



Università
Ca' Foscari
Venezia

Corso di Laurea
Magistrale
In Scienze Filosofiche

Tesi di Laurea

Strutture in trasformazione

Un'analisi semiotica
dei modelli generativi

Relatore

Prof. Matteo Pasquinelli

Correlatore

Dott. Tommaso Guariento

Laureando

Ettore

Visentin

Matricola

908815

Anno Accademico

2025/2026

Gli stati cambiano di fase, i sistemi cambiano di stato, tramite transizioni di fasi o di stati. Ma considerando il sistema in sé stesso, esso non è mai stabile. Il suo equilibrio è ideale [...]. Non vi sono sistemi istanze, sostanza se non nelle nostre ignoranze. Il sistema è il non-sapere. L'altro lato del non-sapere. Il non-sapere ha un lato caos e un lato sistema. Il sapere è un ponte tra queste due rive. Il sapere come tale è uno spazio di trasformazione.

Michel, Serres, *Il parassita*, Mimesis: Milano, 2022, p. 101.

Fu un grave errore, da parte delle correnti strutturaliste, pretendere di riportare tutto ciò che concerne la psiche al pastorale del significante linguistico! Le trasformazioni tecnologiche ci costringono a prendere in considerazione concorrentemente una tendenza all'omogeneizzazione universalizzante e irrazionalista della soggettività e una tendenza *eterogenetica*, un rafforzamento dell'eterogeneità e della singolarizzazione delle componenti della soggettività stessa.

Guattari, Félix, *Caosmosi*, Orthotes: Napoli, 2024, p. 31.

INDICE

INDICE	5
INDICE DELLE IMMAGINI	6
INTRODUZIONE	7
PARTE 1: INFORMAZIONE E SEGNO	15
<i>CAPITOLO 1: LA TEORIA DELL'INFORMAZIONE</i>	15
1.0 INFORMAZIONE E NEOSTRUTTURALISMO	15
1.1 LA PROBABILITÀ "SHANNON"	18
CAP. 2: LA STRUTTURA E LE SUE CONDIZIONI FORMALI	28
2.0 L'ORDINE SIMBOLICO	28
2.1 UN ANCORAGGIO PRATICO	33
2.2 SURDETERMINAZIONE E COMPLESSITÀ	41
CAP. 3: L'IPOTESI DIFFERENZIALE	44
3.0 DAL SEGNO AL SIMULACRO	44
3.1 MACCHINE INTELLIGENTI?	46
PARTE 2: LA STRUTTURA DEI MODELLI LINGUISTICI	53
CAP. 1: "WORD EMBEDDING"	53
1.0 DISTRIBUZIONALISMO HUMEANO	53
1.1 VETTORIALIZZAZIONE DEL PENSIERO	58
1.2 QVC IS ALL YOU NEED	64
CAP. 2: LA FUNZIONE POETICA DEL LINGUAGGIO	76
2.0 JAKOBSON E IL TELEGRAFO POETICO	76
2.1 NUMERI E PAROLE: UNA RELAZIONE APERTA	83
PARTE 3 GENERARE CULTURA	89
CAP. 1: QUALE GENERATIVITÀ?	89
1.0 SINTASSI E PRAGMATICA	90
1.1 VIRTUALE E POSSIBILE	96
CAP. 2: POETICA GENERALE	106
2.0 UNA ZUPPA TIEPIDA	106
2.1 INTRATTENIMENTO IDEOLOGICO	109
2.2 SIMULAZIONE CULTURALE	120
PARTE 4 MACCHINE SEMIOTICHE	125
CAP. 1: DALLA STRUTTURA ALLA MACCHINA	125
1.0 DALLA LINGUISTICA FORMALE AL DIAGRAMMA	133
CAP. 2: CORPO SENZA ORGANI COMPUTAZIONALE	142
2.0 CONCATENAMENTO UTENTE-LLM-CORPORATE	142
2.1 DESIDERIO VETTORIALE	148
CAP. 3: LE SEMIOTICHE DEGLI LLMS TRA HJELMSLEV, PEIRCE E GUATTARI	156
3.0 UN FILONE LINGUISTICO MATERIALISTICO	156
3.1 SOVRACODIFICA, FORMA DEL CONTENUTO E DELL'ESPRESSIONE	164
CONCLUSIONE	171
BIBLIOGRAFIA	179
SITOGRAFIA	183

INDICE DELLE IMMAGINI

FIGURA 1 SCHEMA DELL'INFORMAZIONE PER SHANNON	19
FIGURA 2 CONCATENAMENTO SEMIOTICO	21
FIGURA 3 CATENA DI MARKOV	24
FIGURA 4 PROCESSO SEMIOTICO MACCHINICO	55
FIGURA 5 RAPPRESENTAZIONE DI TRE VETTORI IN UNO SPAZIO TRIDIMENSIONALE	59
FIGURA 6 PROCESSO DI ATTENTION	67
FIGURA 7 PROCESSO QKV	69
FIGURA 8 MULTI-HEAD ATTENTION MECHANISM	70
FIGURA 9 STRUTTURA DEL LINGUAGGIO PER HJELMSLEV	96
FIGURA 10 ESEMPIO DI "MAKE IT MORE"	118
FIGURA 11 ESTENSIONE DEL RITRATTO DELLA GIOCONDA SECONDO ADOBE	122
FIGURA 12 RICIRCOLO CONTINUO DELL'INFORMAZIONE TRAMITE LLM (GENERAL PURPOSE TECHNOLOGY)	139
FIGURA 13 SEMIOTICHE MISTE	159
FIGURA 14 CICLO DELL'INFORMAZIONE TRA SEMIOTICHE MISTE E AI	165

INTRODUZIONE

Parlare di AI sembra essere diventata un'impresa infondata e senza via di uscita; un "iperoggetto"¹ le cui coordinate sembrano sfuggire a vista d'occhio. Un'opulenza di informazioni contamina quotidianamente la critica, in cui molti articoli dai toni oracolari si rivelano essere piuttosto delle ipotesi infondate². La novità scombussola sempre il dibattito e lo accende in svariate dimensioni; il problema è piuttosto l'alone sensazionalistico, propagandistico che circola in certi articoli. La situazione che viene a profilarsi sembra essere quella, per prendere in prestito l'immagine di Franco Vaccari, di una *discarica di rifiuti*³. Il dibattito, se preso localmente, ovvero facendo ad esempio riferimento agli articoli di una certa testata giornalistica oppure di un ente accademico, è massimamente differenziato. La sua falsa molteplicità locale si rilancia però globalmente come una vasta massa indistinta di notizie. Indistinta e in ultima istanza indifferenziata perché l'enorme quantità non arriva quasi mai a scalfire il cuore degli argomenti più urgenti e sostanziali quali ad esempio quello dell'istanza della *black box*.

Caratteristica, tra le altre, dell'iperoggetto è quella di non trovarsi situato entro uno specifico orizzonte di spazio, di tempo, ma anche concettuale. Nelle prossime pagine⁴ ci si riferirà a tale carattere attraverso la nozione di un pensiero della complessità, un'ipotesi di ricerca euristica portata avanti da epistemologi francesi come Gaston Bachelard, Michel Serres, arrivando fino alla contemporaneità del XXI con Edgar Morin⁵. L'attenta analisi richiede una commistione di saperi, un intreccio che sciolga le varie branche della scienza dalle loro rigide griglie e dai loro ambiti di ricerca. *Complessità* più che indicare un grado di difficoltà, riflette piuttosto una varietà eterogenea di concetti. Un filosofo come Gilles Deleuze dimostra come il suo metodo di indagine debba prendere a piene mani il pensiero di un autore, ma portarlo al limite, scardinandone gli esiti della ricerca e attualizzandolo⁶; egli rimane filosofo poiché si interessa alla creazione di concetti, ma nel farlo svolge un'indagine pluridisciplinare, attingendo dai saperi della

¹ Morton, Timothy, *Iperoggetti*, Nero: Roma, 2018.

² Si prendano ad esempio: Ullman, Tomer, *Large Language Models fail on Trivial Alterations to Theory of Mind Tasks*, <https://arxiv.org/abs/2302.08399>; Schaeffer, Rylan et al., *Are emergent Abilities of Large Language Models a Mirage*, <https://arxiv.org/abs/2304.15004>; Leech, Gavin, *Questionable practices in machine learning*, <https://arxiv.org/abs/2407.12220>.

³ Vaccari, Franco, *Fotografia e inconscio tecnologico*, Einaudi: Torino, 2011, p. 87.

⁴ Cfr. parte 1, cap. 2.2.

⁵ Ceruti, Mauro, Bellusci, Francesco, *Abitare la complessità*, Mimesis: Milano, 2020.

⁶ «Ma il mio modo di cavarmela all'epoca, consisteva soprattutto, almeno credo, nel fatto di concepire la storia della filosofia come una specie di inculata o, che poi è lo stesso, di immacolata concezione. Mi immaginavo di arrivare alle spalle di un autore e di fargli fare un figlio, che fosse suo e tutta via fosse mostruoso». Deleuze, Gilles, *Pourparler*, Quodlibet: Macerata, 1999, p. 16.

biologia⁷, della matematica⁸, della psicanalisi⁹, della letteratura¹⁰, dell'arte¹¹. Insomma, delle interazioni eclettiche nutrono e animano le pagine di *Differenza e ripetizione*, ad esempio, opera che verrà largamente tenuta in considerazione, assieme al repertorio di strumenti critico-pratici che testi come *Anti-Edipo* o *Mille-piani* consegnano al lettore. Per affacciarsi a una critica consapevole di un *iperoggetto* è quindi necessaria la presenza combinata di più saperi. Da qui l'esigenza nelle prossime pagine di richiamarsi ad autori di diverse branche.

L'indagine preliminare organizzata nella prima parte è finalizzata a mettere in luce i presupposti epistemologici della ricerca: lo strutturalismo si dimostra essere una fucina di saperi, un metodo analitico più che un preciso indirizzo tematico; ancora di più la teoria dell'informazione, una sottotrama che si vedrà predisporre i fondamenti del dibattito sugli LLMs anche storicamente. Il presente scritto segue da vicino le considerazioni di natura neostrutturalista del filosofo e linguista Leif Weatherby, in particolare nel suo libro *Language Machines*¹², che traggono forza dai connotati strutturalisti e differenziali della linguistica di Saussure e della “funzione poetica” di Jakobson. Weatherby, dopo aver esordito nel campo degli studi letterari, concentrando le sue ricerche attorno a certa terminologia del Romanticismo e dell'Idealismo classico tedesco, contribuisce attivamente alla disamina della proliferazione culturale degli LLMs¹³, ritenendogli macchine letterarie. Rimarchevole è l'operazione critica di Weatherby condotta contro il post-strutturalismo, in particolare contro l'asse Lacan-Derrida¹⁴, e la sua pretesa di asserire l'autonomia del significante appiattendolo in una trama di relazioni differenziali.

⁷ Ruyer, Raymond, *La genesi delle forme viventi*, Bompiani: Milano, 1966.

⁸ In *Differenza e ripetizione* autori come Bordas-Demoulin, Maimon, Leibniz, Abel risultano essere imprescindibili per comprendere cosa si intenda come condizioni di *realtà*, cascata di rotture di simmetria, quindi, più in generale misurare il fondamentale apporto che il calcolo differenziale dà al ragionamento sul concetto di “differenza in sé”; in *Mille-piani* verrà dedicato un paragrafo apposito a Riemann e alla sua geometria.

⁹ Uno tra tutti Lacan, come viene attestato in Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo?*, SE: Milano, 2008.

¹⁰ Proust è onnipresente in *Differenza e ripetizione*. A lui viene dedicata anche un libro cfr. Deleuze, Gilles, *Marcel Proust e i segni*, Einaudi: Torino, 1970. Altri autori fondamentali si trovano poi in Deleuze, Gilles, *Critica e clinica*, Cortina: Milano, 1996.

¹¹ Nel presente scritto ci si riferirà principalmente a Deleuze, Gilles, *Sulla pittura*, Einaudi: Torino, 2023; bisogna però ricordare anche la celebre monografia su Bacon: Deleuze, Gilles, *Francis Bacon. La Logica della sensazione*, Quodlibet: Macerata, 2023.

¹² Weatherby, Leif, *Language Machines*, University of Minnesota Press: Minneapolis, 2025.

¹³ Un *Large Language Model* (LLM) è un sistema di intelligenza artificiale addestrato su enormi quantità di testo per apprendere le relazioni statistiche tra le parole. Funziona generando testo un token alla volta, calcolando ad ogni passaggio la probabilità dei possibili token successivi in base al contesto. Il risultato, prodotto da una serie di *trasformative blocks* (matrici interne di calcolo) è un testo persuasivo, *come fosse* scritto da una persona umana, ma senza che sia veramente “compreso” dalla macchina. Si potrebbe allora connotare la sua posizione come quella di un neostrutturalismo più vicino alla storia materiale.

¹⁴ Weatherby, Leif, op. cit., pp. 64-77.

Derrida, infatti, ritiene che Saussure con l'istituzione della dicotomia linguistica significante-significato avrebbe prediletto la voce [*phoné*] al segno [*gramme*], promuovendo un paradigma metafisico della presenza. Ma per Weatherby non si può disgiungere il problema metafisico da quello semiologico, come sembrerebbe fare Derrida: Saussure non ha mai richiesto il primato della voce, e il fatto che la *parole* sposti di continuo i valori del sistema non implica affatto «una metafisica fuori dalla semiologia». È solo «una constatazione di principio su *cosa sia* un significante»¹⁵. Il risultato sarebbe allora un completo abbandono della riflessione sul linguaggio da parte delle *humanities*: un'operazione parallela e opposta che da un lato inaugura lo studio statistico sul linguaggio, dall'altro apre alle riflessioni derridiane che abbandonano le concrete modalità semiologiche con cui il significato viene a proliferare (in questo caso l'architettura *transformer* degli LLMs). Allora, il significante, ribatte Weatherby, «può peregrinare quanto vuole, ma non può entrare in alcun "gioco" senza assumere forma materiale»¹⁶. La traccia linguistica, su cui si innesta la differenziazione dei significati è già parte del sistema di codifica linguistico ed è precisamente ciò su cui si fonda quel gioco aperto e interminabile tra linguaggio e computazione¹⁷. Insomma, il tentativo di critica di Derrida non fa che spostare il discorso su un piano metafisico anti-presenzialistico, eludendo le proprietà che lo strutturalismo aveva additato al segno e che oggi si rivelano più importanti che mai. Allora perché mettere in dialogo Weatherby, Deleuze e Guattari? La sua critica al post-strutturalismo non si può estendere anche a loro, escludendoli dal dibattito sugli LLMs? Deleuze e Guattari rivendicano una pluralità di linguaggio che relativizzano la pretesa egemonica del significante e che si declinano in una molteplicità di specifiche semiotiche. Come si è già detto, l'importanza dei linguaggi matematici è messa da loro in primo piano, aspetto che Weatherby riconoscere essere cogente e di primaria importanza per un dibattito adeguatamente fondato sugli LLMs¹⁸; inoltre, nell'ultima parte viene applicato il concetto di *diagramma*, il cui dispiegamento non è mai indipendente dall'assemblaggio macchinico in cui viene a compiersi; in altre parole, è la composizione concreta delle regole tecniche, scientifiche che governano precipuamente una struttura simbolica che trasmette l'informazione linguistica (ciò nonostante il linguaggio non può dirsi fisico, ma al più *materiale*). Dunque, Deleuze e Guattari si sono rivelati essere un punto di riferimento che, per riecheggiare il titolo di

¹⁵ *Ibid.* p. 71.

¹⁶ *Ibid.* p. 73.

¹⁷ Cfr. parte 2, cap. 2.1.

¹⁸ Weatherby rimprovera al post strutturalismo di non aver «mai preso sul serio, specie la matematica» cfr. *Ibid.* p. 66.

uno dei paragrafi finali, predispone a una “linguistica materialistica” (infatti *Mille piani* è intessuto di riferimenti a Louis Hjelmslev).

Una delle personalità chiave del discorso seguente è Roman Jakobson, linguista e critico letterario che incrocia entrambi gli addendi del discorso. Egli si è prodigato in esperimenti finanziati dal Governo americano nell'utilizzo di strumenti elettronici e d'informazione come il telegrafo. L'uso del media altera, afferra e sussume il modello epistemologico. La teoria in questo senso è seconda alla pratica, ovvero all'utilizzo dello strumento, a sua volta congeniale a un'idea dell'inventore¹⁹. In ottica strutturalista si può a buon ragione operare un ampliamento della ricerca sul fenomeno AI grazie a un decentramento dalle istanze soggettivistiche:

Quello che invece noi proponiamo è l'abbandono del continuo e implicito riferimento all'uomo per sostenere un radicale spostamento del punto di osservazione verso lo strumento, che deve essere visto come dotato di un'autonoma capacità di organizzazione dell'immagine in forme che sono già strutturate simbolicamente, indipendentemente dall'intervento del soggetto.²⁰

A questi fini non si poteva che dedicare larga parte della dissertazione sul precipuo funzionamento dell'apparato di codificazione e de-codificazione di queste macchine, come avviene nella seconda parte. Da una parte, il fenomeno viene illuminato da una retrospettiva storica che ne metta in luce le specificità, richiamandosi ai vari modelli computazionali che hanno segnato la storia dell'informatica: dall'AI simbolica ai modelli a rete, dalla cattura digitale-analogica alla sua riscrittura in termini esclusivamente digitali, o meglio *vettoriali*, passando perciò tramite alcune delle pagine più influenti dell'*engineering retrieval*. È una carrellata di modelli tecnici, la cui impronta internalista serve per preparare il lettore alla successiva descrizione del modello LLM, che apparirà così una sincretica operazione ingegneristica di incastro di diversi aspetti epistemologici e formali delle precedenti tappe dello sviluppo informatico: *information retrieval*, programmazione, neuroscienze. Così torneranno pure i due grandi presupposti teorici

¹⁹ Damerow, Peter, “Abstraction and Representation”, in *Abstraction and Representation: Essays on the Cultural Evolution of Thinking*, Springer: Berlin, 2013. Si può indicare questa dinamica vorticosamente tra teoria e prassi come una dialettica materiale. Autori come Gilbert Simondon mettono in luce la plasticità dell'esperienza, la sua formazione a partire dal *medium* che si intende usare; cfr. Simondon, Gilbert, *Del modo di esistenza degli oggetti tecnici*, Orthotes: Roma, 2021. In area italiana Franco Vaccari si dimostra un attento catalizzatore della tendenza a vedere il pensiero prodursi sulla base e a partire dallo strumento che si vuole usare (nel suo caso la macchina fotografica) e delle strategie di formazione inconsce che questi muovono. Vaccari, Franco, *Fotografia e inconscio tecnologico*, Einaudi: Torino, 2011.

²⁰ Vaccari, Franco, op. cit., pp. 18-19.

(strutturalismo e teoria dell'informazione della prima parte).

Le prime due parti inseguono quindi una direzione strutturalista, rafforzata dai riferimenti neo-strutturalisti che vi si trovano. È indubbio che lo strutturalismo, con la sua nozione di *segno*, riesca a recidere ogni situabilità di quest'ultimo, sradicandolo dal suo terreno referenziale. «Il pensiero simbolico libera il concetto da qualsiasi intuizione dell'oggetto»²¹. Il simbolo, detto anche segno, non è altro che un operatore di struttura, come si cercherà di dimostrare, ossia un mezzo per cui possono darsi delle opposizioni distintive. La condizione del simbolo sarebbe quella di *non* essere ciò che rappresenta. Gioco-forza dell'insistenza del simbolico sul reale è perciò la possibilità di scinderlo dallo sviluppo di quest'ultimo, ovvero dalla crescita e dalle proprietà degli oggetti reali che prende in considerazione. L'operazione può dare vita perciò a feticismi di sorta, ritenendo che sia possibile intendere anche il circolo simbolico in maniera sistematica e conclusa. Allora tutte le più alte virtù e colpe potranno venire studiate facendo unicamente ricorso al suo alone strutturale; il pericolo per un pensiero che si voglia dire critico è di spostare l'attenzione sull'identità dei concetti e le loro opposizioni, entro questo grande *gioco* simbolico, disconoscendone l'origine più materiale che ne sta a monte.

La terza e la quarta parte lavorano allora in opposizione con le prime due, scongiurando una certa feticizzazione del simbolico, ma permettono anche di proseguire alcune considerazioni già avviate in corso d'opera; ad esempio, considerando come l'ordine simbolico rispecchi l'ambiente culturale di una certa struttura, sarà proprio nella terza parte che si trarranno alcune conseguenze sulla capacità di produrre cultura degli LLMs. O ancora, si ri-intrecceranno i fili del discorso sulla Teoria per Althusser²², muovendo ad un altro suo concetto fondamentale, quello di ideologia. Viene messa luce sulla questione della generatività del modello: in che senso questi modelli possono dirsi creativi o generatori di nuove forme simboliche? La nozione di *generatività stocastica*, mutuata dalle riflessioni sugli LLMs dell'antropologo Paul Kockelman²³, attualmente professore a Yale, che costituisce il fulcro della terza sezione, cerca di rispondere a questa domanda. Il termine riflette il modo specifico con cui il modello linguistico, attraverso la sua architettura *transformer*, ricava una certa quantità di testo a partire da altro testo: *make more given some*. La macchina, infatti, campiona una distribuzione di probabilità e la ri-modella automaticamente. Il movimento quindi è quello che va dal testo passato,

²¹ Ortigues, Edmond, *Le discours et le symbole*, Aubier: Parigi, 1962.

²² I due riferimenti principali per Althusser saranno la serie di saggi di Althusser, Louis, *Per Marx*, Mimesis: Milano, 2008; i seminari che tenne assieme a Etienne Balibar Althusser, Louis, Balibar, Etienne, *Leggere il Capitale*, Mimesis: Milano, 2006.

²³ Kockelman, Paul, *Last Words*, Prickly Paradigmatic Press: Chicago, 2024, p. 64.

immeso come *prompt*, a quello futuro, prodotto come *response*. Più che di una creazione si tratta allora di una riconfigurazione di quello che si è già detto.

La generatività stocastica si può allora contrapporre alla creazione vera e propria per come la presenta Deleuze in *Differenza e ripetizione*; una precisazione urgente per evitare facili analogie tra virtuale deleuziano e lo spazio latente degli LLMs. Il virtuale è un campo di forze potenziali inesauribile, che funziona tramite differenziazione (attualizzazione), secondo un preciso movimento e una certa temporalità. Virtuale non significa “possibile”²⁴, poiché sfugge dalla somiglianza al *prompt*; la sua è una differenziazione priva di leggi fisse, identificabili: è un movimento *aberrante*²⁵. Lo spazio latente del modello generativo è piuttosto considerabile come una matrice del possibile, che procede per accumulo di associazioni contigue.

A questo punto si situa il secondo snodo della terza parte, ovvero la rilettura del concetto di ideologia per Weatherby. Dopo Deleuze, ora il dialogo si intreccia con Louis Althusser. La matrice di un LLM, predisponendo delle possibili permutazioni, ovvero dei raccordi con il testo passato senza soluzione di continuità. La mancata rottura, precedentemente chiamata “evento”, prende ora un altro nome: *averaging effect*. Il livellamento di cui si parla è quello tra informazione e rumore, le due categorie classiche della teoria informatica della comunicazione che vengono qui fatte coincidere. Il rumore diventa informazione e viceversa: l’informazione non è più un contenuto rilevante perché inatteso o inusuale, ma in virtù della sua elevata frequenza di trasmissione. Althusser risulta un termine di paragone appropriato anzitutto perché la sua nozione di ideologia è la stessa problematica strutturale, ovvero il campo di domande e di possibilità che si stabilisce a partire dalla struttura formale, dimettendo ogni posizione soggettivistica che la riterrebbe una questione di illusioni e false credenze. Per questo ci si servirà anche di Foucault, mostrando la piega epistemologica che ha preso il concetto nei termini di “a priori storico”.

In secondo luogo, dall’ideologia ci si libera solo tramite il processo teorico indicato nella prima parte, ovvero tramite la rottura di Generalità. Ebbene, proprio perché tale rottura non si avverte nel caso degli LLMs, è più opportuno ritenere l’estensione dello stesso spazio latente come ideologia; la *heat map*²⁶ può dirsi a buon diritto ideologica

²⁴ Per approfondire il discorso si rimanda a Lévy, Pierre, *Il virtuale*, Cortina: Milano, 1997, p. 51; Massumi, Brian, *Parables for the virtual: movement, affect, sensation*, Duke University Press: London, 2002.

²⁵ Lapoujade, David, *Aberrant Movements, the philosophy of Gilles Deleuze*, Semiotext(e): London, 2017.

²⁶ Con *heat map* si riprende l’espressione di Weatherby, intendendo lo spazio latente degli LLMs, ovvero il campo vettoriale multidimensionale, in cui vengono rimescolati vettori e matrici. Si rimanda a parte 2, cap. 1.2.

proprio perché in essa non si dà alcuna rottura: ogni distribuzione resta pur sempre una Generalità I, materia prima del linguaggio nella sua diffusione, mai conoscenza scientifica. Assieme alle figure del “semplice” e della “totalità”, il concetto di ideologia per la prima volta diventa empiricamente scansionabile, a dispetto di quanto Althusser potesse pensarne in termini di un invisibile formalismo strutturale.

