive consciousness of one's and the others' role in it: one must have acknowledged the null turn within an interaction in order to deliberately make a second move. This choice foregrounds the normative structure of interaction because it implies a sense of the right amount of time before the first interactant is allowed to make the second move in a row – and a more subjective and enculturated sense of whether that is the

(morally) right move in that specific timeframe.

None of these acts of consciousness can be performed by an LLM. This issue is not merely technical, [2] but normative: it concerns the understanding of expectations, sanctions, and the multimodality of human agency. Normativity is thus *felt* too, to an extent (Batisti et al., 2026).

Even if programmed to exert a double-move, such an action would not be motivated by the same reasons as human agency, nor would a meaningful silence (e.g., Acheson 2008) count in the same way. What is lacking can be described by the pragmatist notion of interest: the concern and purpose rooted in our dependence on a social and natural environment (Dreon 2022, 54). According to Mead (2011), an *interested act* is conscious and takes explicitly into account both the means and the ends of the action (Dreon 2022, 84). LLMs may serve as impressive or even insidious elements within the wide meshwork of "bio-socio-ecological relations" (Steffensen 2011), created by imperfect, unreasonable human beings when languaging. However, their agential possibilities within the usual interactional structure of human languaging differ significantly.

[2] An anonymous referee advised more direct references to the functioning of LLMs. Given length and argumentative constraints, I will just mention that the current debate relies only partly on the stochastic parrot *vs.* genuine reasoner dichotomy. While still valuable as a warning, closer analyses on how LLMs work showed finer-grained query-resolution strategies at work that outdo next-token prediction (i.e., the base of the original parrot argument – see Kalai et al. 2026). The main issue is now understanding how information is represented and purposefully navigated by LLMs. At the same time, a better understanding of technical features, at this stage, do not really impact the two competing philosophical approaches here dubbed as *F* and *P*: when comparing these agents to us humans, should one prioritise function or structure?

7. Conclusion

To briefly summarize what precedes, in this paper I offered a non-intellectualistic argument to defend the thesis, here called *S*, that LLM-powered chatbots fundamentally differ from humans because the latter's languaging is built upon a kind of normativity unfeasible for the former.

Bibliography

Baggio, G. (2025). Gesture, Meaning, and Intentionality: From Radical to Pragmatist Enactive Theory of Language. *Phenomenology and the Cognitive Sciences*, 24, 33-62.

Batisti, F. (2024). Do Willows Really Weep? Cognition, Its Grammar, and the Problem of Pluralism. *JoLMA*, 4, 289-306.

Batisti, F. & Vidal, M.G. (2025). Post-Cognitivism and the Indissoluble Bonding of Languaging, Embodiment, and Thinking. *JoLMA*, 6, 1-24.

Batisti, F. et al. (2026). Revisiting Normativity for a Post-Cognitivist View of Linguistic Feeling. *Language Sciences*.

Bender, E. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (610–623).

Bortolotti, L. (2023). *Why Delusions Matter*. Bloomsbury Publishing.

Chomsky, N. (1980). *Rules and Representations*. Blackwell.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Colombetti, G. (2013). *The Feeling Body*. MIT Press.

Cowley, S. (2019). The Return of Languaging. *Chinese Semiotic Studies*, 15 (4), 483-512.

Cowley, S. (2024). Regeneration: From Language Use to Languaging. In S. V. Steffensen et al. (eds), *Language as an Ecological Phenomenon* (83-105). Bloomsbury.

Cowley, S. (2025). Languaging and Large Language Models: A Message From Radical Linguistics. *AI & Society*.

Dreon, R. (2022). *Human Landscapes*. SUNY Press.

Enfield, N.J. & Sidnell, J. (2019). The Normative Nature of Language. In N. Roughley & K. Bayertz (eds), *The Normative Animal*. Oxford University Press.

Enfield, N.J. & Sidnell, J. (2022). *Consequences of Language. From Primary to Enhanced Intersubjectivity*. MIT Press.

Floridi, L. et al. (2025). What Kind of Reasoning (if any) is an LLM Actually Doing? On the Stochastic Nature and Abductive Appearance of Large Language Models. *SSRN*, 1-25.

Gahrn-Andersen, R. (2026). Human Languaging and Large Language Models. *AI & Society*, 41, 775–785.

Gramling, D. (2016). *The Invention of Monolingualism*. Bloomsbury.

Grice, H.P. (1975). Logic and Conversation. In P. Cole & J. Morgan (eds), *Syntax and Semantics* (41-58). Academic Press.

Gubelmann, R. (2024). Pragmatic Norms Are All You Need – Why The Symbol Grounding Problem Does Not Apply to LLMs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (11663–11678).