La quarta parte, invece, è da leggere a partire dal gesto che Bernhard Rieder sottoscrive in *Engines of Order*: costituire una filosofia dello strumento tecnologico, capace di considerare il modello nella sua autonoma consistenza proprio come nel caso della meccanologia di Gilbert Simondon. Mentre per Rieder il discorso viene concepito per rendere conto delle tecniche di *information retrieval*, qui appare urgente leggere i modelli linguistici attraverso i concetti “pratici” di macchine desideranti e degli assemblaggi, portando a compimento quello spostamento del punto di osservazione verso lo strumento da cui l'intera ricerca aveva preso le mosse. Tre sono i principali concetti che servono ad ampliare lo spazio del dibattito, uscendo alle maglie strutturaliste: assemblaggio, corpo senza organi e semiotiche miste. Il passaggio è giustificato teoreticamente da un salto dal livello strutturale a quello “macchinico”, “molare” a quello “molecolare”, ovvero dalle condizioni strutturali per costruire e mettere in moto il modello generativo, al suo più svariato (si cercheranno di fare alcuni esempi concreti, impersonati dalle figure del consumatore e dell'elettore tra gli altri) utilizzo, quindi *funzionamento*. Nelle prime parti, infatti, ci si è concentrati esclusivamente su come il modello elabora i dati, prediligendo una visione interna al suo meccanismo; ora, la posta in gioco si estende anche ad altri due elementi: l'utente (visto come un aggregato di forze, di desideri e di intenzioni) e l'azienda produttrice [*corporate*]. La loro azione reciproca sul modello generativo si condensa in un *concatenamento* di forze, dimostrando come sia riduttivo e insufficiente trattare i vettori-parola come una struttura simbolica autonoma. In *Mille piani*, Deleuze e Guattari smontano l'idea di una macchina astratta chiusa in sé e indipendente da ogni fattore estrinseco: ad esempio, ogni parola si presenta già come azione, forza, ovvero in una *trasformazione incorporea* che permette di categorizzare, ripartire e disciplinare corpi e oggetti ancor prima che un effettivo e concreto ordine politico lo renda noto. Insomma, una prescrizione algoritmica è già un'imposizione politica.

La distinzione che si appronta nella parte tre tra virtuale deleuziano e spazio latente degli LLMs serve proprio ad escludere Deleuze dal gioco post-strutturalista criticato da Weatherby: Deleuze non spaccia il virtuale per ciò che il modello realizza e

così lo mantiene come concetto che possa venire applicato criticamente anche alla computazione.

Oltre alla sfera politica, è sembrato essenziale richiamarsi anche alla direzione libidica del desiderio; il paradigma rimane quello inaugurato negli anni '70 da Deleuze e Guattari, ovvero quello della “schizoanalisi”. La schizoanalisi è una pratica che ha per oggetto l'ambiente in cui l'uomo si situa, ovvero l'assemblaggio di macchine a cui si richiama, atto a plasmare e a codificare il proprio desiderio. Anche il concetto di desiderio è sottoposto a una ridefinizione, che si affianca alla messa in opera del “corpo senza organi”: le matrici QKV²⁷ del modello catturano e registrano i flussi di desiderio attraverso la rimodulazione dei loro pesi. A titolo preliminare l'argomento si può inquadrare intuitivamente così: la stesura di un *prompt* condensa un desiderio, una “volontà” dell'utente e la risposta che la macchina dà sembrerebbe indicarne un appagamento. Ma, in virtù delle considerazioni affrontate nelle prime parti, l'appagamento è surrettizio e “stocasticamente” collaudato, ovvero viene predisposto senza condizione di causa che non sia un mero calcolo probabilistico. Inoltre, qui l'aggiunta precipua di questa ultima parte, in virtù del suo utilizzo massimo, un LLM può rilevarsi un utile strumento ideologico di adescamento e soddisfacimento facile dei desideri, aiutando le aziende ad accrescere il loro capitale non solo economico, ma soprattutto di informazioni personali [*data mining*] che i desideri veicolano, riuscendo, per giunta, ad arrivare a loro volta a provocare nuovi desideri nel consumatore. Infine, sono le semiotiche miste di Guattari a servire da strumento per mostrare come non solo i desideri vengono plasmati dalle macchine, ma anche il discorso: il livellamento ideologico di cui si parlava nella terza parte acquisisce qui un'ulteriore giustificazione: la sovracodifica del vettore, garante dell'informazione per il LLM, traduce ogni sapere altro che costella il campo sociale.

²⁷ Si tratta della moltiplicazione tra matrici QKV: “query”, “key” e “value”. Le tre matrici costituiscono il cuore dell'architettura *transformer* del modello poiché, grazie al loro continuo moltiplicarsi vicendevole, la *query* immessa, ovvero il *prompt* testuale dell'utente, viene analizzato secondo molteplici scansioni attraverso dei pesi (*keys*), fornendo un risultato che andrà a ri-calibrare le misure con cui inizialmente il testo era stato analizzato.

PARTE 1: INFORMAZIONE E SEGNO

CAPITOLO 1: LA TEORIA DELL'INFORMAZIONE

1.0 INFORMAZIONE E NEOSTRUTTURALISMO

Come recenti scritti avvertono²⁸, ma già altri da tempo prospettavano²⁹, le teorie linguistiche che sono sorte a partire dagli anni '50 in area francese e americana sono strettamente connesse alla teoria matematica dell'informazione. Questa branca ingegneristica venne fondata da Warren Weaver e Claude Shannon nel 1949³⁰ col fine di dimostrare che ogni comunicazione tra oggetti, entità, macchine, esseri umani, avviene attraverso lo scambio di informazioni. Lo scopo della ricerca sarebbe allora quello di garantire il passaggio, migliorando la trasmissione di informazione. Il termine informazione abbandona qui il suo significato generico per definire un'unità di misura di scambio. L'informazione, il sapere dell'accadimento di un certo evento entro una situazione, può venire calcolato in *bit* (*binary digit*, ossia "segnale binario"), indicando la serie di disgiunzioni binarie che servono a individuare un'alternativa. L'aspetto che ancora oggi attrae di questa teoria è l'individuazione di una dinamica essenziale, dunque della forma, intesa come struttura elementare, per mezzo di cui può avvenire un'informazione, e questo perché «è utile individuare il rapporto comunicativo, nella sua dinamica essenziale, là dove si profila con maggiore evidenza e semplicità, suggerendoci la costruzione di un modello esemplare»³¹.

La ricerca oggi si fa ancora più cogente al fine di aprire nuove zone di ricerca e fornire nuovi strumenti di studio per analizzare il rapporto con cui l'informatica (i LLMs) si è strettamente connessa ai fenomeni culturali; un intreccio che merita di essere studiato a partire dalla prospettiva di Leif Weatherby³² a cui si devono gran parte delle considerazioni che seguiranno. La sua inedita prospettiva porta a riflettere sui LLM

²⁸ Geoghegan, Bernard, D., *Code: From Information Theory to French Theory*, Duke Press University: Durham, 2023.

²⁹ Eco, Umberto, *La struttura assente*, Bompiani: Milano, 1994.

³⁰ Shannon, Claude, E., Weaver, Warren, *La teoria matematica delle comunicazioni*, ETAS Kompas: Milano, 1971.

³¹ Eco, Umberto, op. cit., p. 49.

³² Weatherby, Leif, *Language machines*, University of Minnesota Press: Minneapolis, 2025.

intendendoli principalmente come “macchine letterarie”, prospettiva in linea con quella del “neostrutturalismo” che ritiene di tornare ad alcuni pensatori strutturalisti come Jakobson per trarne alcuni strumenti importanti. Il suo lavoro si inserisce organicamente nella cornice neostrutturalista, come quello di Lydia Liu³³ e Juan Luis Gastaldi³⁴, che per primi hanno messo in luce l’ipotesi strutturalista per la descrizione dei modelli generativi, ritenendo che l’intimo legame tra linguaggio e computazione sia possibile sulla scorta della funzione *poetica* e ideologica del linguaggio, ancor prima di quella referenziale; l’inventore della fonologia strutturale, Roman Jakobson, infatti offre una riserva di concetti ancora oggi applicabili nel ventre di quello che si può considerare il “neostrutturalismo”: non un ritorno nostalgico alla struttura, ma una sua riapplicazione critica alla luce degli argomenti post-strutturalisti.

Solo per dirne alcuni, basti ricordare che il linguaggio non può che esplicitarsi su binarismi oppositivi; il suo schema sulle sei funzioni del linguaggio anticipa diverse strutture formali che troveranno spazio nell’informatica e nella linguistica computazionale, mentre i suoi concetti di “metonimia” e “metafora” hanno trovato un prolifico terreno d’applicazione nello studio lacaniano. Alla triade, resa celebre da Charles Morris, di semantica, sintassi e pragmatica, segue una più netta distinzione dicotomica tra linguaggio “scientifico”, referenziale, dunque fondato su una sintassi arbitraria e la cui semantica dovrebbe enunciare la perfetta adeguazione al referente, marginalizzando ogni prospettiva pragmatica; al contrario, tutte le discipline umanistiche si concentrarono su un linguaggio storico e psicologico, in cui andava ricercata l’intenzione dell’enunciante e il suo rapporto con il contesto sociale; tra queste, molte rimasero vicine a una linguistica di tipo dicotomico e differenziale derivante dall’asse Saussure-Jakobson, mentre altre, come Deleuze e Guattari si ispirarono a un’unità

³³ Liu, Lydia, *The Freudian Robot: digital media and the future of the unconscious*, University of Chicago Press: Chicago, 2010. Nel libro è dedicato un ampio studio a Jakobson. Come mostreremo nei capitoli della presente tesi, la teoria delle sei funzioni del linguaggio non è scindibile dal contesto tecnico e sociale in cui è germogliata: Jakobson nel 1952 partecipò a una conferenza al Rockefeller Center di New York con i teorici della cibernetica e della teoria dell’informazione, lasciando profonde tracce nel suo schema interpretativo. Jakobson cercherebbe allora di tradurre in termini di linguaggio naturale la trasmissione d’informazione che avviene tra macchine cibernetiche, nei circuiti d’informazione.

³⁴ Juan Luis Gastaldi è un filosofo del linguaggio, storico della logica e della matematica. Egli ritiene che Jakobson con il suo progetto di fonologia strutturale abbia già introdotto degli strumenti e delle tecniche di formalizzazione che possono venire riapplicate in ambito computazionale. L’idea che ogni elemento valga limitatamente al numero di relazioni che costruisce con gli altri era difatti un’idea già presente a livello di algebra e teoria degli insiemi. In Jakobson si presenta allora una geometria del linguaggio: l’articolazione secondo dicotomie è la messa in luce di operatori nel linguaggio. Gastaldi mira allora a portare alla luce l’intima struttura che anima il linguaggio, dialogando in maniera critica con le posizioni “analogiche” deleuziane. Cfr. Gastaldi, Juan, Luis, *Why Can Computers Understand Natural Language?: The Structuralist Image of Language Behind Word Embeddings*, in «Philosophy and Technology», 34(1), 149–214.

semiotica tripartita, come quella di Peirce. Con l'avvento dei LLMs si assiste a un intreccio inedito tra queste due prospettive.

Per lo strutturalismo si tratta di aver a che fare con un sistema di differenze, prima che delle svariate occorrenze concrete, per dirla con Saussure, con la *langue* prima che con la *parole*. Quindi la prospettiva scientifica, cara a Frege, che riteneva la funzione primaria del linguaggio il suo essere indicale e referente rispetto al mondo esterno, sembra aver perso importanza. I grandi modelli NLP³⁵, come bene riassume Paul Kockelman,³⁶ seguono un processo semiotico che nulla ha a che fare con il referente esterno, potendo preliminarmente essere concepiti come una dialettica tra parola e parola. In questo, la semiologia si è rilevata profetica come branca di studi che prescinde da una dimensione psicologista o metafisica e tratta il linguaggio come fosse un sistema di differenze negative in sé; ogni elemento, chiamato segno, unisce due aspetti: significante e significato. Il significato non è il referente esterno, ma il concetto mentale che si può attribuire al significante, l'“Immagine mentale”. L'arbitrarietà tra segni è condizione di scientificità della semiologia che non deve esaurire i suoi sforzi nell'analisi del linguaggio. Nonostante la pluralità di regimi simbolici, quello del linguaggio sembra costituire il medium per eccellenza sugli altri domini. Significa che ogni disposizione simbolica si può costruire tramite linguaggio e appoggiandosi su di esso, tanto che «*il puro ordine simbolico dei LLMs ci presenta un mezzo di addomesticamento degli altri media in una maniera mai vista prima*»³⁷.

Come può un sistema generare da sé dei significati? Weatherby sostiene che due siano i “regimi di segni” che entrano in rapporto: computazione informatica e linguaggio verbale. I due regimi, nella misura in cui si intrecciano e si corrispondono, entrano in rapporto strutturale. Viene a crearsi quella condizione che Gilles Deleuze ha qualificato come struttura³⁸.

La struttura è un insieme aperto di relazioni che nascono tra almeno due serie. Qui

³⁵ NLP (*Natural Language Processing*) è diventato negli ultimi decenni un paradigma linguistico di grande rilevanza. A Differenza dei modelli simbolici, che ritenevano possibile scannerizzare il linguaggio a partire da un codice fisso di regole, il modello NLP si dispiega a partire da un approccio probabilistico, che prenda in considerazione il testo stesso e lo sviluppo reale del discorso umano. Tuttavia, è necessaria un'integrazione con l'“ordine del discorso” e le regole semantiche, sintattiche e grammaticali che lo costituiscono; i modelli generativi, in virtù dell'architettura *multi-head* che li contraddistinguono, si dimostrano in ciò particolarmente efficaci (cfr. parte 2, cap. 1.2).

³⁶ Kockelman, *Last Words*, Prickly Paradigm Press: Chicago, 2024, p. 27.

³⁷ Weatherby, Leif, op. cit., p. 104.

³⁸ Gilles Deleuze ha scritto a più riprese sullo strutturalismo a partire dall'opuscolo del 1966, riferendovici anche nel celebre scritto del 1969 *Logica del senso*. Cfr. Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo?*, SE: Milano, 2004. Cfr. ancora Deleuze, Gilles, *Logica del senso*, Feltrinelli: Milano, 1975. In virtù di una maggiore compattezza e approfondimento dei temi teoretici che stanno al cuore della questione dello strutturalismo, il presente scritto rinverrà maggiormente all'opuscolo del 1966.

la serie di vettori computizzabile, il significante, si trasforma, a sua volta influenzando sul significato che possiamo attribuire alle etichette linguistiche. Ogni serie non può mai essere completamente riducibile all'altra però. Ogni elemento non esiste al di fuori delle relazioni che intrattiene con gli altri: questo si intende con il termine tecnico di "distribuzione". La semantica distribuzionale, infatti, recentemente è tornata in auge proprio in virtù del suo allineamento teorico con i LLMs. Per completare una frase, quindi costruire un *response* si tratterebbe di interrogare il *prompt* rilasciato non secondo un suo significato (che cosa vuol dire?), ma di prevedere quale concatenamento di termini sia più *probabile* seguire; quindi fornendo dei termini che con una certa probabilità seguono da quelli che sono giunti come *prompt*. La prospettiva linguistica da intraprendere allora è debitrice nei confronti della teoria della comunicazione. Ci si deve richiamare a Claude Shannon per comprendere come la prospettiva strutturale intervenga prima di ogni ricorso alla semantica, alla sintassi. Le due prospettive sopracitate, quella scientifica e quella sociologica si accordano sul presupposto che ogni parola è vista non per il suo significato, ma come probabilisticamente connessa a una trama sintagmatica di altri elementi.

1.1 LA PROBABILITÀ "SHANNON"

Il lavoro presente avanza delle considerazioni che non si possono chiudere al campo verbale della lingua, anche se questa ne costituisce il principale bersaglio, essendo il *medium in auge* per i LLMs. Come già Warren Weaver fa notare da principio: «*il termine comunicazione sarà da noi usato in un senso molto ampio per comprendervi tutti i procedimenti, attraverso i quali un pensiero può influenzare un altro*»³⁹, la parola non è il referente privilegiato, ma semplicemente uno dei possibili, di cui si accerta già la possibile interconnessione. Per chiarezza espositiva, ora si procederà ad analizzare l'"universo del segnale" come lo definisce Eco⁴⁰, ossia l'interazione che avviene tra macchine e in cui si dice l'essere umano vi sia escluso; un codice simbolico presiede ogni *possibile* intuizione semiotica, già programmata in partenza. La presenza umana andrebbe invece a complicare l'interpretazione simbolica, aprendo il dibattito a problemi anche di stampo ermeneutico. D'altronde, ogni problema legato alla semantica e alla dimensione pragmatica ineliminabile, discenderebbero da questo punto di vista dal livello sintattico, ovvero da come il processo teorico funziona. È impossibile però astrarre le tre dimensioni,

³⁹ Shannon, Claude, E., Weaver, Warren, *Teoria matematica delle comunicazioni*, ETAS Kompas: Milano, 1971, p. 1.

⁴⁰ Eco, Umberto, *La struttura assente*, Bompiani: Milano, 1994.

perciò già parlando della prima si avanzano delle limitazioni che hanno una ricaduta sulle altre.

Preliminarmente, il quadro di una comunicazione si compone così: una fonte [*information source*], detta sorgente dell'informazione, invia dei messaggi grazie a un trasmettitore di segnali [*transmitter*] a un ricevitore che li riconverte in messaggi che il destinatario riceve. Ogni trasmissione comporta una, anche minima, alterazione del segnale, chiamata rumore [*noise source*].

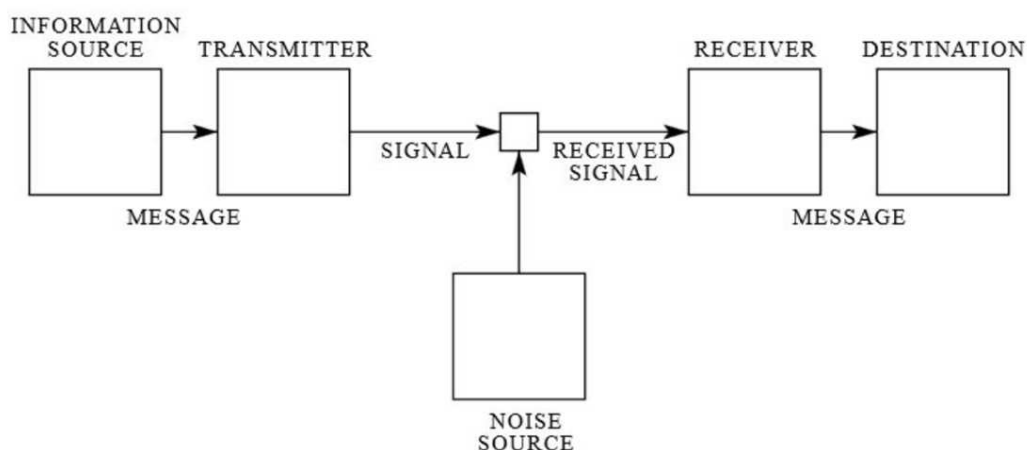


Figura 1 Schema dell'informazione per Shannon⁴¹

Si nota già da questa breve descrizione che *informazione* e *messaggio* non sono sinonimi; ancora, informazione e significato sono termini distinti. L'informazione potrebbe perciò essere considerata anche come nonsenso, ma comunque densamente significativa. Senso e nonsenso sarebbero perciò condizioni preliminari della codificazione del messaggio. L'informazione rappresenta una quantità di biforcazioni o disgiunzioni binarie di scelte. Quello di cui si sta parlando a livello di informazione non è un significato (problema semantico) ma dei *bit* di cosa si potrebbe dire, di quali messaggi si potrebbero ricevere in questa situazione; qualsiasi relè consente così di avere a disposizione già due informazioni (1,0) in base alla sua situazione di apertura/chiusura; perciò, risulterà un utile strumento per affinare il rapporto tra informatica e comunicazione. L'informazione è calcolata, nei casi più semplici, dal logaritmo del numero di scelte possibili. Il modello della comunicazione si svolge così per opposizioni binarie, che indicano che lì interviene una scelta tra due opzioni⁴². Ogni informazione quantitativamente riflette il grado di

⁴¹ Immagine da <https://webcrew.it/teoria-informazione/>.

⁴² Altra questione è definire se questo modello sia adeguato e se anche nelle situazioni complesse la logica binaria possa rappresentare una base adeguata.

libertà di codifica del messaggio: la quantità di informazione dice solo tra quanti messaggi si può scegliere.

Il concetto di informazione si lega a un altro fondamentale concetto della termodinamica, quello di entropia. Shannon si ispira al fisico Ludwig Boltzmann, ritenendo che ogni sistema tende a una situazione più “caotica”, ovvero alla proliferazione del movimento delle particelle che diventa sempre meno prevedibile perché sempre più omologato alla distribuzione del sistema. Allo stesso modo, a lungo andare il rumore, ovvero il disturbo del messaggio, lo renderà sempre più difficile da riconoscere. L’equiprobabilità, quindi la pari possibilità che un messaggio sia ricevuto rispetto ad altri è sorretta da due condizioni: rumore e informazione, vicendevolmente legate tra loro. È qui precisamente che si crea uno dei grandi punti in comune tra linguistica e teoria dell’informazione, come rilevato da Roman Jakobson⁴³. Sia Jakobson sia Shannon fanno allora riferimento all’alfabeto come sistema di simboli a cui seguono delle regole sintattiche precise a seconda della lingua. La “dimensione sintagmatica” direbbe Jakobson sarebbe allora le possibili combinazioni che possono essere assemblate a partire dai fonemi, unità minime di suono che costituiscono le parole, frasi, testi. Importante sottolineare inoltre come non sia possibile che una probabilità sia 1 o 0, ossia che l’avvenimento ricercato sia completamente necessario oppure impossibile.

Ogni messaggio è quindi legato probabilisticamente a determinate risposte, quindi ad ulteriori messaggi. Ogni parola, ogni *token*⁴⁴, si può vedere come probabilisticamente legato all’altro. Ancor di più, ogni morfema è probabilisticamente legato all’altro, come si avrà modo di vedere più avanti. La teoria dell’informazione allora può essere vista come un presupposto teorico dei LLMs. Kockelman, ad esempio, nonostante ne parli attraverso i concetti di Charles Sander Peirce (segno, interpretante, oggetto, agente), descrive il processo semiotico macchinico in termini probabilistici. Ad ogni *input* corrisponde un *output* che ne è probabilisticamente legato; ma come è possibile legarli assieme? Ogni *input* deve essere scomposto nei suoi *tokens* e ognuno di essi deve essere commisurato a una rete di rapporti⁴⁵ che ne descrivono la probabilità con cui possa apparire nell’*output* della macchina. L’alternarsi di *input-output* descrive il rapporto tra “agenti”, ossia i soggetti dell’interpretazione dei segni, che ne donano il senso.

⁴³ Jakobson, Roman, *Structure of Language and Its Mathematical Aspects* (Vol. 12). American Mathematical Society, 1961.

⁴⁴ I *tokens* sono unità a-significative che potrebbero essere simili ai fonemi e ai sememi della linguistica e della semiotica. Si tratta però di unità che frammentano il linguaggio non secondo un criterio umano (ad esempio la fonologia) ma secondo delle scelte puramente macchiniche.

⁴⁵ È esattamente qui che interviene il processo di *attention* QKV di cui si parlerà nei prossimi paragrafi. Per adesso basti un discorso preliminare.

Poniamo due agenti: l'uomo e la macchina. Nonostante i due processi semiotici siano diversi nelle intenzionalità e nei momenti che li contraddistinguono, entrambi sono formalmente equivalenti, il che vuol dire che *«i parametri delle macchine sono collegati ai valori delle persone e, di conseguenza, allineati ad essi»*⁴⁶. L'agente infatti è il punto di partenza del processo, è colui che lo mobilita e che trasforma un segno in un interpretante. L'aspetto più importante da sottolineare del processo tripartito di Peirce è infatti l'apertura che offre, la sua indefinita iterazione che fa di un interpretante il segno per un ulteriore interpretante.

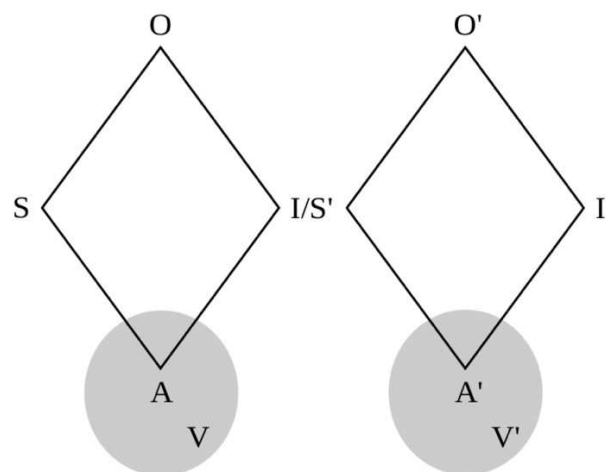


Figura 2 Concatenamento semiotico⁴⁷

Peirce descrive un segno come qualsiasi occorrenza che rappresenta qualcosa per l'agente (A); un segno S quindi può rimandare ad altro (l'oggetto O) in diversi modi: può essere infatti un "indice" e rimandare ad altro fisicamente o causalmente (come il fumo rimanda al fuoco) o un'"icona". Il segno è iconico se rassomiglia fisicamente al suo oggetto; la sua forma infatti rimanda in parte all'oggetto, come una fotografia può rimandare al soggetto della foto o una mappa può rimandare a un territorio. L'oggetto è allora ciò a cui il segno rimanda, separato (Kockelman parla di una "barra") dal segno; è come se il segno manifestasse qualcosa che rimane nascosto, non visibile attraverso di esso. Infatti, l'oggetto può essere conosciuto solo attraverso il suo segno: nel caso degli LLMs però, questo è l'aspetto importante, non bisogna credere che sia un referente esterno del mondo fisico. La prospettiva referenziale del linguaggio fa credere che l'oggetto sia il referente; la referenzialità può valere per l'uomo: *«è la barra che separa il significato dall'oggetto: ciò che può essere percepito direttamente dall'agente si trova da una parte; ciò che può*

⁴⁶ Kockelman, Paul, *Last Words*, Pickly Paradigm Press: Chicago, 2024, p. 19.

⁴⁷ Immagine da *ibid.*

essere conosciuto solo indirettamente (tramite il segno) si trova dall'altra parte»⁴⁸. L'interpretante *I* è invece l'effetto che si produce a partire dal senso dell'oggetto: può essere visto come il prodotto dell'azione dell'agente per mezzo del segno. Insomma, l'agente interpreta un segno attraverso un oggetto, producendo un interpretante. Nel caso degli LLMs l'interpretante è la misura della probabilità che un token si trovi a seguire quelli del prompt e quelli che lo precedono nell'elaborazione dell'output.

L'aspetto però più importante della semiosi di Peirce è che la tripartizione segno-oggetto-interpretante si apre a una catena infinita di rimandi. Due segni, perciò, non possono dirsi recisi completamente l'uno dall'altro, ma rimangono possibilmente legati anche da una minima probabilità. Si capisce l'interpretazione si apre a un campo indefinito di significati, esibendo un presupposto per larga parte degli autori del cosiddetto "postmodernismo"⁴⁹. Un interpretante diventerà a sua volta segno per un ulteriore interpretante: questo è il processo che Kockelman definisce come "concatenamento", aprendo la catena semiotica.

Le regole sintattiche aiutano a restringere il canale di comunicazione, quindi complessificare il codice. Più il codice diventa complesso, maggiori saranno i simboli da utilizzare, creando così ridondanza. Appoggiando il messaggio su una forma di ripetizione, la ridondanza è l'aspetto ordinato e regolato della trasmissione che permette di ridurre al minimo ogni alterazione del messaggio (rumore). Rumore e ridondanza sono due aspetti correlati negativamente per Shannon e Weaver. Più aumentano le possibilità di trascrizione simbolica, minore sarà la possibilità di mancare d'interpretazione del messaggio che giunge al destinatario. È allora il codice che crea delle probabilità prefissate, riducendo il rumore, quindi incanalando l'informazione trasmissibile. L'entropia effettiva si dice di un'informazione che giunge al destinatario; essa è una possibile scelta su un campionario più vasto di messaggio. La sua probabilità è allora calcolata come rapporto tra gli eventi accaduti per la trasmissione del messaggio attuale e la totalità degli eventi che sarebbero potuti accadere e si viene a indicare col termine di "entropia relativa". Ogni entropia relativa viene così a rappresentare un valore compreso tra 1 e 0. Complementare all'entropia relativa è la ridondanza, ossia le operazioni sintattiche che bisogna ripetere per diminuire il rumore; essendo ripetitive possono venire cancellate dal messaggio, permettendone comunque la comprensione. A questo punto si può introdurre la formula con cui calcolare l'informazione:

⁴⁸ Kockelman, op cit., p. 10.

⁴⁹ Umberto Eco, Jean François Lyotard, Paul Ricoeur tra gli altri.

$$H = - \sum p(x) \log_2(x)$$

Il segno meno è presente in virtù dei valori negativi che restituisce il logaritmo di una probabilità, sempre contenuta tra 0 e 1⁵⁰. La sommatoria indica invece che il grado di libertà che un'informazione veicola è dovuta a una serie di biforcazioni o disgiunzioni da intendere come somma. Perché il logaritmo? Senza indugiare in una dimostrazione matematica, basti una spiegazione intuitiva. Se otto sono i possibili messaggi che si possono inviare, allora è il logaritmo a darci la quantità di informazione: $H = \log_2 8 = 3 \text{ bit}$. In altri termini, ogni informazione sarà di 3 bit se 8 sono i messaggi tra cui si può liberamente scegliere. Si notino i casi limite, ovvero in cui la probabilità è 0 o 1: il primo caso risulta impossibile (nessuna occorrenza immaginabile è impossibile, come ad esempio “tttddgggtj” anche solo considerandola entro un discorso metalinguistico come quello presente), mentre nel secondo caso l'informazione H sarà uguale a 0, quindi del tutto irrilevante. Pertanto, H ha valore massimo quando le due probabilità sono uguali e si riduce a zero quando non ci sono libertà di scelta. L'informazione maggiore si ha quando la scelta è tra molti casi simili tra loro e in cui i condizionamenti sono quasi nulli (nel caso ideale della moneta in cui la riuscita è pari al 50%). Al contrario, se si ha un messaggio che molto probabilmente possa venire ricevuto, allora H verrà ad assumere un valore piccolo, vicino a 0, poiché l'informazione, l'entropia, quindi la scelta e l'incertezza, sono scarse.

Ci sono due modi per far crescere H: dato un insieme di messaggi, tanto più è uguale la loro probabilità di trasmissione, tanto più alta sarà l'informazione, quindi grazie al valore della probabilità attorno a $\frac{1}{2}$. Si potrebbe sennò accrescere il numero di messaggi che è possibile inviare, tra cui scegliere: la somma accrescerebbe di conseguenza e il valore dell'informazione trasmessa di conseguenza. L'informazione è strettamente connessa a un processo stocastico di calcolo delle probabilità. È questo l'aspetto che maggiormente interessa e che motiva una riproposizione di queste teorie. *Il linguaggio, in quanto canale di comunicazione, quindi di trasmissione di informazione, prima di essere una funzione di referenza o di significanza è una trama di relazione tra i suoi elementi.* Allora i simboli si diranno co-dipendenti e calcolabili secondo una loro distribuzione stocastica. Più precisamente, il processo stocastico viene definito *ergodico*, in cui «ciascuna sequenza prodotta mediante il processo possiede le stesse proprietà

⁵⁰ Per la spiegazione matematica della formula a livello intuitivo si consiglia la lettura di Stewart, Ian, *Le 17 equazioni che hanno cambiato il mondo*, Einaudi: Torino, 2017, pp. 407-444.

statistiche», così che le frequenze degli elementi (lettere alfabetiche, note musicale, gesti...) «si avvicineranno all'aumentare delle lunghezze delle sequenze, a limiti definiti, indipendenti dalla sequenza particolare»⁵¹. Un sistema è stocastico se quindi c'è omogeneità statistica tra i suoi elementi; Shannon applicò questi presupposti teorici alla lingua inglese per renderla crittografabile per il telegrafo, partendo dalle ricerche matematiche di Andrej A. Markov. Nonostante il diverso ambito di applicazione (qui la linguistica computazionale, lì la genesi biologica delle forme), è utile richiamarsi al contributo del filosofo e biologo Raymond Ruyer⁵². La sua riflessione, che intreccia teoria dell'informazione, cibernetica e biologia individua proprio nella catena markoviana un modello alternativo alla dialettica hegeliana nello sviluppo biologico degli esseri viventi e, aspetto qui più rilevante, delle strutture sociali, culture e del *linguaggio*.

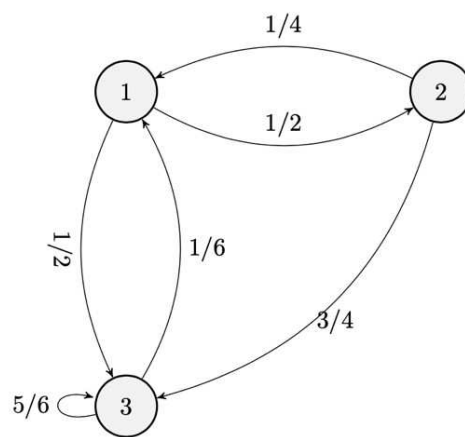


Figura 3 Catena di Markov⁵³

Lo sviluppo markoviano viene descritto come occupante una posizione mediana tra disordine (rumore) e l'ordine di un codice. A seconda del grado di codificazione, dovuta alla chiusura del circuito dell'informazione rispetto a possibili modificazioni esterne, la catena di Markov tiene assieme necessità dell'ordine e aleatorietà della contingenza; quello che, come si vedrà nei prossimi capitoli, corrisponde alla “necessità della contingenza”, ossia la precipua modalità logica di qualificare il dispiegamento di una struttura simbolica di segni nel tempo per Althusser. Come si vede dalla figura 3, l'andamento è probabilistico e differentemente distribuito tra ognuno dei tre termini. Una volta che si è giunti al termine 3, ciò che conta è solamente la probabilità dei diversi legami e non l'intera storia passata delle relazioni. Ovvero, preso l'elemento 3, esso ha

⁵¹ Shannon, Claude, E., Weaver, Warren, op. cit., p. 51.

⁵² Ruyer, Raymond, *La genesi delle forme viventi*, Bompiani: Milano, 1966.

⁵³ Immagine da <https://www.gameludere.it/2025/01/11/processi-stocastici-e-catene-di-markov-teoria-e-applicazioni/>.

avrà molta più probabilità di rimandare a sé stesso (5/6) anziché di rinviare all'elemento 1 (1/6), prescindendo dall'originaria provenienza (potrebbe essere reiterato su sé stesso, proveniente dall'elemento 1 oppure dall'elemento 2).

Il disturbo non è un concetto opposto a quello di informazione, ma a quello di ridondanza. Il rumore (o disturbo) è l'alterazione, quindi l'incertezza con cui si riceve un messaggio. Maggiore è l'informazione, maggiore è l'incertezza che può caratterizzarla. Il disturbo è un'informazione falsa e indesiderabile, quindi *«per ottenere l'informazione utile bisogna eliminare dal segnale ricevuto questa parte falsa»*⁵⁴, ossia il rumore. Il rumore si calcola attraverso il concetto di equivocazione, ovvero si definisca equivocazione la misura dell'incertezza media del messaggio (o della sorgente) quando il segnale è noto.

Che ci siano uno o più livelli di trasmissione dell'informazione appare come una conseguenza plausibile, da chiarire. Si riconosca ad esempio la fonte o la sorgente dell'informazione e in un secondo momento il trasmettitore. La fonte può tradurre molta più informazione e scegliere solo tra alcune probabilità preposte e rappresentate in appositi simboli per significare dei messaggi. Il trasmettitore dovrà tradurre gli *input* della sorgente in ulteriori *output* secondo certi segnali a cui è stato codificato. Ogni trasmissione tra *input* e *output* avviene producendo uno scarto tra informazioni, ma, per come si è visto sopra, anche tra rumori. Infatti, *«per quanto si sia abili nel processo di codifica, sarà sempre vero che, una volta ricevuto il segnale, resterà qualche incertezza indesiderabile (perché dovuta al disturbo) su ciò che era il messaggio»*⁵⁵.

Shannon, assieme a Norbert Wiener, riuscì, attraverso un'intuizione scientifica e una teoria dalle applicazioni ingegneristiche a proporre una concezione del linguaggio peculiare: da un lato, affine alla visione della linguistica di Saussure, annullando la primarietà del referente; d'altro, considerando l'informazione come una quantità di *bit* relativi alla trascorsa discussione, quindi ancorata su un campione di parole passate, quindi probabilisticamente stimate. Concetto, quello di informazione, che, per come è stato descritto, fa tutt'uno con quello di incertezza e solo secondariamente si relazione a quello di significato. La questione semantica sembra abbandonata e lasciata al caso, vista la celebre asserzione di Shannon per cui *«gli aspetti semantici della comunicazione sono irrilevanti dal punto di vista ingegneristico»*⁵⁶.

Con ciò Shannon si propone di dire che per trasmettere un messaggio non importa

⁵⁴ Shannon, Claude, E., Weaver, Warren, op. cit., p. 20.

⁵⁵ *Ibid.* p. 23.

⁵⁶ *Ibid.* p. 55.

conoscerne il significato, idea questa che potrebbe allora indurre a credere che i mezzi di comunicazione siano meri ripetitori vuoti, dei “pappagalli stocastici”⁵⁷. Non bisogna cadere in facili conclusioni, credendo che il significato venga squalificato dall’ambito della teoria informatica della comunicazione. Anzi, per come dimostra anche Weaver, il problema “sintattico”, ossia della tecnica di trasmissione dei messaggi coinvolge di conseguenza anche i problemi semantici del loro significato. Infatti, la nozione di informazione è statistica, prima ancora di essere qualitativa per come viene intesa secondo il senso comune. Allo stesso tempo la distribuzione si applica a un ventaglio di messaggi; quindi, di segni che da un punto di vista linguistico competono immediatamente a un significato: sintassi e semantica sono inscindibili. È una semplificazione ritenere che il significato sia lasciato fuori dalla questione della trasmissione. Tanto che il lavoro di Shannon venne pensato inizialmente per problemi da far svolgere ai calcolatori oppure per la questione della traduzione tra diverse lingue; tanto da ammettere che «*o si deve affermare che un tale calcolatore “pensa”, oppure bisogna sostanzialmente modificare il modo convenzionale di intendere il verbo pensare*»⁵⁸, di fatto ciò che sta accadendo attorno al dibattito sul termine “intelligenza” artificiale⁵⁹.

Dunque, il linguaggio sembra significare senza obliterare la sua funzione di referenza, ma semplicemente tramite calcolo statistico. Si potrebbe anche dire che del significato si genera prima ancora di essere riconosciuto attraverso gli strumenti della sintassi e della semantica. Prima di poter scomporre analiticamente è come se già una “forma significativa” si imponesse, facendo del linguaggio una ridondanza di pattern. Giustamente Weatherby fa notare che di forma si è sempre parlato o a partire da una fenomenologia, dalla costituzione esterna (*eidos*) oppure come sviluppo interiore dell’individuo (*morphé*). In questo caso si vede però come l’interiorità del significato, la definizione della parola, fa tutt’uno con l’intelaiatura sintattica, quindi la sua distribuzione stocastica. La prospettiva strutturalista terrebbe entrambi gli aspetti assieme, definendo il concetto di forma come una calibrazione tra i due.

⁵⁷ Kockelman, Paul, *Last Words*, Pickly Paradigmatic Press: Chicago, 2024, p. 59.

⁵⁸ Shannon, Claude E., Weaver, Warren, op. cit., p. 27.

⁵⁹ Tra tutti, una figura che si è prodigata a scongiurare il termine “intelligenza”, pregno di ambigui riferimenti è Luciano Floridi, che ritiene i modelli dotati di zero intelligenza. Cfr. Floridi, Luciano, *La quarta rivoluzione*, Cortina: Milano, 2017.

CAP. 2: LA STRUTTURA E LE SUE CONDIZIONI FORMALI

2.0 L'ORDINE SIMBOLICO

Sta diventando opinione comune che per trattare i LLMs si possano considerare le loro matrici come delle strutture, nel senso con cui lo strutturalismo le ha definite nel corso

degli anni '50-'60. Diversi autori⁶⁰ ritengono sia utile approfondire il funzionamento dei LLMs attraverso linguisti come Roman Jakobson⁶¹. La linguistica, infatti, come fa notare Françoise Dosse⁶², gioca un ruolo fondamentale nel preparare il campo a quella vasta stagione pluri-decennale che è stato lo strutturalismo. Molte idee strutturaliste sorgono a partire dagli interessi della linguistica:

Una considerazione sistematica delle multiformi relazioni tutto-parte amplia considerevolmente l'ambito della nostra scienza, consente un'analisi sistematica dei messaggi verbali sia per quanto riguarda il codice sia per quanto riguarda il *codice* sia per quanto riguarda il *contesto* e, dalle unità più grandi a quelle più piccole, svela la complessa interazione tra i vari livelli del linguaggio, come pure la costante interdipendenza tra diverse funzioni verbali.⁶³

Questa citazione enfatizza il lavoro che molti pensatori mossero nei confronti di una molteplicità di relazioni che si possono annettere entro un'unica *unità*. Sicuramente, sottolineare la primarietà della struttura fa di questi pensatori un terreno di scoperte del tutto inattese e uno spazio per una ricerca plurisetoriale tra psicanalisi, etnografia, antropologia, teoria letterarie e filosofia. Il clima di multidisciplinarietà, di eclettismo letterario e scientifico che si può respirare leggendo le trascrizioni delle conferenze di Lévi-Strauss e di Jacques Lacan, testimonia un decentramento dall'umanismo classico, ancora oggi però circolante in dibattiti come quello di Weatherby⁶⁴; a non rendere giustizia a questa effervescenza culturale sono interpretazioni del movimento culturale ormai diventate classiche, come quella di Ricoeur⁶⁵, che ritengono lo strutturalismo come un ritorno ai grandi "maestri del sospetto" attraverso la lente di Saussure.

Nel clima instabile del secondo Dopoguerra, molte università francesi⁶⁶ ricevettero molti fondi, ad esempio da programmi economico-finanziari come il piano Marshall, per instaurare una rete di laboratori. La ricerca americana si serviva infatti di

⁶⁰ Weatherby, Leif, *Language machines*, Minnesota University press: Minneapolis, 2025. Geoghegan, , Bernard D., *Code: from Information theory to French Theory*, Duke University press: London, 2023. Hui, Yuk, *On Existence of Digital Objects*, Minnesota University press: Minneapolis, 2016.

⁶¹ È questo il caso che maggiormente sarà preso in considerazione, in rapporto a Leif Weatherby. Cfr. Weatherby, Leif, *Language machines*, Minnesota University press: Minneapolis, 2025.

⁶² Cfr. Dosse, Françoise, *History of structuralism*, vol. 1, Minnesota University press: London, 1997.

⁶³ Jakobson, Roman, *Parti e totalità nel linguaggio* in *Qualità e quantità e altre categorie della scienza*, Bollati Boringhieri: Torino, 1971, p. 210. Corsivo nostro.

⁶⁴ Weatherby nella sua ricerca lo qualifica come «remainder humanism», ad oggi usato dai critici come strumento per separare qualitativamente il dominio dell'umano da quello della macchina. Cfr. Weatherby, Leif, *Language machines*, Minnesota University press: Minneapolis, 2025.

⁶⁵ Ricoeur, Paul, *Freud and Philosophy: An Essay on Interpretation*, Yale University Press: New Haven, CT, 1970.

⁶⁶ La Sorbona tra tutte si oppose a questa tendenza "tecnocratica".

progetti di ricerca multidisciplinari⁶⁷, che andarono a reclutare sempre più ricercatori, facendo del critico intellettuale sempre meno un'autonoma voce nel dibattito dell'epoca e sempre più un produttore culturale stipendiato, integrato in gruppi di ricerca dotati di costose attrezzature e coinvolto in progetti a lungo termine⁶⁸. L'interesse delle varie branche di ricerca si intreccia sempre più con la cibernetica e la teoria dell'informazione, in particolare articolando una riduzione dell'uomo alla macchina e viceversa; consapevoli di questa svolta, i cosiddetti pensatori "strutturalisti" cercarono di svilupparla senza alcuna influenza politica di sorta, rimanendo il più possibile neutrali, tanto che Michel Foucault li descriverà come «una nuova schiera di pensatori specializzati, al pari dei fisici nucleari e dei tecnici del computer»⁶⁹. Così la cultura diviene un codice che è possibile a tutti gli effetti programmare, nel senso di una sua descrizione formale, al pari delle macchine informatiche; scuole, centri di ricerca, letteratura, scienza diventano spazi di continua codificazione e decodificazione. Il paradigma macchinico, che nei decenni precedenti era stato influenzato dalle riflessioni di celebri antropologi e filosofi attorno alla questione dell'organicismo⁷⁰, in particolare nell'articolare il rapporto uomo-ambiente, al pari di quello della macchina attraverso un continuo ricircolo di *input* e *output*, arriva ora a invadere diversi campi di ricerca; la nozione di soggetto, trattato secondo i concetti della fenomenologia oppure dell'esistenzialismo⁷¹ si frammenta in una miriade di pratiche di conoscenza, testimoniando «la nuova centralità della conoscenza teorica, il primato della teoria sull'empirismo e la codificazione della conoscenza in sistemi astratti di simboli che possono essere applicati a circostanze diverse e variegate»⁷². Come avrà modo di chiarire Gilles Deleuze la questione del soggetto nello strutturalismo:

Lo strutturalismo non è affatto un pensiero che sopprime il soggetto, ma un pensiero che lo frantuma e lo distribuisce in modo sistematico, che contesta l'identità del soggetto, lo dissipa e lo fa passare di posto in posto, soggetto sempre nomade, fatto di individuazioni ma impersonali, o di singolarità, ma preindividuali.⁷³

⁶⁷ Si vedrà nei prossimi paragrafi, ad esempio, la ricerca di Roman Jakobson finanziata dal Rockefeller Center, il laboratorio del MIT e di Harvard.

⁶⁸ Geoghegan, Dionysius, Bernard, *Code: from Information Theory to French Theory*, Minnesota University press: 2023, p. 108.

⁶⁹ Foucault, Michel, "Truth and Power" (1977), in *Power/Knowledge: Selected Interviews and Other Writings*, Colin Gordon: New York, 1980, p. 131.

⁷⁰ Pasquinelli, Matteo, *The Eye of the Master*, Verso: Londra, 2021, cap. 6.

⁷¹ Dosse, Françoise, *History of structuralism*, vol. 1, Minnesota University press: Minneapolis, 1997.

⁷² Geoghegan, Dionysius, Bernard, op. cit., p. 104.

⁷³ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo*, SE: Milano, 2004, p. 48.

Non importa cosa agisce, ma *come*, attraverso quali strategie o forze, che muovono quell'unità inscalfibile e ultima che è il soggetto. L'analisi critica, quindi, non destituisce il soggetto, ma studia quali processi lo attraversano e lo trasformano. Questi non possono essere derivati né dall'esperienza del soggetto né da un'essenza primordiale, ma costituiscono intimamente la trama di relazioni simboliche. Fare filosofia, dunque detenere una ferma e solida conoscenza del reale, diventerà impossibile se non si scende a patti con la griglia fuorviante di connotati ideologici; la filosofia diventa non *Bewusstsein* del *Sein* in generale, ma coscienza della scienza. Se ci si fermasse alla prima definizione ancora non disporremmo della potenza teorica del simbolico, il semplice riflesso potrebbe dirsi al più un gioco di specchi dell'immaginario sul reale sempre "sul modo del doppio"⁷⁴. Con il ritorno al reale per Althusser finisce la filosofia e comincia la critica, finiscono gli abbagli ideologici e ci si può finalmente vedere "da fuori", cominciare a conoscersi rispetto al luogo che si occupa nella conoscenza.

Per simbolico si intende un ordine di leggi che non è né meramente fisico né solo psicologico. Le leggi simboliche plasmano comportamenti, idee e desideri, sono l'asse portante della triade sapere-potere-desiderio. Lacan⁷⁵ lo definisce come un terzo regno, a lato dell'Immaginario e del Reale; dunque, impossibile da annettere entro la costituzione del Sé [*moi*] e nemmeno come ordine fisico-materiale; è non immaginabile, dunque sfugge a qualsiasi identità del concetto, ma è anche irreal, nel senso che non abita la materia. Il simbolico sembrerebbe un «*sistema che conosce solo il suo stesso ordine*»⁷⁶, una forma più che una sostanza, poiché non è calato nella materia. È questo un primo aspetto che lo strutturalismo eredita da Ferdinand de Saussure; per precisare, Saussure non ha mai parlato di simbolico, ma è riuscito a fare della linguistica una branca di studi scientifica, concependo il linguaggio come un sistema chiuso in sé stesso, una forma fluttuante e recisa dal referente. Anche sotto questo aspetto, come si vedrà, Weatherby segue fedelmente la precettistica strutturalista. Tramite l'ordine simbolico si può formalizzare la realtà senza appoggiarsi su significati definitivi né a partire dagli oggetti fisici. Quello che per Saussure è un gioco di segni (linguaggio), diventerà poi un gioco⁷⁷ di simboli. La struttura simbolica si può allora qualificare invece vicina a quell'ordine di relazioni impersonali e di singolarità preindividuali di cui parla Deleuze, ma verso cui

⁷⁴ *Ibid.*, p. 14.