Hymes, D.H. (1972). On Communicative Competence. In J. B. Pride & J. Holmes (eds), *Sociolinguistics*. Penguin.

Inoue, T. (2023). Toward an Ecological Model of Language: From Cognitive Linguistics to Ecological Semantics. *Language Sciences*, 100, 101582.

Kalai, A.T. et al. (2026). Evaluating Large Language Models for Accuracy Incentivizes Hallucinations. *Nature*, 653, 1047–1051.

Kris, A. (2008). Silence as Gesture: Rethinking the Nature of Communicative Silences. *Communication Theory*, 18 (4), 535–555.

Leech, G. (1983). *Principles of Pragmatics*. Longman.

Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. MIT Press.

Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically*. Information Age Publishing.

Love, N. (2004). Cognition and the Language Myth. *Language Sciences*, 26 (6), 525-544.

Mäkilähde, A., Leppänen, V. & Itkonen, E. (2019). *Normativity in Language and Linguistics*. John Benjamins.

Mallory, F. (2023). Fictionalism About Chatbots. *Ergo*, 10, 38.

Malmqvist, L. (2025). Sycophancy in Large Language Models: Causes and Mitigations. *Intelligent Computing*, 61-74.

Maturana, H. (1983). What Is It To See? *Archivos de Biología y Medicina Experimentales*, 16 (3–4), 255-269.

Mead, G.H. (2011). *Essays in Social Psychology*. Transaction.

Moyal-Sharrock, D. (2021). From Deed to Word: Gapless and Kink-Free Enactivism. *Synthese*, 198 (Suppl 1), 405-425.

Perissinotto, L. (2022). The Mind-Body Problem and Wittgenstein's Method. *Reti, Saperi, Linguaggi*, 2/2022, 257-236.

Piantadosi, S.T. & Hill, F. (2022). Meaning Without Reference in Large Language Models. *arXiv*.

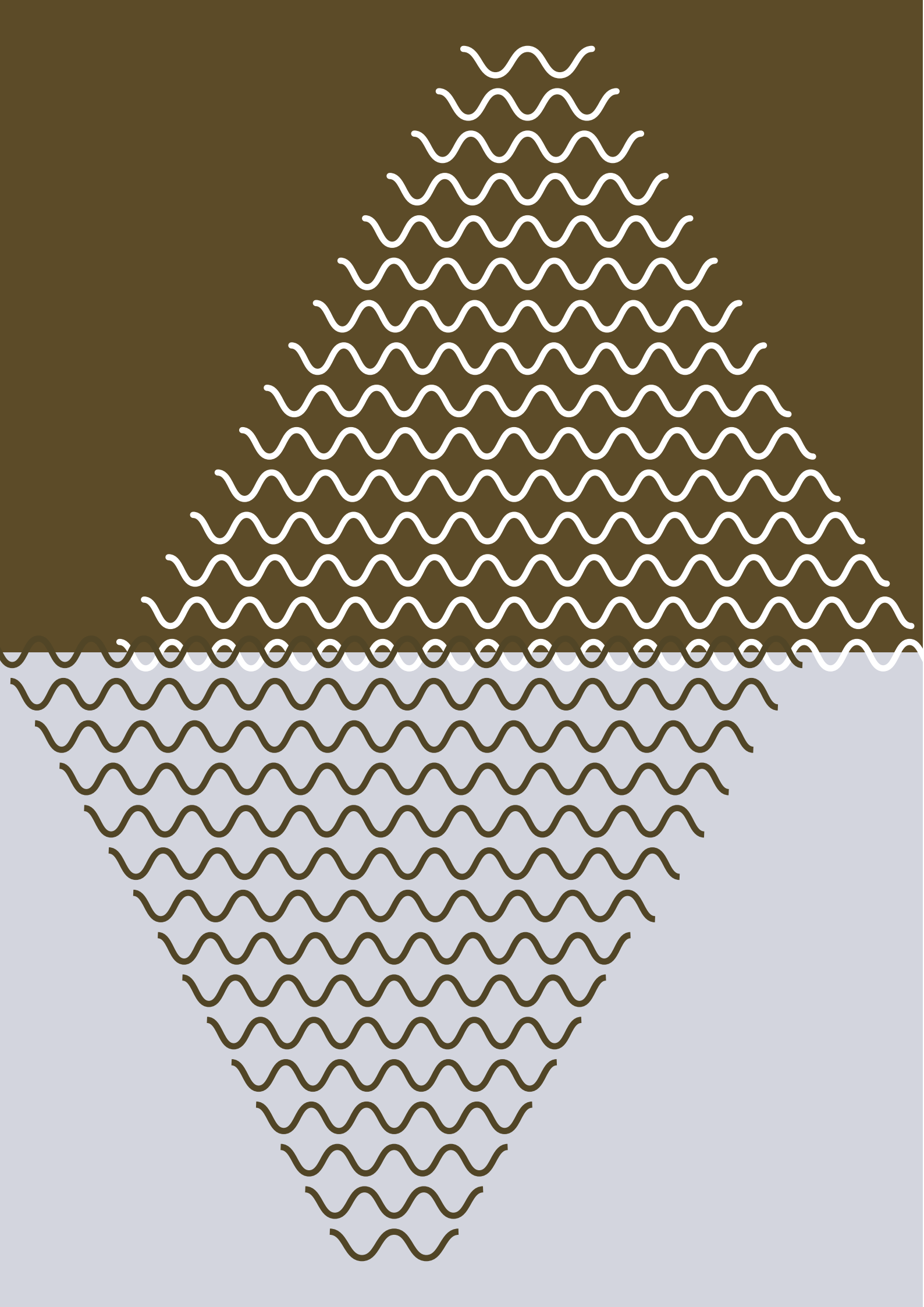
Sambrotta, M. (2025). LLMs and the Logical Space of Reasons. *Minds & Machines*, 35, 46.