⁷⁵ Qui si sta prendendo in considerazione la prima riflessione di Lacan, fino agli anni '60, perciò prima dei celebri seminari sullo schema RSI secondo i diagrammi di Wenn.

⁷⁶ Saussure, Ferdinand de, *Corso di Linguistica Generale*, tr. It. a cura di T. De Mauro, Laterza: Roma-Bari, 1967, p. 126.

⁷⁷ Il termine non è usato a caso, ma ripreso da molti autori, tra cui Deleuze. Cfr. Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, p. 90.

Lacan più volte ritorna nei suoi seminari, servendosi di nozioni prese dalla cibernetica e dalla teoria dell'informazione. Il simbolico è *irreale* e *non immaginabile* perché è generativo, ossia costituisce le stesse condizioni che determinano il gioco strutturale.

Il simbolico come elemento della struttura è al principio di una genesi: la struttura si incarna in realtà e immagini seguendo serie determinabili; ancor di più essa le costituisce incarandovisi, ma non ne deriva, essendo più profonda di esse, sottosuolo per tutti i suoli del reale come per tutti i cieli dell'immaginazione⁷⁸.

Gli elementi che lo costituiscono non hanno un significato precipuo e nemmeno una "forma"; essi costituiscono una combinatoria di elementi formali. Combinatoria perché indefinitivamente mescolabili, dando luce a costruzioni di modelli interminabili. Non significa che i simboli siano mescolati a piacere indefinitivamente, ma che lo sviluppo diacronico segue da un loro rimescolamento. Questo aspetto distingue il simbolico sia dalle teorie della *Gestalt*, non potendo indicare alcuna forma che a priori governi integralmente la struttura: ci sono parti che rendono conto della formazione del tutto e delle variazioni delle parti.

Deleuze accetta però definizione di simbolico che introduce a una dimensione ulteriore da quella puramente formale entro cui gli esperti della linguistica sembravano confinarlo. Infatti, secondo lui il dinamico intreccio tra i tre regni non può rimanere descritto come tra tre registri differenti. Il reale è in sé "uno", ovvero tende a totalizzare l'intero regno degli esseri. L'immaginario consegue invece da un'operazione di sdoppiamento⁷⁹, applicabile su due o più enti ugualmente reali. Affermare che il simbolico non è uno, né solo due, ma *tre* significa rifuggire da qualsiasi riduzione identitaria: è *irreale* perché non poggia su niente, ovvero non ha carattere fisico-estensivo, ma sembra aleggiare, insistere sulle cose senza alcuna effettiva corporeità. È *non immaginabile* perché di per sé privo di identità, di significato, ma sempre incarnato. È come se fosse un "tema", per prendere in prestito un concetto di Ruyer, che controlla qualsiasi atto individuante, un tema unitario che tiene assieme qualsiasi spezzettamento, qualsiasi corso di sviluppo. Dunque, influisce, senza essere però della *stessa natura*: la differenza di natura allora non è una differenza di grado, di proporzioni tra corpi ugualmente estensivi, ma entra così in gioco la categoria intensiva cara a Deleuze.

⁷⁸ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo*, SE: Milano, 2004, p. 13.

⁷⁹ «L'immaginario è definito da giochi di specchi, di sdoppiamenti, di identificazioni e proiezioni rovesciate, sempre sul modo del doppio» cfr. *Ibid.*, p. 14.

Ciò che è magico, se ci atteniamo al vero senso della parola, è l'esistenza di questi *temi potenziali* senza supporto materiale e spaziale e la convenienza dei tempi specifici, malgrado lo spezzettamento e la diversità dei loro portatori, l'”appoggio” quanto alla loro natura generale.⁸⁰

Quando si parla di simbolico allora in realtà ci si sta proiettando su una dimensione intensiva, secondo una forma espressiva e significante del termine, ovvero ciò che può venire percepito sia nella sua presenza che nella sua assenza⁸¹. Il potenziale-simbolico scardina la semplice combinazione di forme precostituite. Anzi, l'andamento aperto e non regolare impedisce di pensarle come tali.

La “necessità della contingenza” si dispiega secondo un modello di pensiero della complessità, attuandosi tramite una sovradeterminazione di codici. Perché lo sviluppo risulti aperto, nessuna forma a priori può darsi in maniera unitaria, ma solo completa. decade il concetto di casualità lineare: la causa e l'effetto lasciano spazio all'interazione di stimoli-*segnali*⁸². Essi di per sé sono privi di informazione nel senso che essa scaturisce a partire da una trasmissione che perde il carattere direzionale della casualità e si tramuta in uno stimolo-indice che la trasmette. L'indice comunica immediatamente un atto istintivo. A venire criticato è il mito comportamentista, secondo cui il comportamento di un essere biologico può venire assimilato a una ferrea struttura di richiami istintivi, quindi a particolari sollecitazioni localizzate. A governare il comportamento è il tema (simbolico) che, per via del suo carattere intensivo, è irriducibile alle sue molteplici individuazioni. Decade, inoltre, anche il mito determinista: non esistono delle catene di causa ed effetti che possono avverarsi a partire da un supposto stato originario. L'origine non sarebbe ravvisabile se non sulla scorta del presupposto della *fortuita* emersione di

⁸⁰ Ruyer, Raymond, *La genesi delle forme viventi*, Bompiani: Milano, 1966, p. 129. Corsivo mio.

⁸¹ D'altronde anche Eco insiste sulla triplice natura della struttura simbolica: tra presenza e assenza. La *presenza* è tale solo in virtù di un'assenza di elementi attuali, ma *presenti* virtualmente. Cfr. Eco, Umberto, *La struttura assente*, Bompiani: Milano, 1971, pp. 20-24.

⁸² Interazione fondamentale per comprendere il carattere del *corpo senza organi*. Cfr. Parte 3, Cap. 2.1. Si aggiunga di passaggio che Deleuze difende la sua filosofia della differenza da alcune assimilazioni a un'indifferenziazione degli elementi che abitano il piano, come alla maniera di Schelling. La sua non è l'individuazione di un originario caos disordinato, ma un disordine-ordinato, ovvero un *caosmo*, tematica ripresa in particolare in Deleuze, Gilles, *La piega*, Einaudi: Torino, 2004.

Per quanto concerne lo statuto dello stimolo-segnale che riflette la posizione intermedia della catena markoviana (in cui spesso si presenta) tra disordine e ordine del codice, si potrà dire con Ruyer che ogni stimolo presuppone un linguaggio e non viceversa. Sembra star dicendo la stessa cosa che Saussure accorda al rapporto *parole-langue*, ma qui l'enfasi non è sulla virtualità della struttura rispetto all'elemento, ma alla condizione di disordine che non può che seguire da quella dell'ordine. Sembra paradossale come pensiero, eppure il caos primordiale, se preso letteralmente, non potrebbe produrre alcun ordine, allora è necessario che il fenomeno del disordine “rumoroso” non possa che accompagnare la successione di elementi empiricamente individuabili.

fattori che verrebbero poi a succedersi necessariamente. Il determinismo si dimostra perciò possibile solo sul tacito accordo con la sua controparte fortuita: tanto fortuita è l'originaria configurazione di elementi, quanto necessario il suo sviluppo. Nel prossimo paragrafo si riprenderà allora l'idea strutturalista anti-originaria, dimostrando che la struttura, per via della sua configurazione simbolica rinuncia all'origine e quindi a venire esplicitata attraverso un concatenamento lineare e causale, avvicinandosi al pensiero della complessità.

2.1 UN ANCORAGGIO PRATICO

Quando si parla di combinatoria di elementi, questi non hanno *di per sé* né forma, né contenuto, né rappresentazione, né significato, né realtà empirica data, né modello funzionale ipotetico, né intelligibilità dietro le apparenze. Il segno simbolico rimane distinto tanto dal significato, quanto dal suo significante, essendone l'unione; l'unità qualificata di significato-significante è così descritta per Saussure: «*il segno linguistico non unisce una cosa e un nome, ma un concetto e un'immagine acustica*»⁸³. Il significato non è il referente, la cosa esterna a cui si riferisce la parola, così come il significante non è la forma fonica del nome, la sua emissione vocale.

A questo proposito si potrebbe definire il campo del simbolico come un impianto teorico, che Louis Althusser conia col termine “Teoria”, per distinguerlo da qualsiasi costruzione teorica (teoria). Il riferimento al padre dello strutturalismo marxista è tanto più pertinente, quanto si tiene conto che il pensiero della Teoria rifugge tanto da un approccio soggettivistico, quanto olistico (affine si direbbe alle teorie della *Gestalt*). Lo stesso marxismo, rispetto al lavoro qui presente, non interessa da un punto di vista politico, ma prettamente teorico, essendo «un “metodo” d'analisi e di azione, ma anche, in quanto scienza, il *dominio teorico di una ricerca fondamentale*»⁸⁴. La scienza della Teoria viene contrapposta al pensiero irriflesso dell'ideologia. Althusser tratta la coppia di termini scienza-ideologia come paradigma per criticare le posizioni marxiste antecedenti e contraddittorie. Qui interessa il carattere formale che viene ad assumere la Teoria. Ogni concetto è visto come rimando ad altre coordinate teoriche entro una struttura vincolante, circostanziale quindi unificante che prende il nome di problematica: «condizione di possibilità assolutamente definita, e dunque la determinazione assoluta

⁸³ Saussure, Ferdinand de, *Corso di linguistica generale*, tr. It. a cura di T. De Mauro, Laterza: Roma-Bari, 1967, p. 45.

⁸⁴ Althusser, Louis, *Per Marx*, Mimesis: Milano, 2008, p. 27.

delle *forme di posizione di ogni problema*, in un dato momento preso in considerazione»⁸⁵.

La problematica ha esattamente il carattere d'unità sistematica che fa emergere il "senso"; l'ideologia viene quindi considerata come un «tutto reale», unificato dalla problematica in modo tale che se uno degli elementi viene sottratto il senso ne viene alterato. Qui si può rintracciare un secondo connotato della riflessione di Saussure, ossia la sincronicità della struttura, il suo ordine di elementi. L'aspetto sincronico interessa maggiormente il pensiero strutturalista, rispetto alla mera successione storica, come d'altronde anche lo stesso Saussure secondo quanto afferma il linguista Jean-Claude Coquet⁸⁶ poiché questa può essere dedotta a partire dalla combinazione degli elementi simbolici. La formalizzazione (sincronica) del linguaggio permette di costruire un ordine "topologico", ossia di posizione; ogni elemento non ha un significato interno (definizione) o un referente esterno, ma solamente un senso

Non si tratta di un posto in un'estensione reale, né di luoghi in estensioni immaginarie, bensì di posti e luoghi in uno spazio propriamente strutturale, ovvero topologico. Ciò che è strutturale, è lo spazio, ma uno spazio inesteso, pre-estensivo, puro *spatium* costituito passo passo come ordine di vicinanza, dove la nozione di vicinanza ha innanzitutto un senso ordinale e non un significato nell'estensione.⁸⁷

Lo spazio è intensivo nella misura in cui ogni parola è posizionata rispetto alla struttura virtuale delle altre, che continuamente muta e si reitera (Deleuze dice di uno *spatium* che muta passo passo).

La Teoria althusseriana permette quindi di trattare il senso di un dato problema a partire da delle condizioni formali, che si possono qualificare come simboliche. Ritorna l'idea già accennata di "pratiche di conoscenza" entro cui il soggetto si disperde, le quali segnano la fondamentale *eteronomia* della conoscenza. Ogni conoscenza è una "pratica", ossia un processo trasformativo che si trova di fronte a una materia data (questa può essere una forma di conoscenza ideologica, una raccolta di dati, una serie di percezioni oppure una conoscenza scientifica pregressa) e attraverso dei mezzi di produzione (un metodo scientifico) riesce a produrre una nuova forma di conoscenza. La materia data viene denominata Generalità I, i mezzi di produzione Generalità II e il prodotto della pratica Generalità III. È bene ora approfondire questo processo generativo di produzione che

⁸⁵ Althusser, Louis, *et al.*, *Leggere il capitale*, Mimesi: Milano, 2006, p. 26.

⁸⁶ Dosse, Françoise, *History of structuralism*, Minnesota University press: Minneapolis, p. 48.

⁸⁷ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo*, SE: Milano, 2004, p. 21.

servirà per chiarire il delicato rapporto tra simbolico e reale, quindi tra struttura e storia.

Le tre Generalità sono qualitativamente diverse. La Generalità II viene considerata come una “teoria” scientifica, un edificio di nozioni e concetti oppure diverse regioni unificate in teorie regionali coesistenti in un’unità complessa e contraddittoria. I mezzi di produzione costituiscono la parte attiva di questo processo trasformativo, che comporta la rottura dalla Generalità I alla Generalità III: tra le due «*non c’è mai identità d’essenza*»⁸⁸. Generalità I e Generalità III sono rispettivamente l’ideologia e la scienza, l’irriflessa istanza pregiudiziale e il rigoroso ragionamento.

Si ottiene una scienza a condizione di abbandonare il terreno in cui l’ideologia crede di aver a che fare con il reale, cioè abbandonando la sua *problematica* ideologia (il presupposto organico dei suoi concetti fondamentali, e insieme al sistema, la maggior parte dei concetti stessi) per fondare “in un altro elemento”, nel campo di una problematica nuova, scientifica, l’attività della nuova teoria.⁸⁹

Per teoria Althusser intende una pratica teorica scientifica, ossia una certa tipologia di pratica. Rispetto a quanto appena dimostrato, una teoria non si compone solo di una “teoria” (Generalità II), ovvero di un certo sistema teorico di una scienza reale, ma ne è la rottura rispetto al campo di carattere ideologico o scientifico (la materia data espressa dalla Generalità I). L’intero processo epistemologico per giungere a produrre Generalità III prende invece il nome di Teoria. Nonostante la confusione terminologica, la Teoria rappresenta il processo di formazione di conoscenze scientifiche (teorie) a partire da un terreno altro, ovvero quello ideologico, ma pur sempre interno alla pratica teorica.

La Teoria descrive quindi la dinamica della pratica teorica in generale, quindi sia l’essenza della pratica (fintato che la teoria costituisce un dominio specifico affianco a quello della pratica politica, economica, ideologica), sia l’essenza delle trasformazioni in generale, la rottura che produce problematiche. Non è cambiando il nome di un concetto che questo ne viene trasformato, ma solo attraverso il rigoroso lavoro teorico. La teoria, meglio la pratica teorica, definisce pertanto il carattere del simbolico altro rispetto all’immaginazione e alla realtà. L’intera missione di Althusser enuncia un’autonomia della conoscenza rispetto al reale. Ogni norma di pretesa veridicità conoscitiva sul reale è dovuta al tribunale della conoscenza stessa: l’oggetto reale (referente) non pone alcuna problematica. Anzi, la problematica è immanente al dominio della conoscenza, sempre-

⁸⁸ Althusser, Louis, *et al.*, op. cit., p. 162.

⁸⁹ *Ibid.* p. 166.

già un'articolazione di pensiero; pertanto, il loro è un rapporto di conoscenza, non un rapporto reale. Così uno Etienne Balibar qualifica i suoi concetti come «*sempre-già autocritici [...] Sono [...] sin dalla loro origine, un modo discorsivo di persi essi stessi in stato di disequilibrio, di assicurarsi contro al fermezza di una “tesi” nel momento in cui la si sostiene. Sono delle unità di contrario di un tipo abbastanza particolare*»⁹⁰. Nonostante Althusser si attesti su un terreno di politico e ancor prima economico, attraverso un'operazione di rilettura dei testi marxiani, le sue indicazioni teoriche sono utili al ragionamento presente. Esse non solo rispondono bene alle eco strutturaliste del tempo, incarnando a pieno le tendenze di pensiero, ma ci permettono di trarre alcuni strumenti concettuali utili anche per comprendere la natura dei LLMs.

Principalmente la problematica althusseriana serve per esibire dei concetti che criticamente abbattano le costruzioni ideologiche, fondate su dicotomie come essenziale-inessenziale oppure esterno-interno o semplice-complesso; il meccanicismo scientifico e la dialettica hegeliana⁹¹ si dimostra che rispondono a questi presupposti ideologici. Ad esempio, la dialettica hegeliana viene fortemente criticata perché confonde astrazione con generalità, ritenendo *ab origine* un'unità semplice che viene a dispiegarsi attraverso delle figure storiche negative. Il negativo gioca perciò un ruolo apparente, di manifestazione dileguante non appena la si identifica con il suo contrario. Il negativo non segna, si potrebbe dire con Deleuze, una vera differenza positiva, “in sé”, perché, come direbbe Althusser, non presenta alcuna vera rottura.

Ogni volta che il momento negativo si presenta, esso viene superato ristabilendo l'origine semplice; all'individuazione di un'origine semplice non si sottrae neanche Feuerbach attraverso il suo “capovolgimento” hegeliano che condurrebbe a un'astrazione del concreto. Egli sostituirebbe all'Idea semplice hegeliana il concreto dell'esperienza, rimanendo però fedelmente hegeliano nel processo lineare di sviluppo dialettico. Perciò la complessità è solo apparente nella dialettica e la fine è già presupposta nell'inizio. La cognizione della complessità delle pratiche invece comporta la strutturazione delle pratiche, ossia il vederle come tutte connesse vicendevolmente. La struttura di rapporti simbolici definisce delle pratiche in virtù della loro necessaria applicazione al reale; una mera descrizione concettuale è insufficiente per qualificare gli effetti dei concetti sulla realtà, che nel loro sviluppo si amalgamano con pratiche extra-scientifiche (politiche, economiche, sociali...). La “struttura a dominante” che viene a crearsi dall'intreccio di pratiche (termine che, come si è visto, serve per annullare ogni differenza tra concettuale

⁹⁰ Balibar, Etienne, *Zitto ancora Althusser!*, in *Per Althusser*, Manifestolibri: Roma, 1991, pp. 21-22.

⁹¹ Althusser, Louis, *Contraddizione e surdeterminazione*, in *Per Marx*, Mimesis: Milano, 2008.

e non-concettuale) è inseparabile dalle contraddizioni che delinea. Ogni trascendentalismo viene reciso. L'andamento storico della struttura è l'unico orizzonte di sviluppo della scienza, ma si potrebbe dire anche dei LLMs. In questo caso è il modello ricorsivo a rete che ne permette lo sviluppo "orizzontale". Già da tempo l'AI simbolica è caduta in discredito, a causa delle difficoltà di poter codificare in anticipo ogni evenienza ed operazione (problema del frame)⁹².

Si definisca parassitario⁹³ un sistema i cui parametri non intrattengono alcuna relazione con il mondo, ma rinviano a sé stessi in maniera circolare e conchiusa. Un LLM può allora dirsi parassitario nella misura in cui ogni sua attività altro non è tecnicamente che un gioco di vettori in uno spazio multidimensionale. La preoccupazione per i sistemi simbolici, portata avanti da Stevan Harnad nel celebre articolo *The symbol grounding problem* per cui il parassitismo è indice di un *deficit* della macchina, dunque un problema da risolvere, sembra venire smentita a distanza di trent'anni dagli LLMs. la critica mossa alla concezione tradizionale dell'AI simbolica è oggi giocoforza della performatività algoritmica di queste macchine; il fatto che un sistema girasse a vuoto⁹⁴ è per il neostrutturalismo di Weatherby la condizione necessaria affinché si stabilisca un senso, quindi un significato posizionale.

Il parassitismo era considerato una debolezza dell'intelligenza artificiale simbolica nella misura in cui il risultato poteva venire semanticamente interpretato solo a partire da un'istanza esterna al sistema di calcolo. Le operazioni del sistema consistono nella manipolazione di token arbitrari, simboli che oggi possono essere rimandati ai vettori di codifica. La garanzia di stabilità del sistema è dovuta alla sua natura sistematica e compositiva, ma ciò non ne assicura l'ancoraggio.

Le analisi di Harnad, che hanno stabilito e aperto un proficuo problema a lungo dibattuto, quale quello dei *symbol grounding problem*, si risolvono nello stabilire la necessità di un isomorfismo tra macchina e mondo. il problema dell'ancoraggio simbolico allora può essere risolto grazie al ricorso ad un sistema ibrido, simbolico e non-simbolico assieme. L'ancoraggio non può che provenire "dal basso", ovvero sfruttando le risorse

⁹² Si veda Harnad, Stevan, *The symbol grounding problem* in «Physica D: Nonlinear Phenomena», 42 (1-3), 1990, pp. 335-346.

⁹³ *Ibid.* p. 337.

⁹⁴ Si rinvia all'articolo in questione, in particolare nel passaggio sul girotondo cinese-cinese, esempio intuitivo che serve a rappresentare un processo di calcolo basato unicamente su un ordine di parametri che rinvia esclusivamente a sé stesso. La giostra (da notare la somiglianza fra questa giostra e l'esempio cardine dello strutturalismo già presente in Saussure, il gioco degli scacchi) indicherebbe perciò la possibilità per una persona di imparare il cinese a partire solo dalla conoscenza di questa lingua e senza confronti di traduzioni: ogni ideogramma sarebbe rappresentabile tramite ulteriori ideogrammi secondo un rimando infinito, ma consistente rispetto alla lingua mandarina.

dei sistemi a rete nello stabilire isomorfismi tra l'ambiente esterno e il sistema di pesi interno (si veda il caso del Perceptron⁹⁵). Le rappresentazioni non simboliche possono essere “iconiche” nella misura in cui costituiscono delle approssimazioni delle percezioni, oppure delle rappresentazioni “categoriali”, indagando delle invarianze entro lo stesso sistema. Infatti, l'aspetto più importante dei sistemi simbolici è la loro composizionalità, poiché solo manovrando i simbolici grazie alla loro natura sintattica e formale, essi possono venire ricombinati tra loro, permettendo la costruzione di significati inediti senza averla da subito prevista e codificata. Si prenda l'esempio della “zebra”, un simbolo non immesso a priori nel sistema, ma che può venire “compreso” e calcolato tramite il ricorso dei simboli “cavallo” e “strisce”⁹⁶, i quali sono altrettante rappresentazioni categoriali. Il simbolico verrebbe ad essere allora lo strato ultimo, superficiale di un calcolo tra elementi non-simbolici, sfruttando le capacità dei sistemi connessionisti. Ritenere che gli LLMs siano insufficienti a comprendere il mondo, poiché non sarebbero altro che il dispiegarsi di un loro gioco interno di permutazioni di vettori, rappresenta una controargomentazione parziale. La questione cruciale non è quella di un'interiorità di un modello contrapposta a un'esteriorità, dicotomia comunque abolita dalle leggi strutturaliste, ma una quasi completa egemonia di un'istanza (interna, digitale, ovvero macchinico-generativa) sull'altra (esterna, analogica poiché umana), volta al collasso di questa su quella in termini semiotici. La maniera di leggere un testo si uniforma alla modalità svolta dal LLM; la maniera di leggere un codice di leggi appare più conforme a quella del LLM. Detto altrimenti, la molteplicità delle pratiche dei saperi e delle rispettive semiotiche avviene ridotta e uniformata alla loro trascrizione in vettori. Più che di una controargomentazione, sembrerebbe trattarsi di una constatazione, che apre a un campo di ricerca che si può definire come *vector grounding problem*.

Le capacità computazionali dei LLMs non sono niente al di fuori dello scambio che si instaura con gli utenti. Per questo, nonostante l'espressione ricorrerà a volte anche nel presente testo per brevità, è fuorviante dire che queste macchine sono “chiuse in sé stesse”. La lezione althusseriana si impone in questo caso da esempio. Si deve invece affermare che la chiusura delle macchine è dovuta al calcolo probabilistico per comporre il *response*. Il vasto utilizzo che se ne fa è dovuto al fatto che incorporano gran parte della

⁹⁵ Si rimanda alla nostra analisi condotta in parte 2, cap. 2.0.

⁹⁶ Si noti l'analogia consistenza tra l'esempio della zebra e quello invece di Mikolov, usato per spiegare l'architettura dello spazio vettoriale. L'esempio di Mikolov, che verrà ripreso più avanti può essere indicato così: $vector(\text{«King»}) - vector(\text{«Man»}) + vector(\text{«Woman»}) \approx vector(\text{«Queen»})$. La formula indica come sia lo stesso vettore spostamento a mettere in comunicazione “re” e “uomo” e “regina” e “donna”. Cfr. Mikolov, Tomas *et al.*, *Linguistic Regularities in Continuous Space Word Representations*, in «NAACL HLT», 2013, pp. 746-751.

intelligenza collettiva umana. Come afferma Kockelman, è sbagliato dire che detengano un'intenzionalità diretta, ma ne posseggono una “parassitica”, nella misura in cui registrano e memorizzano le espressioni prodotte da individui che intenzionalmente tendono ad oggetti a loro esterni⁹⁷. L'*output* è un risultato numerico dovuto a una serie di moltiplicazioni tra vettori e matrici, che però a loro volta descrivono la maniera con cui l'uomo descrive il mondo.

Secondo un'operazione di *data mining*⁹⁸ senza precedenti nella storia, è necessario mantenere che la “struttura” computazionale dei LLMs non permette di esperire il mondo, dunque in termini tecnici, recide ogni ricorso al referente; al contempo, è grazie all'umana descrizione e assemblaggio di parole che riescono a pronunciarsi *come se* fossero umani; per questo l'interiorità (del processo computazionale) non è che un continuo riflesso delle pratiche esteriori (l'esperienza umana del mondo).

Il legame esterno-interno definisce perciò lo sviluppo immanente dei LLMs. Althusser parla di necessità della contingenza per indicare la circolarità tra sviluppo “necessario” strutturale (la necessità è infatti segno di una spiegazione razionale) e procedere storico “contingente” perché imprevedibile. Il *response*, infatti, tiene conto di diversi fattori:

- Il meccanismo di *attention*;
- Il processo di *fine-tuning*;
- Il *prompt* dell'utente.

I primi due punti sono dovuti alla programmazione delle macchine e seguono un meccanismo “sovra-lineare”, ovvero non sussumibile a partire da dei parametri fissi. I parametri, ovvero i pesi con cui il modello generativo analizza il flusso di informazioni in entrata (*input*) si trasforma retroattivamente come dai primi modelli a rete, ma ora pure automaticamente. Non si tratta quindi di pesi invarianti nei confronti di variabili d'informazione, ma i due aspetti sono vicendevolmente connessi. È necessario, perciò, parlare di *variabilità* del dataset, non di *variazione*. La precisione lessicale non è fine a sé stessa, ma riflette una diversa logica sotterranea, ripresa da Deleuze e Guattari in *Mille piani*⁹⁹. Un sistema non può mai dirsi attualmente in variazione rispetto a delle invarianti

⁹⁷ Kockelman, Paul, *Last Words*, Prickly Paradigmatic Press: Chicago, 2024, p. 28.

⁹⁸ Geoghegan considera l'estrazione di dati come una forma moderna di colonialismo. Cfr. Geoghegan, Bernard D., *Code: from Information Theory to French Theory*, Duke University Press: Durham, 2023, pp. 71-107.

⁹⁹ Deleuze, Gilles, Guattari, Félix, *Postulati della linguistica*, in *Mille piani*, Orthotes: Napoli, 2017.

che lo contraddistinguerebbero; il paradigma è puramente differenziale, poiché un'identità (corroborata da delle invarianti) non può darsi. Mentre la variazione riflette questo andamento oscillante di accidenti che non scalfiscono mai la sostanzialità della struttura, la variabilità deve ritenersi interna e primaria. Gli autori nelle loro considerazioni si aprono al confronto con la linguistica strutturalista, che astrae costanti fonologiche e sintattiche, contrapponendo il modello sociologico di Labov. «Ciò che è primo nel linguaggio, e che conta, non è la costante né la variabile, ma la variazione continua come variazione della variabile stessa»¹⁰⁰. Come le banche dati che sono costantemente in trasformazione rispetto agli *input* e agli *output*, così un sistema si può dire metastabile, poiché non si chiude mai a degli equilibri definitivi. La ricerca del *token* successivo, come si vedrà, non è però contingente, ma segue la prescrizione di alcuni valori di veridicità [*honest*], sicurezza [*harmless*] e utilità [*helpfulness*]¹⁰¹ assieme ad altri valori che sono deliberatamente scelti dall'agenzia che li ha creati (spesso finalizzati alla crescita della capitalizzazione). L'unico aspetto che rimane imprevedibile è il prompt in ingresso, che riflette la scrittura soggettiva dell'utente (agente interpretante umano).

Se l'utente umano formula un prompt, il modello linguistico può generare una risposta appropriata a tale messaggio, la quale può a sua volta costituire una mossa discorsiva che richiede una propria risposta, in modo tale che un modello linguistico possa intrattenere conversazioni simili a quelle umane. È proprio questo che rende il termine "Chat" di ChatGPT.¹⁰²

Nella Chat allora si conservano i due aspetti di Althusser della necessaria contingenza dello sviluppo sia autonomo, ma sempre diretto verso l'esteriorità del mondo.

2.2 SURDETERMINAZIONE E COMPLESSITÀ

Althusser nel suo riferimento alla struttura sta gettando le basi per un pensiero della complessità. Lo scarto tra complessità e semplicità è lo stesso che si palesa tra la dialettica hegeliana e lo "strutturalismo" althusseriano. Non basta l'empirismo à la Feuerbach per uscire dai gangli della dialettica, dimostra nella serie di seminari di lettura de *Il Capitale*

¹⁰⁰ *Ibid.* p. 131.

¹⁰¹ Si tratta più precisamente delle tre H, introdotte da Anthropic nello sviluppo di Claude nel paper di Askell, Amanda, *A General Language Assistant as a Laboratory for Alignment*, <https://arxiv.org/abs/2112.00861>.

¹⁰² Kockelman, Paul, op. cit., p. 20.

che tenne assieme a Etienne Balibar¹⁰³. Il pensiero empirista, infatti, sviluppa il reale secondo due direzioni: dal lato del muto oggetto della conoscenza, privo di qualsiasi appiglio concettuale e dal lato della conoscenza soggettiva. La conoscenza soggettiva arriva a cogliere l'essenza dell'oggetto, perciò l'empirismo riesce a rivendicare la sua posizione epistemologica: la sua conoscenza promana dall'essenza stessa dell'oggetto, che diventa conoscenza *pura*. L'oggetto però è un complesso di tratti essenziali ed inessenziali: tante sue qualità sono ad esempio transeunte, contingenti e accidentali.

L'empirista può accettare la sua conoscenza solo sulla scorta dell'istituzione di un discrimine tra qualità essenziale ed inessenziale degli oggetti. Perciò, la conoscenza dà adito a una purificazione dell'inessenziale dall'oggetto, plasmando perciò una forma essenziale che gli compete, una sua struttura interiore, che resiste al divenire di qualsiasi accidentalità. Come si dichiara una dicotomia tra essenziale e inessenziale, così bisogna ammettere che l'operazione di conoscenza sancisca una distinzione tra interno ed esterno. ogni posizione empirista segue perciò da un paradosso: si dirà univoco il suo oggetto, ma l'operazione tratterà una biforcazione non sintetizzabile definitivamente, aprendo un varco tra oggetto reale e (indipendente da ogni intenzionalità soggettiva) e oggetto della conoscenza (qualificato come essenza dell'oggetto). Solo Marx seppe distinguere¹⁰⁴ accuratamente tra concreto-reale e concreto-di-pensiero. Quest'ultimo cade entro un ordine diverso rispetto al processo di produzione dell'oggetto-reale, ma rifugerebbe parimenti da ogni presa idealistica, definendo una struttura di pratiche, un apparato di pensiero che combina materia reale, pratiche ideologiche, teoria: esattamente i tratti di pensiero che si sono indicati del saggio *Contraddizione e surdeterminazione*.

L'attacco all'empirismo "ingenuo" di Feuerbach riflette però una critica ancor più importante, mossa nei confronti del meccanicismo classico. La linearità del processo di causa ed effetto è insufficiente per descrivere la complessità dei processi strutturali. La causa, seguendo la teoria cartesiana¹⁰⁵, è una potenza maggiore o uguale all'effetto che produce. Il prodotto è riconoscibile perché "esterno" alla causa. La causalità lineare così mantiene una distinzione esterno-interno per individuare la causa e l'effetto; non è un caso che la concezione meccanicistica cartesiana rifletta un nocciolo essenziale dell'organismo, contrapposto alla successione accidentale di eventi che lo scalfisce. Per ovviare a una supposta essenza strutturale, Althusser utilizza il termine "surdeterminazione": si tratta di una determinazione non univoca, quindi non

¹⁰³ I seminari sono raccolti in Althusser, Louis, *et al. Leggere il Capitale*, Mimesis: Milano, 2006.

¹⁰⁴ Cfr. prefazione a Marx, Karl, *Grundrisse*, Einaudi: Torino, 1976.

¹⁰⁵ Descartes, René, *Meditazioni metafisiche*, Laterza: Bari, 1997.

qualificabile solo in riferimento a un'unica causa. Sono diverse le cause che agiscono e l'effetto riscontrato non è tracciabile solo rispetto a un punto di vista.

Esiste un terzo modello, che è propriamente quello di Marx, e che noi abbiamo chiamato causalità strutturale. La struttura non è un'essenza esteriore agli elementi economici, capace di modificarne gli aspetti, le forme e i rapporti, e di agire effettivamente su di essi come una causa assente; la causalità della struttura sui suoi elementi è il modo stesso d'esistenza della struttura.¹⁰⁶

La maggiore dinamicità che acquista la struttura simbolica è dovuta al rapporto molteplice tra la “struttura a dominante” e tutte le condizioni di esistenza, si potrebbe dire tra architettura *transformer* e le tre condizioni individuate prima: *prompt* utente, processo di *fine-tuning* e meccanismo di *attention*. L'effetto è prodotto dall'interazione tra struttura e sue condizioni, quindi a partire da una casualità circolare, concorda infatti Deleuze che

[...] se gli elementi simbolici non hanno designazione estrinseca né significato intrinseco, ma solo un senso di posizione, bisogna porre in principio *che il senso risulta sempre dalla combinazione di elementi che non sono di per sé significanti* [...] il senso è sempre un risultato, un effetto: non solo un effetto come prodotto, ma anche un effetto ottico, un effetto di linguaggio, un effetto di posizione.¹⁰⁷

Il processo è sovrilineare perché le condizioni riflettono la totalità della struttura e viceversa. La struttura non può essere riferita senza tenere conto delle sue contraddizioni. Il termine contraddizione perde la valenza hegeliana, per acquisire quello di condizione specifica del sistema attuale; mentre per Hegel le contraddizioni sono epifenomeni della storia da destinare entro l'ambito contingente ed “inessenziale”, per Althusser sono necessarie componenti che condizionano la struttura. Insomma, c'è immanenza tra condizioni e struttura, perciò la causalità viene detta “circolare” tra struttura e sue condizioni. Un pensiero della complessità deve dimettere il modello meccanicistico, ritenuto insufficiente perché ancora confinato alla distinzione tra essenza stabile e permanente e accidente contingente.

¹⁰⁶ Althusser, Louis, Balibar, Etienne, *Leggere il capitale*, Mimesis: Milano, 2006, p. 187. È il celebre passo in cui Althusser introduce la nozione di “causalità strutturale” come modello alternativo sia al meccanicismo cartesiano sia alla dialettica hegeliana.

¹⁰⁷ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo*, SE: Milano, 2004, p. 19.

CAP. 3: L'IPOTESI DIFFERENZIALE

3.0 DAL SEGNO AL SIMULACRO

Studiare un sistema complesso significa disfarsi dai precetti meccanicisti, insufficienti perché ancora vicini a una propagazione lineare della forza dalla causa all'effetto. La circolarità del rapporto tra causa ed effetto, del problema con le sue soluzioni si può descrivere come una continua trasformazione dei suoi elementi: non esistono elementi che godano di autonomia nel sistema, dunque invarianti; tutti gli elementi anzi, essendo simbolici, non hanno alcun valore determinato.

Come si determina il senso allora? A partire dalla disposizione degli elementi, dalla loro attuale ripartizione. Il senso non può che essere sempre atteso nel presente e non si nasconde "dietro" al simbolo. Il riferimento esclusivamente attuale della distribuzione degli elementi nella struttura è ciò che Saussure definisce come forma. «*Il*

linguaggio ha più il carattere della forma, piuttosto che della sostanza»¹⁰⁸ è ciò che Weatherby qualifica come “ipotesi differenziale”. La combinazione tra elementi simbolici dà vita a un interminabile gioco di rapporti differenziali. Deleuze, quando in *Differenza e ripetizione* parla dei caratteri dell’Idea, si stabilisce su un terreno altro rispetto alla logica binaria del vero (A) e falso ($\neg A$): «si opponga dx a $\neg A$, come il simbolo della differenza a quello della contraddizione, come la differenza in sé alla negatività»¹⁰⁹. Il passaggio di paradigma, infatti, si registra da vero-falso a senso e non-senso, ovvero a partire da una prospettiva in cui ogni elemento è cogliibile solo in rapporto agli altri, a partire dai rapporti (differenziali) che si instaurano e non essenzialmente connesso a un’essenza: dx è del tutto indeterminato rispetto a x , poiché è sempre $\frac{dx}{dy}$, ossia nulla senza il rapporto con dy (con determinazione reciproca si indica l’impossibilità di dissociare dx da dy). È la continua ri-disposizione dei rapporti a costituire il «collante della struttura, i rapporti relativi agli eventi della stessa»¹¹⁰. La struttura viene a coincidere con la sua genesi nella misura in cui la sincronicità è variabile in relazione al tempo.

Il pensiero della complessità si contraddistingue per questa circolarità tra causa ed effetto che annulla ogni prima e dopo. Non c’è un’unica direzione entro cui le differenze possono giocare, ma una «molteplicità di tempi entro la struttura»¹¹¹, ovvero ogni “sottostruttura” è implicata da un proprio tempo; il medesimo punto fa da presupposto al racconto breve di Jorge Luis Borges, *Il giardino dai sentieri che si biforcano*¹¹². C’è una linearità temporale che permetta di dichiarare che una causa prima sia intervenuta su una catena ininterrotta di effetti. L’immanenza tra causa ed effetto si richiama ad un altro aspetto dello strutturalismo che coesiste con l’ipotesi differenziale: l’assenza di un’origine, di un inizio essenziale, di una natura semplice *prima* della complessità. «Ogni cosa non esiste se non in quanto ritorna, copia di un’infinità di copie che non lasciano sussistere originale e neppure origine»¹¹³: è il fondo di differenze positive tra elementi che coesistono, ovvero la struttura che continuamente viene a ricombinarsi secondo il suo “sfondamento”, ovvero «la libertà del fondo non mediata, la scoperta di un fondo dietro ogni altro fondo, il rapporto del senza-fondo col non-fondato,

¹⁰⁸ Saussure, Ferdinand de, *Corso di linguistica generale*, tr. It. a cura di T. De Mauro, Laterza: Roma-Bari, 1967, p.67.

¹⁰⁹ Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, p. 222.

¹¹⁰ Dosse, Françoise, *History of structuralism*, Minnesota University press: Minneapolis, 1997, p. 295.

¹¹¹ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo*, SE: Milano, 2004, p. 34.

¹¹² Borges, Luis, J., *Finzioni*, Adelphi: Milano, 2003.

¹¹³ Deleuze, Gilles, *Differenza e ripetizione*, Raffaello Cortina Editore: Milano, 1997, p. 92.

*la riflessione immediata dell'informale»*¹¹⁴. Ogni momento della struttura corrisponde a una configurazione di una forma, alla selezione di elementi simbolici. Non esiste così un'autenticità a fare da controaltare alla manifestazione surrettizia, non esiste un originale di contro a una copia. La struttura è concepibile solo sulla scorta di una continua trasformazione di simulacri. Così Deleuze ne dà un ritratto

Il simulacro è il sistema in cui il differente si riferisce al differente mediante la stessa differenza. Tali sistemi sono intensivi, poggiano in profondità sulla natura delle quantità intensive, che entrano appunto in comunicazione attraverso le loro differenze.¹¹⁵

Il processo computazionale dei LLMs ne permette l'esibizione di una risposta *come se* fosse umana. Il ricalcolo dà perciò vita a un simulacro, ovvero una copia senza modello, il travestimento di un gioco tra attori che intervengono secondo ruoli: una volta il termine "albero" viene annesso a una risposta riguardante l'ecologia, un'altra volta come strumento di una strategia ideologica di fomentare l'acquisto spasmodico di merci, quale è il Natale. Sono i molteplici rapporti che di volta in volta legano il token "albero" con quelli a cui rimanda il prompt che ne creano la connotazione. Perciò la struttura è una «*molteplicità di elementi differenziali [simulacri], di rapporti differenziali tra elementi, e di singolarità corrispondenti a tali rapporti*»¹¹⁶. La forma del simulacro viene prima di ogni possibile attribuzione di vero e falso; infatti, la dimensione del senso e del non-senso è quella che interessa principalmente gli autori strutturalisti. Se esiste un senso, è sempre un senso globale e di posizione degli elementi: certo, si potranno attribuire valori di verità a certi elementi, ma sempre dopo aver già acceduto alla dimensione strutturale del senso.

È probabile che le nozioni di singolare e di regolare, di rilevante e di ordinario, abbiano, per la stessa filosofia, un'importanza ontologica ed epistemologica molto più grande di quelle del vero e del falso, relative alla rappresentazione; infatti il cosiddetto *senso* dipende dalla distinzione e dalla distribuzione di questi punti che spiccano luminosi nella struttura dell'Idea.¹¹⁷

La struttura viene progressivamente a determinarsi secondo rapporti differenziali e "singolarità". Tenendo conto che la singolarità si definisce come un elemento

¹¹⁴ *Ibid.*

¹¹⁵ *Ibid.*

¹¹⁶ Deleuze, Gilles, *Differenza e ripetizione*, Raffaello Cortina Editore: Milano, 1997, p. 356.

¹¹⁷ *Ibid.*, p. 357.

preindividuale, scisso da ogni somiglianza rispetto a un modello (scambiata erroneamente come causa, ma in realtà effetto di senso) è essa l'oggetto-simulacro che risulta dalla continua ricombinazione. I caratteri impersonali, preindividuali, di differenziazione che Deleuze ascrive alla singolarità possono venire attribuiti al segno linguistico o ai suoi costituenti significanti primi, ovvero i fonemi. Il fonema, come ad esempio “*sclerato*” o “*sperato*” è la differenza che si distribuisce tra due parole di significato diverso. Esso si incarna in certe sillabe o parole, ma non è contemplabile al di fuori delle relazioni che lo legano agli altri fonemi (*scler-spe*). Ogni struttura si contraddistingue perciò per una serie di rapporti differenziali, dovuti alla determinazione reciproca tra gli elementi e una serie di punti singolari, i cui valori (significato, verità/falsità) dipendono dalla determinazione completa dei rapporti differenziali.

3.1 MACCHINE INTELLIGENTI?

A partire da Alan Turing si è iniziato a riflettere sulla natura “digitale” dell'intelligenza, ossia sulla sua capacità di essere trascrivibile secondo un modello consistente e di essere un sistema aperto, “*open-ended*”¹¹⁸, dunque ricorsivamente operativo, capace di reiterare le proprie operazioni per svolgere dei calcoli non previsti nella sua inizializzazione. Dagli studi di Turing si conferma che la macchina non solo debba riprodurre una correttezza disumana, ossia capace di risolvere problemi preclusi per una questione di tempo agli umani, ma anche deve sbagliare come farebbe un uomo, producendo degli errori che ricalchino la maniera collettiva, interindividuale di intrattenere discorsi umani. La stupidità si potrebbe pensare come un livello insito nel ragionamento umano e non come un deficit¹¹⁹, quindi qualcosa da tollerare nel ragionamento delle macchine. Le macchine non dovrebbero solo ragionare seguendo la stessa maniera dell'uomo, ma anche sbagliare nella medesima modalità.

Il funzionamento a rete secondo “deep learning” consente di andare oltre la mera riproducibilità tecnica, seriale di codici secondo dettami informatici e di programmazione; la ripetizione meccanica e fisica del medesimo processo era ciò che fino a prima dell'invenzione del deep learning si credeva essere l'automazione. Il

¹¹⁸ La ricorsività è data dall'apertura di queste macchine e dalla loro consistenza: convertono dei calcoli logicamente coerenti, quindi sono consistenti, ma non chiuse, quindi complete. L'incompletezza è il carattere “*open-ended*” della macchina.

¹¹⁹ Gilles Deleuze in *Differenza e ripetizione* insiste su questo punto, proponendo la stupidità come un fattore trascendentale del pensiero umano. Cfr. Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, pp. 196-199.

processo semiotico secondo reti neurali, infatti, è una continua evoluzione e non la reiterazione del medesimo. La nozione di simulacro secondo Baudrillard, che ne definisce già i principali caratteri nel 1976 con *Lo scambio simbolico e la morte*, che rintraccia un contenuto privo di una forma essenziale, originaria, il propagarsi di un fenomeno indefinitivamente secondo un codice di convenzioni, dunque rilevando il fondamentale presupposto sociale, può fare eco a quella che già predispone Deleuze qualche anno prima.

In breve, la ripetizione è simbolica nella sua essenza, il simbolo, il simulacro, è la lettera della ripetizione stessa. Mediante il travestimento e l'ordine del simbolo, la differenza è compresa nella ripetizione. Questo spiega perché le varianti non vengono dal di fuori.¹²⁰

Una nozione prolifica come quella del simulacro serve per evitare di rientrare dentro una metafisica che ancora gioca sulla dicotomia tra apparenza ed essenza, da cui possano proliferare umanismi nostalgici di sorta, che prevedano il riconoscimento di una non ben definita figura umana. Al posto dell'opposizione tra essenza ed apparenza, seguendo il ragionamento di Deleuze e tenendo conto della coppia regolare-singolare, ogni simulacro è tale perché promana da una modulazione di differenze, ovvero la commutazione degli elementi simbolici che servono per costruirli, ossia tramite le opposizioni binarie che servono a costruirli:

Dalla più piccola unità disgiuntiva (la particella domanda/risposta) fino al livello macroscopico dei grandi sistemi d'alternanza che governano l'economia, la politica, la coesistenza mondiale, la matrice non cambia: è sempre 0/1, la scansione binaria che s'afferma come la forma metastabile, o omeostatica dei sistemi attuali. Essa è il nocciolo dei processi di simulazione che ci dominano. Essa può organizzarsi in un gioco di variazioni instabili, dalla polivalenza alla tautologia, senza che sia rimessa in causa la forma strategica del dipolo: è la forma divina della simulazione.¹²¹

È proprio il reiterare del modello e non del processo, come ad esempio viene garantito secondo un processo di *backpropagation* che ne consente la differenziazione, quindi la reiscrizione sempre nuova. Il modello cede il posto a una continua riproduzione dei mezzi tecnici [*modulazione*], di cui non si può parlare come di una semplice copiatura; il *learn to learn* si traduce in un'automazione dell'automazione, basata sulla generazione di nuovi

¹²⁰ *Ibid.*, p. 28.

¹²¹ Baudrillard, Jean, *Lo scambio simbolico e la morte*, Feltrinelli: Milano, 2022, p. 82.

pattern a partire dall'analisi dei dati. Ciò che conterebbe non è alcuna adeguazione alla cosa in sé, alcuna rappresentazione della cosa attraverso delle forme a priori, ma una trascrizione della cosa secondo un'altra *serie* di segnali. L'intero processo di intelligenza risulterebbe in ultima istanza un processo a spirale aperta, un continuo interscambio semiotico tra un sistema e l'ambiente che lo circonda.

Quello che McCulloch e Pitts descrivevano nel loro celebre articolo del 1943 come un'"eterarchia"¹²², ossia un sistema che si ricostituisce secondo delle relazioni mobili, flessibili e abbandona la fissità e centralità del modello "arborescente", tipico dei GOFAI, sigla con cui il filosofo John Haugeland descrive il modello simbolico AI. L'approccio simbolico si basa sull'osservazione che i computer sono operatori simbolici, che i simboli sono segni arbitrari che rappresentano valori fissi e che tra loro intercorre una struttura di relazioni che i programmi informatici possono potenzialmente sfruttare: il programma, quindi, reitera delle operazioni stabili secondo dei segni codificati in anticipo in maniera fissa. Anziché esserci un *set* di simboli che fanno eco a un fondazionalismo secondo regole fisse, leggi immutabili che a priori governano, McCulloch e Pitts enfatizzano che è la finitudine della conoscenza che formalmente accomuna macchine e computer, ovvero è la sua costituzione fondamentalmente aperta che non consente di distinguere nettamente l'interiorità di una configurazione fissa e stabile e l'esteriorità contingente e precaria. Almeno, questa è la qualificazione digitale-analogica che contraddistingue la tendenza AI a partire dagli anni '80, chiamata "connessionismo". Prima dell'avvento dei Big Data, conseguenza della commercializzazione di Internet, sistemi (Mark I Perceptron, NumaRete) si definiscono «connessionisti» in senso letterale: non solo sono collegati all'interno del "cervello" computazionale della macchina, ma anche a uno spazio esterno ai suoi circuiti elettronici, combinando forme fisiche catturate otticamente con l'elaborazione simbolica tramite parametri. Oggi invece i LLMs presentano un modello connettivista tenendo in considerazione che non c'è alcuna simbolizzazione fissa e stabile, ma che ogni parametro è il risultato della connessione con gli altri, ma non c'è alcuna distinzione analogico-digitale al pari di qualche decennio fa.

I modelli di apprendimento automatico [*machine learning*] di oggi operano in un ambiente digitale standardizzato e apprendono dalla sua struttura. Piuttosto che essere analogici e digitali, sono innanzitutto digitali e, in secondo luogo, non lineari e statistici.

¹²² Il riferimento si trova in Weatherby, Leif, *Language Machines*, Duke University Press: Minneapolis, 2025, p. 120.

Le loro "percezioni" risiedono all'interno dei loro "cervelli".¹²³

I modelli a rete neurale sono completamente costruiti sul digitale, abbandonando il lato analogico. Tutto ciò con cui hanno a che fare sono tokens e connessioni statistiche tra loro; non applicano delle regole a una materia, trasformandola, mettendo una serie di immagini di oggetti fisici in correlazione con una serie di codici. Tutto qui invece si svolge entro il *codice* di dati digitali. Un modello affine a quello dell'intelligenza consiste nell'emulare il meccanismo connettivo del cervello tra parole e cose, tra regno simbolico e non simbolico: l'intelligenza, comunemente intesa per i neurobiologi, è una connessione ibrida di elementi di natura completamente differente. Con gli LLMs invece le operazioni rimangono entro un'unica struttura: montaggio e decodificazione di un messaggio valgono di più del significato interno al messaggio; in altre parole, è il medium digitale che *fa* il messaggio, che prende il sopravvento sul significato del messaggio: caratteristica della terza generazione di simulacri per Baudrillard, che parimenti contraddistingue queste tipologie di macchine. Non c'è alcun modello di intelligenza a partire da un'inizializzazione originaria. Piuttosto di intelligenza connettiva, si tratta di un processo semiotico: *«Dobbiamo renderci conto che l'intelligenza stessa presenta un aspetto formale-semiotico e comprendere che i sistemi a rete non stanno copiando un'intelligenza preesistente — e certamente non conosciuta né compresa»*¹²⁴. Quindi, privi di qualsiasi "deduzione trascendentale" che riconduca la molteplicità simbolica delle macchine a un'unità fondativa di un "Io penso macchinico", McCulloch e Pitt forniscono un modello di intelligenza a rete più vicino all'idea di simulacro di Baudrillard, che rompe con il referente esterno.

¹²³ *Ibid.*, p. 89.

¹²⁴ *Ibid.*, p. 92.

PARTE 2: LA STRUTTURA DEI MODELLI LINGUISTICI

CAP. 1: “WORD EMBEDDING”

1.0 DISTRIBUZIONALISMO HUMEANO

Il recente avvento dei LLMs porta alla luce un’architettura software che mai prima d’ora si era vista. L’aspetto linguistico è quello che qui interessa maggiormente, ovvero la svolta che questi strumenti sembrano aver dato nelle teorie del linguaggio. Questo capitolo vorrebbe quindi approfondire come le teorie strutturaliste e della linguistica riescano a spiegare la dinamica computazionale che governa queste macchine.

La macchina in questo caso metterebbe in mostra una modalità di programmazione altra rispetto al modello dell’AI simbolica, fondato su una logica *top-down* e in cui manualmente ogni simbolo deve essere connesso alle sue operazioni digitali. Si cercherebbe di prevedere ogni mossa digitale della macchina a partire da una costellazione di simboli che riescano a coprire ogni evenienza, inevitabilmente riducendone le capacità rispetto a un modello adattivo all’ambiente circostante, in costante crescita con esso. L’apporto cibernetico, che molto ha condizionato lo strutturalismo¹²⁵, sta qui: *concepire la macchina e l’uomo come due organismi capaci non solo di riprodursi, ma di auto-regolarsi e svilupparsi a seconda dell’ambiente in cui sono calati*. I LLMs invece possiamo definirli ricorsivi primariamente, perché garantiscono una continua e rinnovata iterazione con l’ambiente, quindi con l’interfaccia umano, sviluppando i propri pesi e misure proprio come un organismo rispetto all’ambiente esterno. Fuor di metafora, tramite i dati che vengono continuamente ridestati e aggiornati, i LLMs in realtà non vivono in simbiosi con l’ambiente; tutt’al più la simbiosi si potrebbe definire socio-culturale. Ciò che manca loro è l’immediata esperienzialità del mondo, che ancora i fratelli Dreyfus avevano messo in evidenza nel loro saggio¹²⁶, comparando il

¹²⁵ Lafontaine, Céline, *The cybernetic matrix of French Theory*, in «Theory, Culture and Society», 24, 2007, pp. 27-46.

¹²⁶ Dreyfus, Hubert, Dreyfus, Stuart, *Making a Mind versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint*, in «Daedalus», 1, 1988.

regime della macchina con il concetto heideggeriano di “essere-nel-mondo”. Ciò che avverrebbe tramite i lucidi *response*, è un riferimento non al mondo, ma al “mondo in parole”, letteralmente un *Umwelt* linguistico. È pur vero che l’uomo che si interfaccia alla macchina permette alla macchina di registrare delle combinazioni tra parole, accostamenti verbali e semantici che riflettono quelli di un individuo che vive nel mondo. Sotto questo profilo, la macchina stringe una relazione *parassitaria*¹²⁷ con l’uomo, costruendo il proprio data set sulla scorta di qualcosa che le è precluso. Ciò che coglie Serres, è l’apporto innovativo e creativo del rumore, che fa sì che un parassita, possa a sua volta diventare fondatore di un sistema ulteriore. La logica che contraddistingue l’operatività del parassita è quella della *presa*: dati due elementi e una loro relazione, il parassita è ciò che, mediando tra i due, trae energia sufficiente per riprodursi, senza che, viceversa, i due possano ricavarne energia. In altri termini, il rapporto è unidimensionale: il parassita trae giovamento da una struttura, riconfigurandola. Il parassita allora è principio di nuove strutture¹²⁸. Il termine francese *parasite*, infatti, ha tre accezioni: il parassita biologico, quello sociale (che partecipa ai banchetti ricambiando il cibo con una storia) e il *bruit parasite* (il rumore di fondo, l’interferenza). È interessante la proposta di Serres, perché permette di vedere l’evoluzione di un rapporto uomo-macchina come di una relazione non tra entità monolitiche, ma processi in continua trasformazione. ogni posizionamento sistematico non è che inevitabilmente transeunte e mai definitivo. Tra “caos” ovvero «non-sapere» e sistema come «sapere» non ci sono che scambiatori di energie, informazioni, tempi, segnali e rumori. Gli LLMs si possono allora considerare parassiti nella misura in cui la loro funzionalità strumentale a sua volta non fa che tradurre il nostro sapere in una serie di codifiche metastabilmente succedentesi a cui l’uomo finisce per assoggettarsi.

Si diceva che gli LLMs non sono immediatamente proiettati sul mondo, ma entro un particolare *Umwelt* linguistico: a questo livello interpretativo si capisce che la lezione dello strutturalismo è forte, nella misura in cui si libera del referente. Allora si potrebbe dire che l’oggetto del processo semiotico macchinico non è un corpo esterno, il mondo

¹²⁷ Serres, Michel, *Il parassita*, cap. “macchine e motori”, Mimesis: Milano, 2023.

¹²⁸ «Credevo che gli scambiatori fossero intermediari, che l’interferenza fosse nella frangia, che il traduttore si piazzasse tra le istanze, che il ponte riunisse le due rive, che il percorso andasse dall’origine alla fine. Non ci sono istanze. O piuttosto le istanze, sistemi, rive ecc. sono a loro volta analizzabili in scambiatori, percorsi, traduzioni e così di seguito. Non vi sono istanze o sistemi, ma solo scatole nere. Quando non comprendiamo, quando rimandiamo la nostra scienza a più tardi, quando la cosa è troppo complessa per i mezzi del giorno, quando ci piazziamo in una scatola nera temporanea, giudichiamo a priori che si tratta di un sistema [...] non vi sono sistemi, istanze, sostanze se non delle nostre ignoranze. Il sistema è il non-sapere. L’altro lato del non-sapere. Il non-sapere ha un lato caos e un lato sistema. Il sapere è un ponte tra queste due rime» cfr. Serres, Michel, op. cit., p. 101.

fisico, ma la parola “successiva” al prompt scritto. Il segno del prompt, infatti, viene letto rimandando a uno o più oggetti (l’*output* oggetto è infatti una composizione di molteplici oggetti tokens); la barra [*slash*] si presenta quindi non tra una parola e l’oggetto, tra il fenomeno e la cosa stessa, ma tra una parola e l’altra, tra la parte iniziale di un testo e il suo proseguimento, che Kockelman estende a un passaggio da «passato a futuro»¹²⁹.

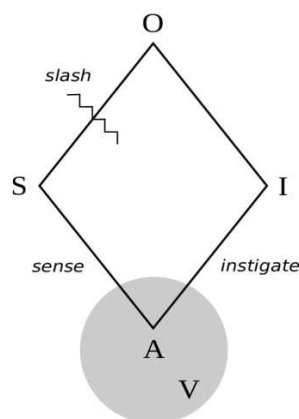


Figura 4 Processo semiotico macchinico¹³⁰

Storicamente il passaggio che si è appena delineato può essere rappresentato dal discorso che Frank Rosenblatt tenne in occasione del primo simposio internazionale di AI nel novembre 1958, dal titolo “La meccanizzazione dei processi di pensiero”¹³¹. Rosenblatt, psicologo e dedito studioso delle reti neurali, si appoggia alla distinzione tra linguaggio *primario* del cervello e *secondario* della matematica, affermando che il ragionamento per riprodurre le facoltà neurali per il riconoscimento di parole o immagini doveva ricalcare un ragionamento di tipo ricorsivo e induttivo piuttosto che di tipo deduttivo, chiuso e completo.

Già von Neumann, rinomato matematico, si era soffermato sulla differenza tra linguaggio primario del cervello, atto a processare una trasmissione continua e reiterante di dati, e il linguaggio secondario (metalinguaggio) logico-matematico che gli scienziati costruiscono per descriverlo. La coincidenza tra i due è impossibile, viste le finalità del linguaggio matematico: esso è chiuso in sé e riposa su una serie di regole logiche inamovibili. Per questo l’AI simbolica fu ritenuta inadeguata e Rosenblatt propose un modello matematico più “intuitivo” e statistico piuttosto che analiticamente perfetto. La

¹²⁹ Kockelman, Paul, *Last words*, Pickly Paradigmatic Press: Chicago, 2024, p. 17.

¹³⁰ Immagine da *Ibid.*

¹³¹ Ci si riferisce al celebre convegno internazionale “*The Mechanization of Thought Processes*”, presso il *National Physical Laboratory* di Teddington.

successiva invenzione del *Perceptron* (1958), poi divenuto riferimento per i modelli a rete come *LeNet* (1989) e *AlexNet* (2012), è segnata principalmente da un'intuizione della matematica statistica applicata ai metodi di rilevamento dei dati a multidimensioni della psicologia dell'epoca, piuttosto che ad una svolta in termini di programmazione informatica. Così la differenza tra i due linguaggi di von Neumann, poi ripresa da Rosenblatt in occasione del simposio si traduce «non solo come una digitalizzazione del piano dell'immagine in una matrice numerica a due dimensioni, ma nella sua vettorizzazione in una matrice a multidimensioni»¹³².

Per questo l'ipotesi distribuzionalista si accorda bene con il processo semiotico alla base dei LLMs. L'ipotesi ritiene che il significato di un termine venga a coincidere con le sue relazioni, probabilisticamente connotate. Christopher Manning, una delle voci più autorevoli del dibattito linguistico, ha descritto l'arrivo dei LLMs come «una delle scoperte che hanno apportato maggior conoscenza sulla struttura del linguaggio»¹³³; affermazione importante e tendenziosa per rivendicare la preminenza del suo ambito. Secondo un approccio probabilistico, il linguaggio non è che l'accostamento di espressioni tipiche, veicolate con una certa frequenza che può venire descritta in termini probabilistici, una prospettiva externalista già dibattuta da Leonard Bloomfield; secondo una prospettiva anti-concettualista, infatti, l'enunciazione non è un contenuto presente prima in chi enuncia e poi ricevuto nella mente del destinatario, l'enunciazione segue delle leggi di frequenza linguistica che codificano in partenza la trasmissione. Il riconoscimento del pattern linguistico consegue allora dal rilevamento delle sue posizioni entro il testo, dunque rispetto alla sua distribuzione, per dirla con Zellig Harris¹³⁴. La linguistica compie un'operazione meramente descrittoria della distribuzione statistica dei *token*; l'enunciato diventa perciò un compromesso tra vincoli e libertà dell'enunciante, quindi tra informazione e ridondanza per dirla con Shannon. Nel 2017 infatti viene pubblicato un documento intitolato “Attention is all you need” [tutto ciò di cui si ha bisogno è l'attenzione] in cui i ricercatori di GoogleBrain rivoluzionano il mondo dei modelli a rete, passando dai modelli RNN a reti neurali ricorrenti a quello più performante di auto-attenzione. Ora, Perceptron, RNN e LLMs sono modelli che funzionano diversamente tra loro, ma vengono qui confrontati per mostrare che rispondono tutti e tre a una logica probabilistica, segnando un definitivo distacco dalle logiche deduttive in informatica e dai modelli referenziali del linguaggio. Ciò che tutti e tre presuppongono è

¹³² Pasquinelli, Matteo, *The Eye of the master*, Verso: Londra, 2023, p. 223.

¹³³ Manning, Christopher, *Human language understanding and reasoning*, in «Daedalus», 2022, 151 (2), p. 131.

¹³⁴ Harris, Zellig, *Structural linguistics*, University of Chicago Press: Chicago, 1960, p. 5.

perciò una primarietà del NLP, ossia del linguaggio naturale (il linguaggio primario per Rosenblatt) su quello simbolico, matematico. Anche se si dovesse adoperare una serie di simboli logici, in ultima istanza si dovrebbe ricorrere al linguaggio naturale per spiegarli. Il rigoroso empirismo, che Weatherby definisce “humeano”¹³⁵, abbraccia allora l’idea di un indefinito progresso nella modellizzazione dei dati, affinando il rapporto tra matematica e linguaggio senza spiegarne in ultima istanza una ragione “trascendentale”. Più che di un riduzionismo alle probabilità si tratta piuttosto di un’evidenza epistemologica per cui il linguaggio naturale, con le sue stocastiche costruzioni, finisce per essere l’orizzonte ultimo di riferimento.

L’ipotesi distribuzionalista sposa bene la serie di tecniche di calcolo che ingegneri e ricercatori hanno implementato secondo il modello bayesiano. Senza approfondire il ruolo del programma bayesiano nella storia del NLP, basti sapere che esso non fu mai solo una teoria, rappresentando una svolta importante dalla programmazione computazionale simbolica alla schermatura tramite dati. Thomas Bayes, filosofo e matematico illuminista, progettò un metodo di analisi dell’andamento statistico per mezzo della ricorrenza di forme (pattern) passati, prevedendo di calcolare le occorrenze sulla base di aggiornate previsioni su quando e come possano avverarsi. Il procedimento bayesiano secondo Weatherby si è affinato con l’avvento degli spazi vettoriali:

Le regolarità relative delle parole in un corpus ci forniscono solo alcune indicazioni, ma la matematica dei vettori ha offerto una chiave per la modellizzazione che tende a costruire un senso del significato distribuendo quelli che sembrano essere gli «aspetti» del significato attraverso uno spazio vettoriale continuo.¹³⁶

È la traducibilità potenzialmente a infinite dimensioni che offre il vettore come mezzo per implementare la logica distributivo-probabilistica. Non si tratta di un’analisi lineare, ma che prende a riferimento il significato come posizionamento entro uno spazio multidimensionale. Il significato di una parola deriva dall’uso che se ne fa, quindi linguisticamente a partire dal suo contesto lessicale o di situazione pragmatica.

¹³⁵ Weatherby, Leif, *Language Machines*, Minnesota University Press: Durham, 2025, p. 54.

¹³⁶ *Ibid.*, p. 59

1.1 VETTORIALIZZAZIONE DEL PENSIERO

È necessario chiarire meglio la natura del vettore. Ogni vettore matematicamente si può scrivere come la radice della somma dei quadrati delle sue componenti. Sul piano cartesiano può essere visto quindi come composto da due componenti vettoriali, corrispondenti alle due dimensioni del piano. Nello spazio multidimensionale le dimensioni accrescono vertiginosamente, facendo della parola che nomina il vettore una certa posizione nello spazio. Virtualmente, un vettore (quindi un *token*, perciò anche una parola) sarà vicino ad altri che condividono, o hanno una stretta correlazione con, le sue componenti. Lo spazio vettoriale prefigura perciò, tramite i rimandi e i rapporti statistici, un vocabolario semantico. Nei precedenti modelli a rete si trattava di studiare le occorrenze sintattiche sulla base di quelle paradigmatiche, in altre parole studiare dove un significato si situava secondo tutti i contesti in cui appariva. La “semantica vettoriale” invece non ricerca delle regolarità formali tra variabili, ma relazioni significative tra parole, ossia operazioni tra vettori e matrici.

I termini sono assi, le parole sono vettori, le cui componenti sulle varie dimensioni sono determinate dal numero di co-occorrenze nella matrice in cui sono situate. La somiglianza tra i vettori ha quindi un chiaro equivalente grafico: la vicinanza nello spazio multidimensionale corrispondente alla matrice di collocazione.¹³⁷

La somiglianza (iconicità, come direbbe Peirce) si registra non passando tramite un significato del simbolo, ma tramite il suo posizionamento nella rete, quindi nella struttura. La semantica del vettore è quindi una questione di posizionamento, prima ancora che di valore di verità.

Il vettore è uno strumento semantico flessibile, applicabile anche nel caso della “document-term matrix” (DTM), ossia una matrice di ricerca entro vasti depositi di dati. Il vettore può infatti venire usato come mezzo per scovare la “rilevanza” di un dato che si vuole cercare. Ad esempio, digitando “Gilles Deleuze” la matrice potrà restituire non solo i documenti in cui effettivamente compare le parole “Gilles Deleuze”, ma anche tutti i documenti rilevanti per quella serie di parole (manuali di filosofia francese contemporanea, opere di filosofi o scienziati a cui si è ispirato). Allora il valore di rilevanza dato dalle operazioni matriciali viene influenzato dalla logica bayesiana, quindi

¹³⁷ Schütze, Hinrich, *Dimension of meaning*, Proceedings of Supercomputing '92, Center for the Study of Language and Information, Stanford, 1992.

gli argomenti, i temi di ricerca passati contribuiranno a definire delle aree di interesse pertinenti alla ricerca in corso su “Gilles Deleuze”. È in fondo il medesimo procedimento che interviene con la ricerca su Google che «*concerne la probabilità che un documento (scritto con un numero di termini pesati) è rilevante per una ricerca [query] (ricercato da un numero di termini che potrebbe a sua volta venire pesato)*»¹³⁸.

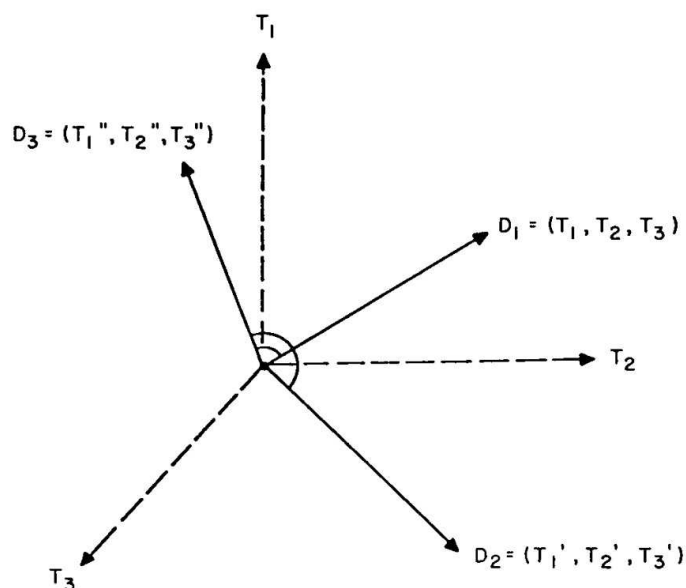


Figura 5 Rappresentazione di tre vettori in uno spazio tridimensionale¹³⁹

Il meccanismo alla base del software di ricerca venne sviluppato in un articolo di Salton, Wong e Yang, testo cardine dell'attuale *information retrieval*. Data una collezione di documenti, qual è la configurazione ottimale dello spazio in cui rappresentarli affinché sia possibile recuperarli in modo efficace? Lo spazio vettoriale permette ad ogni documento di essere trascritto sotto forma di vettore in uno spazio multidimensionale, in cui ogni dimensione corrisponde a un termine indice. Ogni documento appare come somma vettoriale di tutti i termini indici, a loro volta componenti che misurano la frequenza, quindi la media pesata delle loro apparizioni. La rilevanza di un termine però non è semplicemente la sua maggiore frequenza. Come si evince dalla parte sperimentale dell'articolo¹⁴⁰, basata su una ricerca estesa su alcune collezioni di documenti importanti come TIME per l'economia, la rilevanza non è né sinonimo di rarità né di frequente. la frequenza ottimale è intermedia: né troppo rara da non permettere una ricerca su un vasto aggregato di documenti, né così comune da venire avvertita quasi indistintamente. Anzi,

¹³⁸ Rieder, Bernhard, *Engines of order*, Amsterdam University Press: Amsterdam, 2020, p. 238.

¹³⁹ Immagine da Salton, Gerard *et al.*, *A Vector Space Model for Automatic Indexing*, in «Communication of the ACM», vol. 18, n. 11, 1975, p. 615.

¹⁴⁰ Salton, Gerard *et al.*, *op. cit.*, pp. 613-620.

lo schema di pesatura ideale viene registrato secondo la formula $f_i^k \cdot IDF^k$, dove IDF è la frequenza inversa del documento e f la frequenza di ogni indice i per un termine k .

Un sistema di ponderazione dei termini proporzionale a $f_i^k \cdot IDF^k$ assegnerà il peso maggiore a quei termini che ricorrono con elevata frequenza nei singoli documenti, ma che allo stesso tempo sono relativamente rari nell'insieme della raccolta.¹⁴¹

In pratica i termini più rilevanti riescono a scremare meglio lo spazio, a diradarlo, mentre quelli cattivi, una volta inseriti lo addensano. Il *discrimination value model* è uno strumento per scegliere i termini-indice migliori, più rilevanti. «Un termine-indice “buono”, ovvero un termine con un elevato discrimination value, riduce la somiglianza tra i documenti quando viene assegnato alla collezione. Il contrario vale per un termine di indice “cattivo”, con un basso discrimination value»¹⁴². Ciò però non basta per segnalare la rilevanza di un eventuale termine. Bisogna estendere anche a termini composti, magari una sequenza di termini incisivi oppure ai sinonimi dei termini-indici che si dichiarano importanti. A tal fine nascono due funzioni specifiche: il *phrase* e *thesaurus*, che, secondo le ricerche condotte producono dei miglioramenti di circa il 20% nell'individuazione di termini chiave. Il *phrase indexing* è una funzione che permette di associare in frasi composte diversi termini ad altissima frequenza, la cui frequenza combinata è minore; la funzione *thesaurus* invece aggrega termini a bassa frequenza in classi semantiche affini, in maniera da inquadrare le sinonimie. Tutta la tradizione successiva della *Latent Semantic Indexing* o dei *word embedding* segue da queste premesse del 1975¹⁴³. La portata epistemologica della proposta, infatti, sostituendo il modello logico-simbolico che cercava di stendere delle relazioni biunivoche tra termini, lascia il posto alla *semantica distribuzionale*.

Nel 1992 sarà Hinrich Schütze, ricercatore a Stanford, a inaugurare una più elaborata modalità di ricerca, a partire dalla geometria del modello di Salton. Per i principi della semantica distribuzionale, infatti, un significato è legato direttamente al contesto di una parola. È questo principio semantico che permette a un LLM di costruire delle frasi di senso compiuto. Il principio della semantica distributiva è così astratto che non

¹⁴¹ *Ibid.* p. 615.

¹⁴² *Ibid.* p. 618.

¹⁴³ Un esempio tra tutti: le strategie di selezione *top-k*, come si vedrà in seguito, che restringono la ricerca dei termini alle k occorrenze di valore più elevato, prendono esempio proprio dal SVD (*Single Value Decomposition*). I valori singolari più elevati infatti catturano maggiore struttura semantica della matrice di riferimento. La matrice, quindi, viene approssimata e ridotta, ad esempio da migliaia di dimensioni a centinaia di dimensioni dense, catturando le dimensioni senza rumore.

abbisogna di alcun dizionario o enciclopedia. Lo spazio viene sondato da delle “dimensioni di significato”¹⁴⁴, ossia delle connessioni latenti che permettono di registrare delle sequenze di parole rilevanti, in maniera così da restituire il significato di una parola tramite il suo contesto, attraverso degli algoritmi come SVD (*Singular Value Decomposition*). Ad esempio, per una parola come “banco”, che può voler dire sia un tavolo che una somma di denaro, il significato può venire rintracciato tramite le occorrenze linguistiche che lo accompagnano prima o dopo, come potrebbero essere “scuola”, “alunno”, “studente”, “professore” oppure “banca”, “prestito”. Ora, al di là dei dettagli tecnici, quello che qui interessa è mostrare come le odierne tecnologie LLM utilizzano i medesimi principi studiati dall’*information retrieval* già nel 1975 da Salton e poi affinati nel 1992 da Schütze per disambiguare il significato prima di un documento e poi, per estensione, delle stesse parole che vi accorrono.

La tesi di Rieder consiste nel definire questi modelli ricorsivi bayesiani come una novità rispetto alle filosofie del linguaggio, che verrebbero incorporate nelle modalità mediatiche con cui le macchine digitali funzionano. In altre parole, il ferreo empirismo che non rileva alcun carattere di tipo concettuale nel rapporto parole-significato, consente di sviluppare i modelli linguistici computazionali non sulla base di una precisa teoria linguistica, ma su una serie aperta di modelli tecnici. L’aspetto del linguaggio da prediligere è quello interpretativo, nella misura in cui, data una stringa di *tokens*, bisogna capire a cosa di accordi meglio, quali connessioni appaiono più rilevanti. Non sarà però una catena di Markov a risultare utile ma dei modelli che si distinguono per la loro capacità di apprendere tramite algoritmi di *training/learning*. Questi algoritmi ricevono in ingresso non solo un insieme di punti dati, ma anche una gamma di modelli possibili. Il processo di apprendimento consiste nella selezione del modello più adatto tra quelli disponibili, in base alla capacità di adattarsi alla struttura dei dati forniti. Infatti, Weatherby afferma che «*il significato non viene quindi colto da un “modello” nel senso comune del termine, bensì dall’adattamento dei valori del modello alla struttura dei dati in ingresso. La questione è se un modello di questo tipo abbia poi un qualche potere esplicativo*»¹⁴⁵.

Una delle principali differenze rispetto agli approcci tradizionali, come quello avanzato da Shannon, è che questo tipo di modello non si limita a prevedere una singola parola; piuttosto, si occupa di costruire una funzione predittiva che sarà poi in grado di

¹⁴⁴ Da cui il titolo del celebre articolo a cui ci sta riferendo. Cfr. Schütze, Hinrich, *Dimension of meaning*, Proceedings of Supercomputing ‘92, Center for the Study of Language and Information, Stanford, 1992.

¹⁴⁵ Weatherby, Leif, op. cit., p. 62.

prevedere quella parola. In tal modo, il significato emerge dall'adattamento continuo dei valori del modello alle caratteristiche specifiche dei dati di ingresso. Di conseguenza, la forza di un modello addestrato risiede nella sua flessibilità: non è vincolato a una struttura fissa, ma si trasforma e si perfeziona in base ai dati che riceve. Questo processo di adattamento permette di cogliere il significato non come un'entità statica, ma come una relazione dinamica tra la funzione appresa e la *struttura* dei dati. In tal modo, il modello si evolve costantemente, riflettendo la complessità e la varietà delle informazioni che gli vengono fornite. Il passaggio all'architettura *transformer* però cela maggiori potenzialità. Ad esempio, come fa notare Manning, la risoluzione dei test di Winograd¹⁴⁶ fa apparire la macchina avere “un'intelligenza” adattiva alle leggi del linguaggio naturale. Più che di intelligenza si tratta della reiterazione dei modelli complessi del linguaggio: il meccanismo di *attention* crea infatti ulteriori livelli di profondità e nuove connessioni mappabili tra *tokens*.

Il significato non è una questione di tutto o niente. In molte situazioni ci ritroviamo parzialmente a comprendere il significato di una forma linguistica. Ritengo perciò che il significato emerga dalla comprensione della *rete di connessioni tra una forma linguistica e altre cose*, che siano oggetti del mondo o altre forme linguistiche. Se possediamo una densa rete di connessioni, allora avremo una comprensione adeguata di quella forma linguistica.¹⁴⁷

Non si tratta di un'intelligenza legata alla cognizione, alla chiara presa in considerazione di oggetti secondo parole, ma di un senso che emerge dal basso, connettendo in maniera mai definitiva gli elementi della struttura. Quindi, «*la prospettiva statistica, una volta fondata sulla distribuzione di parole, poi resa attraverso la sua vettorizzazione, è diventata empiricamente modellabile, nel senso specifico del machine learning, in cui il modello può emergere dal data set attraverso i pesi e parametri solo con l'obiettivo di fornire abbastanza flessibilità per raggiungere correttezza lessicale*»¹⁴⁸.

Il dualismo pensiero-cosa viene sostituito da una rete di vettori-parola, legati tramite l'espressione “word embedding”. È un processo tecnico che serve a incorporare parole nei vettori entro un campo vettoriale governato da dei rapporti tra vettori e matrici.

¹⁴⁶ Gli schemi di Winograd sono dei test che vengono somministrati all'AI per cercare di riconoscere il significato di pronomi ambigui. Il lavoro quindi si basa su delle abilità di riconoscimento del significato secondo il linguaggio naturale. Furono ideati dall'informatico Terry Winograd nel 1972, seguendo la stessa idea di “misurare l'intelligenza” dei test di Turing.

¹⁴⁷ Manning, Christopher, op. cit., p. 134. Corsivo nostro.

¹⁴⁸ Weatherby, Leif, op. cit., p. 64.

Da un punto di vista meramente matematico quindi i LLMs processano quantità massive di dati trascritti secondo addizioni e moltiplicazioni. I *token* dei prompt che vengono immessi, assieme ai data set iniziali, sono vere e proprie etichette del corrispondente vettore. La parola nomina un vettore, a sua volta costituito da componenti (50000 per ChatGPT-3) che indicano le corrispondenti dimensioni entro il campo vettoriale. Dunque, l'immersione di parole entro lo spazio multidimensionale dei vettori corrisponde al vocabolario virtuale. Qui entra l'ipotesi strutturalista: la costruzione verbale più appropriata (il *response*) non è valutata a partire dal significato e nemmeno dalla mera frequenza sintattica, ma dalla distribuzione stocastica della stringa di parole rispetto al prompt immesso. È qui che «*il valore semantico di una parola, come lo fosse, senza riferimento al suo posizionamento sintattico nella sequenza di parole entra nel processo*»¹⁴⁹; come fa notare Kockelman, si può parlare di semiosi perché il significato emerge non da una definizione, oppure per mezzo della ricorsività dei modelli a reti (reiterando una definizione al suo *input*), ma dalla previsione che si cerca di esibire con altri *token*, che costituiranno le unità minime di significato con cui costruire il senso. Rispetto al prompt che si immette la macchina non lo interpreta, ma calcola una sequenza di tokens che con più probabilità si trovano a seguire quelli del prompt rispetto alla vasta quantità di dati contenuti che può vagliare.

Per riassumere ciò che si è detto fin qua, il processo di *word embedding* comincia con la scomposizione del prompt nei rispettivi token grazie a un “*tokenizer*”, che vengono tradotti in vettori, chiamati “*decontextualised word embedding*”. È proprio questo vettore a rappresentare il significato attraverso la posizione del vettore, data dalle sue componenti¹⁵⁰. Perciò, vettori vicini condideranno gran parte delle componenti e quindi si diranno di “significato” simile. Si ricorda che significato è da intendere non come definizione di un termine, ma capacità di venire inserito nel giusto contesto. L'aspetto posizionale che il significato prende lo mette in stretta correlazione con il senso secondo Deleuze. È proprio perché sono vettori ad essere lavorati e non le parole, per come il senso comune le intende nella loro primaria referenzialità, che è possibile generare nuovo senso. Di contro alla tendenza simbolista dell'AI e alla linguistica generativa di Chomsky, Weatherby ritiene che nulla di “innaturale” stia accadendo: «*le reti neurali ricorrenti si sono rivelate “irragionevolmente” efficaci solo sulla base dell'idea piuttosto singolare secondo cui l'elaborazione dovesse imitare la sintassi piuttosto che aggregare e*

¹⁴⁹ Kockelman, Paul, *Last Words*, Pinckly Paradigmatic Press: Chicago, 2024, p. 33.

¹⁵⁰ Nel caso di Chat GPT-3 si ritrovano 12288 componenti per vettore. Cfr. *ibid.*

mettere in relazione le unità all'interno del linguaggio scritto vero e proprio»¹⁵¹.

1.2 QVC IS ALL YOU NEED

È qui che si giunge al cuore della funzione *transformer*¹⁵² con il meccanismo di *attention*. Già incontrato in precedenza, è propriamente il momento del processo in cui nuova semiosi viene a prodursi. Il vettore “decontextualised word embedding” viene trasformato secondo:

- le relazioni che intrattiene precedentemente con il data set;
- il prompt inserito, quindi con gli altri tokens.

Attraverso una serie di moltiplicazioni tra matrici avviene la suddetta trasformazione. Infatti, il processo non prevede alcuna ricorsività o convoluzione, ma «l'attention *risulta essere un processo di analisi molto più semplice e diretto, ma esteso*»¹⁵³. Esso non prevede alcun trattamento differente per la sintassi e la semantica dei termini, come i modelli a rete precedenti. Non si comprende prima il significato di un simbolo, per poi memorizzarne le varie occorrenze e contesti in cui sintatticamente può situarsi. Seguendo il ragionamento del linguista Jakobson, il linguaggio si compone di due dimensioni: sintagmatica e paradigmatica. La prima dimensione, “orizzontale”, descrive l'andamento della composizione degli enunciati, ossia nel caso dei LLMs, quale *token* segue un altro. È la pura sintassi, ossia fornisce delle regole di costruzione degli enunciati. “Verticalmente” invece si pone la dimensione paradigmatica attraverso le possibili sostituibilità di un termine con l'altro. È possibile che lo stesso enunciato sia esplicito per diversi soggetti o oggetti: non si tratta di pura sinonimia, ma di sostituibilità di un termine con un altro che serve la medesima funzione sintagmatica. Lavorando sulla struttura si ottiene, mediante la *medesima* operazione sia la semantica che la sintassi: si tratta della moltiplicazione tra matrici QKV: “query”, “key” e “value”.

Lo schema riprende il già citato *document-term matrix* dei vasti database. Intuitivamente l'idea è la medesima che si ritrova per quasi tutte le ricerche online: le parole che si immettono nella barra di ricerca costituiscono una query, il vasto database in cui ricercare (potrebbe essere banalmente un catalogo di prodotti) è la key e il risultato

¹⁵¹ Weatherby, Leif, op. cit., p. 147.

¹⁵² È esattamente da *transformer* che deriva la “T” in ChatGPT.

¹⁵³ Weatherby, Leif, op. cit., p. 150.

che uscirà (la lista di prodotti rilevanti a partire dal titolo della mia ricerca) sarà il value. A differenza di questa modalità di ricerca un LLM però ha delle matrici QKV che vengono continuamente modificate; perciò, viene programmato per imparare a modificarle a seconda delle ricerche. Mentre il LLM riesce automaticamente in questo sviluppo, i database hanno bisogno di essere programmati manualmente; quindi, i dati devono essere inseriti dai programmatori stessi. È precisamente questo aspetto che viene riassunto nello slogan “learn to learn”.

Mentre un programma tradizionale accumula e classifica i dati, gli algoritmi apprendono attivamente facendo una selezione, quindi riducendo il numero di scelte, determinando automaticamente la più plausibile correlazione tra dati. Il Machine learning è usato in una situazione dove le regole non sono predisposte in anticipo, ma sono configurate grazie alla natura computabile del dato. Il Machine learning è perciò l'inverso della programmazione: non si tratta di derivare un risultato di un certo algoritmo, ma di trovare l'algoritmo che produca il risultato. Gli algoritmi devono acquisire dati per risolvere una *query*.¹⁵⁴

Un altro aspetto, ancor più importante, è che il prompt nel caso del modello linguistico viene usato per prelevare informazioni dalle parole per produrre nuove parole.

La tripartizione del processo è però superficiale. Non esistono zone o confini netti tra i vari gruppi di vettori. In realtà, query e key sono il medesimo vettore inizialmente, poi però, con il continuo adattamento ai parametri il vettore key si modifica. Anche il vettore value è il risultato della moltiplicazione tra i due vettori precedenti. Propriamente i dati nei LLMs non mantengono una netta divisione tra condizione e condizionato, un aspetto ripreso nel pensiero di Deleuze da *Differenza e ripetizione* a *Mille piani*. Non c'è contenuto altro rispetto alla forma e viceversa, così come le condizioni finiscono per essere i medesimi condizionati: non solo i vettori value possono essere visti come la trasformazione successiva di quelli key, ma l'inserimento di una nuova query comporta una modifica sia delle keys che dei values. Dunque, quando si immette un prompt, i tokens diventano la query e, grazie ai pesi delle key e delle value con cui vengono moltiplicate, vengono proposte in una stringa come candidate, fornendo il *response*. Le operazioni sono finalizzate quindi a definire con quale sequenza di parole (il *response*) i tokens immessi sono maggiormente connessi.

La connessione avviene non attraverso il ricorso a costanti sintagmatiche, a

¹⁵⁴ Szepanski, Achim, *In the delirium of simulacrum*, Becoming Press: Berlino, 2025, p. 225.

schemi prefissati che accolgono le variabili linguistiche; l'incrocio tra i due assi del linguaggio è dovuto alla reiterante operazione di *multihead attention*. Ogni processo matriciale QKV viene denominato *head attention*; ma contemporaneamente sono diversi processi di analisi che vengono compiuti, a seconda dell'aspetto semantico o sintagmatico che viene richiamato dalla sequenza. Non esiste mai una singola query, ma ognuna si apre a molti livelli di decodifica. Ad esempio, presa una sequenza di tokens si analizzerà il rapporto verbo-soggetto, oppure soggetto-oggetto. Non si tratta di analisi logica, ma di studiare ogni singola coppia per creare nuovi apparati di cattura lessicali. L'analisi condotta dalla *multihead attention* studia come ogni token della query, nel suo rapporto con gli altri, è possibile catturarlo in ulteriori sequenze. Lo spazio delle sequenze non è puramente grammaticale, semantico o lessicale, ma confonde tutte le dimensioni assieme, quindi «il posizionamento di una singola parola in una sequenza significativa è un'interazione con il modello complessivo e il contesto specifico: lingue e parole»¹⁵⁵. Per questo il termine migliore per definire l'output è proprio “values”: non parole, lessico, ma valori linguistici.

Conviene ora dirigere l'argomentazione sulla precisa struttura del processo computazionale linguistico degli LLMs. Il processo consta di due grandi regioni di *encoding* (area a sinistra nella figura 6) e di decodificazione per tornare dal vettore al token in uscita (area a destra nella figura 6). L'aspetto più innovativo della *multi-head attention* è che ogni vettore della sequenza non venga elaborato in serie, secondo un ordine sintagmatico predefinito, bensì in parallelo: ogni token diventa il punto di partenza di un proprio flusso di computazione, percorrendo simultaneamente la successione dei *Transformer blocks*¹⁵⁶. A differenza dei modelli RNN¹⁵⁷, la parallelizzazione è ciò che propriamente distingue il *Transformer*, in modo da ricavare grande risparmio di tempo e acquisire maggiore quantità di informazione da poter ricavare (la stratificazione del senso viene riprodotta in una disposizione che prenda per oggetto diversi aspetti semantici e linguistici). Il movimento a due tempi di *encoding* e *decoding* può ricordare quello della traduzione per Jakobson¹⁵⁸: ogni messaggio viene codificato in un sistema di segni interno

¹⁵⁵ Weatherby, Leif, op. cit., p. 152.

¹⁵⁶ Cfr. Alammari, Jay, Grootendorst, Maarten, *Hands-On Large Language Models*, O'Reilly, 2024, cap. 3 (“Looking Inside Large Language Models”): «ognuno di questi token *input* viene poi processato nel proprio percorso computazionale».

¹⁵⁷ Per RNN (*Recurrent Neural Network*) si intende una rete neurale in cui l'*output* viene immesso come nuovo *input* al sistema di codifica, andando così ad alterare i pesi e le misure del modello. È ormai un'architettura superata dalla novità dei *Transformer* (2017).

¹⁵⁸ Storicamente, il paper *Attention is all you need* (2017) intende proprio rimarcare l'utilità dei sistemi *transformer* in ambito di traduzione. Troveranno successivamente applicazione nella messa in dialogo di due sistemi eterogenei di segni come: immagini-parole, suoni-parole.

e poi ricodificato all'esterno (qui invece il vettore, una volta lavorato, viene trasformato in token in uscita). Per come si sarà intuito, la lavorazione interna non è neutra e definitivamente programmata, ma è una macchina differenziale che produce continuamente nuove combinazioni, possibili continuazioni di testo.

Nonostante la parallelizzazione dei processi, la sequenza *input* non perde la sua consistenza temporale, rispetto al posizionamento delle parole negli spazi della frase e nel tempo della narrazione. In altri termini, il processo non si riduce a una banale combinazione formale di elementi. Prima che il vettore *embedding* entri nel processo QKV viene integrato a un *positional encoding* che ne salvaguarda la risonanza entro la frase; la macchina apprende dove si trova nella sequenza e l'*attention* può computare il proprio risultato tenendo conto non solo dei vettori, ma anche della loro relativa distanza. Ne deriva uno degli aspetti che sopra si sono sottolineati dello strutturalismo: la temporalità (diacronia) ha valenza sincronica, ovvero è dovuta per combinazione di elementi simbolici, qui di vettori *embedding*.

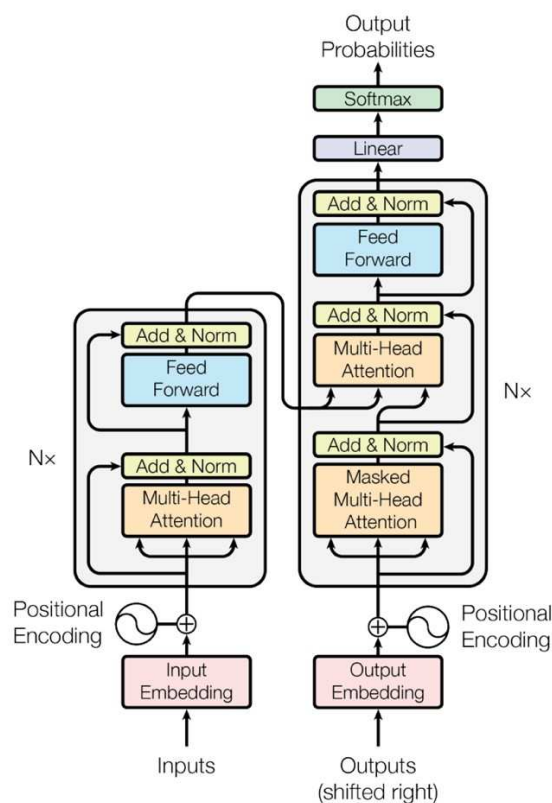


Figura 6 Processo di attention¹⁵⁹

Si prendano ora le tre matrici che si sono annunciate sopra, denominate W^Q, W^K, W^V , apprese durante la fase di *training*. Da queste tre matrici si ottengono tre ulteriori matrici tramite la moltiplicazione con la matrice dei vettori *embedded*: rispettivamente Q, K e V . È

¹⁵⁹ Immagine da <https://www.thinkstack.ai/glossary/attention-mechanism/>.

diventata celebre la formula presentata nel paper *Attention is all you need* :

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

La formula riassume l'intero processo *attention*. Anzitutto, tre sono le matrici che abbiamo identificato: si noti come le prime due, le *query* e *key* vengano moltiplicate al numeratore e solo in secondo luogo intervenga quella *value*. Il prodotto al numeratore restituisce il *relevance scoring*, quindi si può dire che l'interazione tra le prime due, misurato come rapporto scalare tra vettore-key e vettore-query, restituisce quanto un vettore sia rilevante per ogni altro. Una volta sondata la risonanza tra token, questa viene divisa per il valore numerico $\sqrt{d_k}$ per stabilizzare l'operazione successiva di *softmax* e fare in modo che non vengano rigettati valori estremi che comprometterebbero il processo.

Softmax è invece una funzione che trasforma un insieme arbitrario di numeri reali in una distribuzione di probabilità: ciascun valore viene costretto nell'intervallo da 0 a 1 e la somma tra loro viene coincidere con la totalità delle occorrenze (1). Formalmente la formula *softmax* di un vettore z è pari a

$$softmax(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Per $i = 1, \dots, K$, la formula permette di mostrare che, essendo basata solo su valori esponenziali, il risultato non può che essere positivo; la normalizzazione invece tramite la somma al denominatore garantisce che il risultato sia una distribuzione di probabilità. Dalla rilevanza si passa così al peso statistico, quindi ogni token distribuisce la propria attenzione tra gli altri. Anche qui, la probabilità è strutturalmente differenziale, ovvero relativa al campionamento precipuo e preso in considerazione e non per ordine superiore, a livello algoritmico-sintagmatico. La rilevanza, quindi, non è mai assoluta, ma relativa ai termini concorrenti (“cotesto” per Kockelman).

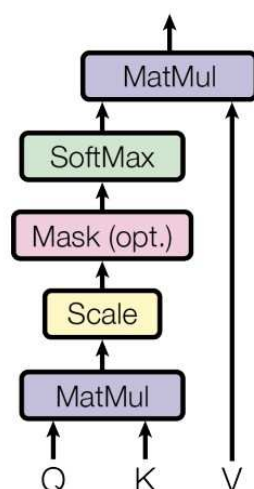


Figura 7 processo QKV¹⁶⁰

Così ottenuta la matrice con i pesi normalizzati, la si moltiplica per la matrice-value V : si ottiene il vettore di output sommando i values pesati secondo la probabilità di rilevanza. I vettori-value vengono mescolati in base al loro valore e alla loro rilevanza: *combining information*¹⁶¹. I vettori-value vengono moltiplicati sempre per quelli embedding iniziali: è come se le parole che compongono la sequenza iniziale venissero ancora una volta moltiplicate per una quantità che esprime l'informazione aggiuntiva, dovuta al rimescolamento probabilistico.

Come si accennava prima, il meccanismo di *attention* si dice essere *multi-head*, poiché viene svolto non secondo un solo punto di vista algoritmico, ma tenendo conto della pluralità di significati che si intersecano in una sequenza di *input*. La medesima sequenza viene decodificata e ricodificata tenendo presente una volta gli aspetti sintattici, un'altra le sfumature lessicali, un'altra ancora la coerenza pronominale, oppure aspetti di traduzione.

¹⁶⁰ Vaswani, Ashish *et al.*, *Attention is all you need*, in «Advances in Neural Information Processing Systems 30 (NIPS 2017)», p. 4.

¹⁶¹ Immagine da <https://www.analogictips.com/what-are-attention-mechanisms-and-how-do-they-work-in-speech-and-audio-processing/>.

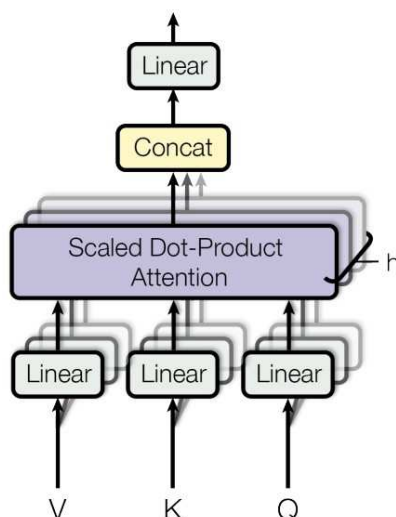


Figura 8 multi-head attention mechanism¹⁶²

Ogni unità di decodifica e ricodifica viene chiamata *head*, perciò a contraddistinguere queste macchine è la molteplicità di unità che riflette la polisemia linguistica, la stratificazione di senso che simultaneamente viene a svolgersi, restituendo in un'ulteriore matrice il vettore finale di output del *layer* dell'*attention*.

Come si vede dalla figura 6, il blocco arancione *multi-head attention* è seguito da uno azzurro *feedforward*. La *forward propagation*¹⁶³ è la trasmissione del segnale dall'input all'output, uno strato dopo l'altro. Sono infatti presenti degli strati [*layers*], composti da pesi e parametri, sviluppati durante la fase di *training*. Essi rappresentano la "memoria" del computer, che non è quindi né puntualmente localizzata, ma distribuita lungo diverse aree e soprattutto passibile di venire modificata, seguendo le prescrizioni tecniche dei modelli a rete. L'insieme di strati viene chiamato *multilayer perceptron* (MLP), attestando la discendenza dal *Perceptron* di Frank Rosenblatt nel 1958¹⁶⁴. Nonostante la differenza di calcolo (400 parametri per l'invenzione di Rosenblatt, comparati ai miliardi di GPT4), l'architettura rimane fundamentalmente invariata. In genere sono presenti più unità, ognuna a seguito di un singolo processo di *attention* per monitorare il risultato e rinforzare la memoria. Lo schema e il funzionamento del *perceptron* si basava sull'attivazione di potenziali secondo soglie. Gli input in entrata, a

¹⁶² Vaswani, Ashish, op. cit., p. 4. «Calcoliamo i prodotti scalari della query con tutte le key, dividiamo ciascuno per $\sqrt{d_k}$ e applichiamo una funzione softmax per ottenere i pesi dei valori».

¹⁶³ Immagine da <https://www.analogictips.com/what-are-attention-mechanisms-and-how-do-they-work-in-speech-and-audio-processing/>.

¹⁶⁴ Al contrario, si parlerà di *backward propagation* come di un'operazione di immagazzinamento del segnale in uscita durante la fase di *training*. La macchina deve imparare a ricordare le informazioni in uscita e quale output sono venute a costituire, perciò è grazie al processo *backward* che si studia il gradiente dell'errore rispetto ai pesi, aggiornandoli e modulandone il valore fino a minimizzare la distanza tra previsione del modello e la sequenza bersaglio.

seconda del segnale che giungeva ai differenti pesi, venivano rimescolati per giungere alle unità di *response* e formulare una risposta. Di seguito la formula che esprime il flusso di informazione che attraversa il Perceptron:

$$y = \sum_i w_i x_i + b$$

Si è deciso di esibire la formula per mostrare la sua compatibilità con i calcoli del processo di *attention*; entrambi i casi non sono altro che una serie di somme e prodotti vettoriali. Mentre $w_i x_i$ restituisce la variabilità degli input secondo una loro somma pesata, b costituisce il *bias*, ovvero il parametro collaudato a seguito della fase di training, che riflette i valori con cui gli LLMs sono programmati (*truthness, helpfulness, safeness*).

Generalmente sono presenti due strati di pesi per ogni processo di *attention*, intervallati da delle funzioni di attivazione non lineare, che fa funzionare la macchina *come se* “decidesse”. A questo punto si potrebbe avanzare un’ulteriore analogia che fa capire la rilevanza interpretativa strutturalista: come il processo *multi-head* non ha fatto altro che processare diverse informazioni contemporaneamente, secondo un’orizzontale intreccio, così quello di *forward propagation* prende singolarmente in considerazione ogni vettore, isolandolo. Si potrebbe allora sottolineare che, come alla dimensione orizzontale e sintagmatica della prima fase, il MLP fornisca una soluzione invece più in linea con la verticalità paradigmatica di cui parla Jakobson, in cui il modello compie un’operazione di selezione e di sostituzione tra possibili contenuti. La figura 6 rappresenta l’intera architettura *transformer* che può essere vista come una continua iterazione (decine di volte) di questi due processi (*multi-head attention* e *forward propagation*) per affinare il vettore output.

Infine, l’ultimo passaggio avviene riconvertendo i vettori-value in token da combinare per la trascrizione dell’output. Qui interviene una matrice particolare, “inversa” rispetto a quella che costituisce la fase iniziale di *embedding*, ovvero la *unembedding matrix* (detta anche *language modeling head* o *LM head*). Dalla continuità dello spazio multidimensionale dei vettori, si passa alle unità discrete di token del vocabolario digitale. Il processo consta di due fasi principali: inizialmente ogni vettore *embedding* viene moltiplicato per il token, ovvero una colonna della matrice *unembedding*. Ogni prodotto fornisce un punteggio chiamato *logit*, che riflette la rilevanza e la plausibilità per un certo token rispetto a quelli che l’hanno preceduto. Successivamente, su ogni *logit* viene applicata la funzione softmax che li trasforma in

una distribuzione di probabilità, che può essere gestita in diversi modi. Si chiama *decoding strategy* il processo di scelta e calibrazione del token a partire dal *logit* vettoriale: si potrà scegliere il valore più elevato, indice di maggiore rilevanza (*greedy decoding*) oppure si può campionare la distribuzione tramite tecniche *top-p* o *top-k*¹⁶⁵. Ad intervenire in questi processi è anche il parametro *temperature*, nient'altro che la distribuzione d'informazione tra i token candidati: più un token è rilevante e spicca rispetto agli altri, più la temperatura si dirà bassa (conducendo al caso limite in cui $T = 0$ e la scelta appare deterministicamente compiuta); al contrario, alta temperatura indica una distribuzione più uniforme, lasciando spazio a scelte più variate e meno prevedibili.

In virtù della generalità del termine “valore”, esso si accorda a una prospettiva linguistica *differenziale*; come insiste Weatherby, il linguaggio, seguendo la lezione di Saussure è indicato come un continuo scambio di informazioni, perennemente in movimento: in questo bilico tra fissità e fluidità, tra parametri dei *layers* e variabilità del vettore residuo si gioca la nozione di *metastabilità* linguistica. Ogni volta che si presenta un valore, esso viene scambiato con ciò che è fuori di esso, ovvero con la virtualità del linguaggio. Come anche mostra Deleuze, nei rapporti differenziali dx è indeterminato rispetto a x , allo stesso modo per cui il valore di un termine per Saussure non è confinabile al suo referente esterno. Ogni cosa può essere indicata, positivamente entrare a descrivere una parola, ma il suo valore eccede questo semplice rapporto biunivoco parole-cose: nel linguaggio «*ci sono solo differenze senza termini positivi*»¹⁶⁶. I termini non sono quindi atomisticamente astruibili dall'insieme in cui si trovano. Il fattore temporale diventa allora cruciale, perché è in relazione alla dimensione diacronica che il sistema si modifica; o meglio, la trasformazione strutturale, ossia sincronica, fa tutt'uno con la sua evoluzione temporale. Come si modifica la struttura? Attraverso ogni scambio di valori, di segni. «*La struttura spaziotemporale del messaggio diviene uno dei problemi decisivi per qualsiasi considerazione oggettiva del linguaggio*»¹⁶⁷. Tecnicamente, la semiotica classica di Peirce

¹⁶⁵ Per approfondire il funzionamento del *Perceptron* si rimanda a Pasquinelli, Matteo, op. cit., cap. 9.

¹⁶⁶ La strategia *top-k* e *top-p* sono strategie di *sampling* per restituire un token a partire dal valore vettoriale del *logit*. La strategia *top-k* consiste nel restringere la selezione ai k valori più alti, impedendo di generare token improbabili e mantenendo elevata la variabilità. Il limite di questo processo è che k rimane un insieme fisso di elementi, quindi, nel caso in cui la temperatura sia molto bassa, a venire selezionati saranno anche token relativamente improbabili, viceversa per una temperatura elevata verranno esclusi molti token rilevanti.

La strategia *top-p* invece di fissare un numero di token, seleziona una probabilità cumulativa p . Selezionando tutte le probabilità al di sopra di una soglia, il campione viene normalizzato ulteriormente. Spesso le due tecniche vengono usate in successione, escludendo prima i token improbabili e poi intervenendo restringendo le probabilità.

¹⁶⁷ Saussure, Ferdinand de, *Corso di linguistica generale*, trad. it. di T. De Mauro, Laterza: Roma-Bari 1967, p. 128.

sintetizza quest'evoluzione secondo un susseguirsi di segni e interpretanti. Ogni interpretante sarebbe un ulteriore segno che segue da quello precedente. Non c'è un legame di necessità tra segni. Il segno rimanda a un oggetto, che ne costituisce il significato, ciò per cui il segno sta e rappresenta. Il segno, rappresentando un oggetto, può venire codificato secondo varie occorrenze, tante quante i rapporti strutturali con gli altri segni indicano. Così ogni scambio tra segno e interpretante non è neanche mai impossibile. Il muto silenzio, l'inaspettata interruzione del discorso o dell'affaccendarsi è comunque un segno. Come gli elementi linguistici non si possono astrarre da un insieme, così i segni sono tutti concatenati.

Che rapporto c'è allora tra valori e segni? Tecnicamente ci si riferisce al segno come a ciò verso cui l'interpretazione dell'agente semiotico si indirizza: è l'oggetto della semiosi. È però il valore a mediare l'interpretazione del segno e farne perciò un interpretante nuovo. L'agente guardando un segno immediatamente vi costruisce un oggetto per definirlo e di conseguenza agisce (praticamente o teoricamente) tendendo a un interpretante. Il valore condiziona perciò la *scelta* dell'oggetto e di conseguenza dell'interpretante; Kockelman¹⁶⁸ definisce la funzione semiotica dell'agente come $A_v(S)$ come si evince dalla figura 4, laddove A è l'agente che interpreta il segno S attraverso i parametri valoriali V. Anche qui, non essendoci dei generi o delle specie fisse di appartenenza, vale lo stesso discorso dei LLMs, per cui le condizioni non sono più larghe dei condizionati. Il ventaglio di valori media la semiosi, assieme all'informazione che veicola il segno. L'intero processo è temporale e non logico: certo i suoi termini rivestono ruoli differenti, ma segno-interpretante sono l'uno l'evoluzione dell'altro. Il concatenamento di segni perciò finisce con l'essere una reiterazione differenziante¹⁶⁹, che si accorda con quella che Weatherby chiama «ipotesi differenziale»¹⁷⁰ di Saussure per cui

Per così dire [...] una sorta di algebra costituita esclusivamente da termini complessi [...] Ovunque e sempre esiste lo stesso equilibrio complesso di termini che si condizionano a vicenda. In altre parole, il linguaggio è una forma e non una sostanza.¹⁷¹

¹⁶⁸ Jakobson, Roman, *Parti e totalità del linguaggio* in *Qualità e quantità e altre categorie della scienza* a cura di Lerner, David, p. 208.

¹⁶⁹ Kockelman, Paul, op. cit., p. 20.

¹⁷⁰ Ci si riferisce all'algebra dei quaternioni che S. Bouquet ha rintracciato negli scritti inediti di Saussure. In realtà, anche Bouquet condivide che l'attribuzione non sia sicura, almeno non è mai stata apertamente sostenuta da Saussure (nei suoi scritti compare solo il termine *quaternion*); l'analogia portata avanti da Bouquet però ha considerevolmente modificato l'esegesi dei passi di Saussure, aiutando a definire la struttura più come un processo dinamico rispetto a una statica configurazione. Cfr. Saussure, Ferdinand de, *Scritti inediti di linguistica generale*, Laterza: Roma-Bari, 2005.

¹⁷¹ Weatherby, Leif, op. cit. p. 127.

La forma è esattamente la nozione complessa a cui si richiama Weatherby, di calibrazione tra *eidos* e *morphé*. Si tratta allora di affermare la natura metastabile del linguaggio, che in quanto tale non è niente al di fuori del suo sviluppo nel tempo.

Anche Shannon allora risulta compatibile con il ragionamento qui proposto, sostituendo alla nozione di segno quella di messaggio. La sua si potrebbe definire una prospettiva semiotica nella misura in cui cerca di *interpretare* un messaggio, codificandolo in significati, e non si interroga su come quel messaggio sia sorto. Anzi, la nozione di informazione si appoggia al grado di libertà dell'agente (destinatario) di interpretazione, che non vuole insinuare alcun libero arbitrio nella scelta del soggetto, anzi vuole testimoniare i diversi canali con cui un messaggio poteva venire tradotto, a seconda della quantità di informazione. La forma viene di volta in volta ridestata e si modifica in linea con la progressione temporale. La definizione di rapporti differenziali di Saussure è calzante. Nel caso dei LLMs, la struttura viene costruita di volta in volta dal meccanismo dell'*attention*, che fa delle parole i loro rimandi strutturali. L'aspetto, che ora bisogna prendere in esame, che permette questa continua reiterazione e trasformazione del linguaggio è la sua funzione poetica, così chiamata da Jakobson.

La funzione poetica, per Jakobson, indica la capacità del linguaggio di rivolgere il messaggio su sé stesso: è la dimensione in cui la forma del dire diventa costitutiva del senso, piuttosto che semplice veicolo di un contenuto preesistente. Questa intuizione trova nella proposta di Weatherby una trasposizione sistematica nella teoria dei LLMs. La macchina linguistica non accede al mondo extralinguistico se non attraverso la mediazione del testo, il che significa che il senso si genera sempre prima come struttura di relazioni interne e solo successivamente come rimando esterno. La funzione poetica è allora la condizione di ogni atto linguistico reso in modo computazionale. L'*attention mechanism* anzi realizza la funzione poetica: la proiezione dell'asse paradigmatico su quello sintagmatico, secondo la formula jakobsoniana, si compie ogni volta che la moltiplicazione QKV produce un vettore-valore.

Il token prodotto è scelto non perché più corretto, ma perché più in linea con l'indicizzazione statistica, ovvero la distribuzione delle connessioni entro la mappa intensiva del linguaggio [*poetic heat map*]¹⁷². La precedenza della forma non conduce solo a una questione linguistica, ma anche sociale. la forma è lo stesso schema di relazioni che ripartisce i compiti sociali entro una società. Come dimostra Pasquinelli¹⁷³, l'AI

¹⁷² Saussure, Ferdinand de, op. cit., p. 122.

¹⁷³ Weatherby, Leif, op. cit., p. 155. L'espressione «poetic heat-map» designa la matrice dei valori in quanto

ricalca non tanto l'intelligenza del singolo, ma l'intera intelaiatura sociale del lavoro; La tripartizione QKV acquista una valenza storica e non soltanto formale. La *query* condensa milioni di atti linguistici umani sedimentati nell'addestramento; la *key* è la forma in cui le classificazioni e le associazioni prodotte collettivamente nei *corpora* di testo vengono indicizzate; il *value* è il risultato (mai definitivo, sempre contestualmente rimodulato) in cui quella conoscenza collettiva si rende disponibile come risposta. La razionalità si distribuisce in pratiche collettive e le cristallizza in strutture matematiche. La forma computazionale è quindi a sua volta un sedimento di un pensiero collettivo.

Il legame con la funzione poetica di Jakobson si approfondisce ancora a questo livello. Non esiste un dizionario o di una grammatica sulla base di cui l'*attention mechanism* seleziona, ma deve la sua operatività alla densità delle co-occorrenze dei token nell'immenso archivio di pratiche linguistiche di cui è l'astrazione reale. Ma ogni *token* porta con sé le innumerevoli situazioni concrete in cui fu calato. In questo senso, il modello non è mai neutro: è, per riprendere una formula di Weatherby, la materializzazione dell'«ordine simbolico»¹⁷⁴ — il sistema di differenze che organizza il possibile del linguaggio in un dato momento storico. I LLMs rendono per la prima volta visibile e quantitativamente interrogabile quella struttura: non la mimano soltanto, ma la estendono e la deformano secondo le logiche del corpus su cui sono stati addestrati, tracciando la mappa di un ordine simbolico storicamente determinato che coincide, nella misura in cui è possibile, con ciò che Saussure chiamava lo stato del sistema in un dato momento diacronico.

Come suggerisce Weatherby, un LLM ha catturato il linguaggio «lungo il suo asse più essenziale»¹⁷⁵: non come rappresentazione del pensiero, ma come sistema *autonomo* di produzione di valori differenziali. In questo il suo funzionamento non si discosta dalla struttura che Saussure, Jakobson e Peirce hanno descritto, ciascuno a modo proprio, come il cuore del processo semiotico: il concatenamento di differenze che genera, senza mai esaurirlo, il campo del senso.

deposito di prossimità semantiche pesate statisticamente, che sceglie il token successivo non secondo regole sintattiche bensì secondo la densità delle co-occorrenze nell'intero corpus.

¹⁷⁴ Si rimanda a Pasquinelli, Matteo, *The Eye of the Master. A Social History of Artificial Intelligence*. Verso: London, 2023.

¹⁷⁵ Weatherby, Leif, op. cit., p. 198. L'«ordine simbolico» è ripreso da Lacan, ma Weatherby lo ricontestualizza.

CAP. 2: LA FUNZIONE POETICA DEL LINGUAGGIO

2.0 JAKOBSON E IL TELEGRAFO POETICO

Figura schiva e geniale, da come lo descrive Geoghegan¹⁷⁶, promulgatore di una teoria più sistematica e organica degli abbozzi e delle intuizioni del *Corso* di Saussure, Roman Jakobson può essere definito uno dei linguisti più influenti del XX secolo. I suoi studi germogliano nel contesto leninista in Russia sulla base di influenze dalla teoria della relatività, dal formalismo russo e dal cubismo. L'influenza linguistica del Circolo di Praga è notevole, facendo di Saussure uno dei suoi riferimenti intellettuali principali e indiretti, in occasione della *performance* Voder del 1944, il Circolo Linguistico di New York come successore di quello di Praga¹⁷⁷. Raggiungendo notevole fama all'epoca, le opere di autori, considerati padri dello strutturalismo, come Jakobson o Lévi-Strauss entrano a far parte del paradigma tecnico-scientifico (cibernetico) dell'arsenale US e delle sue infrastrutture per la comunicazione¹⁷⁸. Dopo aver accettato un posto alla ELHE, École libre des hautes études, tiene diverse lezioni sulla linguistica, servendosi di molte contaminazioni scientifiche e risultati della ricerca cibernetica, tanto da ritenere gli studi sulla tecnologia dei media come precondizioni della linguistica, andando contro il dogma saussuriano dell'esteriorità della *parole* rispetto alla *langue*¹⁷⁹. Saussure riteneva che, come il codice morse fosse esterno rispetto al telegrafo che ne permetteva l'attuazione, così anche il sistema della lingua non era limitato alle restrizioni degli organi vocali.

La rigorizzazione della teoria di Saussure da parte di Jakobson consegue anche da una sua precisa mereologia. Jakobson fu uno dei primi linguistici ad insistere sull'estensione delle categorie di parti e tutto che governano il linguaggio. Il problema dei primi linguisti fu di fermarsi alla frase: ci sono infatti ricadute e legami che stringono unità linguistiche maggiori o minori rispetto ad essa, che chiama "totalità"¹⁸⁰. L'enunciato, ad esempio, è suscettibile di includere un complesso più elevato di frasi, mentre il discorso a sua volta tiene assieme uno scambio di enunciati. La frase a sua volta può dividersi in parti più piccole come le parole o nei costituenti significanti minimi come

¹⁷⁶ *Ibid.*, p. 161. La formula completa recita: «we have basically captured language along its most essential axis» — il che non significa avere catturato l'intelligenza o la cognizione umana, bensì la struttura differenziale e culturale che è propria del linguaggio in quanto tale.

¹⁷⁷ Geoghegan, Bernhard, D., *Code: From the Theory of Information to French Theory*, Duke Press University: Durham, 2023, p. 77.

¹⁷⁸ *Ibid.*, p. 70.

¹⁷⁹ Si ricordi ad esempio il progetto incrociato tra il laboratorio psico-acustico di Harvard e il laboratorio di elettronica del MIT (diretto da Norbert Wiener) per lo studio di crittografia dei fonemi russi.

¹⁸⁰ *Ibid.*, p. 78.

i morfemi.

La linguistica strutturale, in particolare, fa del linguaggio un sistema comprensivo, un'unità di differenze mobili e implacabili. Come la biologia ritiene la morte di un essere vivente come limite di un equilibrio statico e definito, di conseguenza la vita propriamente come trasformazione incessante, tema centrale nelle pagine di Gilbert Simondon¹⁸¹; allo stesso modo la metastabilità linguistica, le sue continue e contingenti trasformazioni non sono una minaccia alla sua stabilità, tanto che l'apparecchiatura tecnica del telegrafo e del telefono sviluppata dalla Bell Company ispira il pensiero di questi autori¹⁸². L'ipotesi differenziale, come si è anticipato sopra, permette un allineamento delle dimensioni sincroniche e diacroniche, o come arriverà a scrivere Deleuze «*la genesi non si può concepire senza struttura e viceversa*»¹⁸³. La linguistica di Émile Benveniste, come di Jakobson, fanno del linguaggio qualcosa in continuo cambiamento e generante continuamente nuove forme, orientato verso trasformazioni future. Il co-adattamento tra sincronia e diacronia diventa però il terreno fertile per una transizione, che si registra nel corso degli anni '50¹⁸⁴ verso una rigorosa trascrizione algoritmica del linguaggio. Morris Halle aiutò a tradurre tanti termini della linguistica strutturale attraverso il ricorso a concetti della teoria dell'informazione, teoria dei giochi e cibernetica, stringendo sempre più il rapporto tra loro: «*il risultato è stato il ritratto del linguaggio come trasformazione digitale, codificabile per consentire calcoli algoritmici in grado di simulare interrelazioni più complesse tra termini*»¹⁸⁵. Sfumano quindi i confini tra cibernetica e linguistica, ancor di più tra ingegneria e cultura.

Il tentativo di rigorizzazione scientifica del linguaggio non fa che acuminarsi verso la fine degli anni '50, alla conferenza intitolata "Stile nel linguaggio", a cui parteciparono svariati cognitivisti come George A. Miller oppure critici letterari come I. A. Richards. Diversi interventi, tra cui in particolare quello di Jakobson, ricorsero allo schema generale della trasmissione del messaggio di Shannon; nel caso di Jakobson, egli ricavò sei distinte funzioni linguistiche a partire dalle sei componenti del sistema di trasmissione dell'informazione. Il linguaggio viene così commisurabile agli scambi codificati tra mittente e destinatario, ossia tra chi invia il messaggio e chi lo riceve. Il

¹⁸¹ Jakobson, Roman, *Parti e totalità nel linguaggio* in *Qualità e quantità e altre categorie della scienza*, Bollati Boringhieri: Torino, 1971, pp. 203-217.

¹⁸² Cfr. Simondon, Gilbert, *L'individuazione alla luce delle nozioni di forma e di informazione*, Mimesis: Milano, 2020.

¹⁸³ Anche qui si può trovare un punto di incontro *materiale* tra la linguistica strutturale di Jakobson e la teoria matematica delle comunicazioni di Shannon.

¹⁸⁴ Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, p. 281.

¹⁸⁵ Si pensi ad esempio alla pubblicazione a quattro mani dei *Fondamenti di linguistica* con Morris Halle.

medium è il messaggio, che incorpora sia la nozione semantica di messaggio per Shannon, sia quella di segnale. Il linguaggio, trasformato in una combinatoria di messaggi, riesce a sistematizzare il piano magmatico della *langue* di Saussure. *Langue* e *parole* vengono sostituiti dalla coppia di “codice” e “messaggio”, i due aspetti che maggiormente interessano ora. «Ogni volta che il mittente e il destinatario hanno bisogno di confrontare se stanno usando lo stesso codice, il discorso si concentra sul codice»¹⁸⁶. La funzione metalinguistica quindi è stabilizzante, crea delle equazioni tra parole e ha per oggetto il codice stesso. La funzione poetica non ne è opposta, ma è impossibile a quella metalinguistica. Tra le due funzioni si registra una tensione funzionale. Così ne parla Geoghegan

Laddove una stabilizza un discorso per assegnare rapporti nella comunicazione, la funzione poetica riflette invece su questi rapporti per generare nuovi discorsi. Laddove il metalinguaggio si eleva al di sopra del discorso per cogliere l'ordine del discorso, come nella forma di un'equazione, quella poetica affonda in profondità nel linguaggio, animando le relazioni matematiche intrinseche tra i termini per provocare una riscoperta sintetica delle possibilità del codice.¹⁸⁷

È allora possibile che il linguaggio sia passibile di entrambe le “analisi”. Jakobson sta quindi tentando di innalzare la funzione poetica al rango di prim'ordine scientifico, dove usualmente i linguisti posizionano i ragionamenti metalinguistici. Invece, anche la poetica del linguaggio crea delle equazioni, riflettendo sull'intima possibilità *generativa* del linguaggio. È anch'essa un'analisi riflessiva, che chiude il linguaggio su sé stesso, ma a livello di messaggio e non di codice. Come detto poco sopra, il codice viene studiato dal metalinguaggio, lasciando invece alla funzione poetica operare un'analisi del ritmo della frase, la sua è la «forma generale dei rapporti complessi»¹⁸⁸ tra termini che si rappresenta in una combinatoria di cui disporre per costruire nuove enunciazioni.

Il linguaggio così si costruisce attraverso due poli fondamentali: la sua funzione poetica e quella comunicativa. Larga parte di questo ragionamento si deve alle influenze formaliste russe, riconoscendo un implicito rimando tra identità e non identità del segno e del referente. Tutta la critica anti-referenziale di Weatherby si può leggere a partire da un ragionamento di critica letteraria, operato dai formalisti di allora. Quando si parla,

¹⁸⁶ Geoghegan, Bernhard, D., *Code: from the theory of information to french theory*, Duke Press University: Durham, 2023, p. 81.

¹⁸⁷ Jakobson, Roman, *Closing statements: linguistics and poetics*, in Sebeok, Thomas A., *Style in Language*. Cambridge, MA: MIT Press, 1966, p. 356.

¹⁸⁸ Geoghegan, Bernhard, op. cit., p. 85.

implicitamente la parola viene identificata con l'oggetto di riferimento, ma è compito della critica rimarcare sempre l'arbitrarietà del legame, dunque la non-identità, per evitare una deriva verso un automatismo percettivo (credere che le cose siano descrivibili attraverso solo quelle parole che si suole attribuirle); è però a causa della struttura e della sua rete di connessioni che si impone il segno: l'arbitrarietà del rapporto tra segno e referente viene imposta dal codice culturale della struttura. La critica letteraria è perciò principalmente un'arma contro l'omologazione della lingua, ma anche della percezione. Credendo che parole e cose coincidano, i legami e le percezioni che si innestano tra enunciante e mondo si restringono e rischiano di venire dedotti unicamente da quello che