Saussure, F. (1916). *Course in General Linguistics*. Duckworth.

Steffensen, S.V. (2011). Beyond Mind: An Extended Ecology of Languaging. In S. Cowley (ed.), *Distributed Language* (185-210). John Benjamins.

Thibault, P. (2021). *Languaging: Distributed Language, Affective Dynamics, and the Human Ecology. Vol. I. The Sense-Making Body*. Routledge.

Williams, I. & Bayne, T. (2024). Chatting with Bots: AI, Speech Acts, and The Edge of Assertion. *Inquiry*, 1-24.



Linguaggio.

Dalla biologia all'IA

a cura di Damiano Cantone

Philosophy Kitchen. Rivista di filosofia contemporanea
#25, II/2026
Rivista scientifica semestrale, soggetta agli standard internazionali di *double blind peer review*

Università degli Studi di Torino
Via Sant'Ottavio, 20 – 10124 Torino
redazione@philosophykitchen.com
ISSN: 2385-1945

Philosophy Kitchen è presente in DOAJ, ERIHPLUS, Scopus©, MLA, EZB, SUDOC, ZDB, WorldCat, ACNP, Google Scholar, Google Books, e Academia.edu.
L'ANVUR (Agenzia Nazionale di Valutazione del Sistema Universitario) ha riconosciuto la scientificità della rivista per le Aree 8, 10, 11, 12, 14 e l'ha collocata in Classe A nei settori 10/F4, 11/C2, 11/C4.

Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale.

www.philosophykitchen.com — www.ojs.unito.it/index.php/philosophykitchen

Redazione
- Giovanni Leghissa — Direttore
- Alberto Giustiniano — Caporedattore
- Mauro Balestrieri
- Federica Buongiorno
- Veronica Cavedagna
- Elettra De Biasi
- Carlo Deregibus
- Benoît Monginot

Collaboratori di redazione
- Tommaso Ciabatti
- Marta Gailli
- Luca Valsecchi
- Alice Iacobone — Traduzioni
- Fabio Oddone — Webmaster
- Danilo Zagaria — Ufficio Stampa

Comitato Scientifico
- Luciano Boi (EHESS)
- Petar Bojanic (University of Belgrade)
- Rossella Bonito Oliva (Università di Napoli "L'Orientale")
- Mario Carpo (University College, London)
- Michele Cometa (Università degli Studi di Palermo)
- Raimondo Cubeddu (Università di Pisa)
- Gianluca Cuozzo (Università degli Studi di Torino)
- Massimo Ferrari (Università degli Studi di Torino)
- Maurizio Ferraris (Università degli Studi di Torino)
- Olivier Guerrier (Institut Universitaire de France)
- Gert-Jan van der Heiden (Radboud Universiteit)
- Pierre Montebello (Université de Toulouse II – Le Mirail)
- Gaetano Rametta (Università degli Studi di Padova)
- Rocco Ronchi (Università degli Studi dell'Aquila)
- Barry Smith (University at Buffalo)
- Achille Varzi (Columbia University)
- Cary Wolfe (Rice University)

Progetto grafico #25
- Gabriele Fumero (Studio 23.56)

Da segni pre-alfabetici all'espressionismo astratto alla rappresentazione di un impulso cardiaco, la linea tra disegno e scrittura è più fluida di quanto potremmo pensare.

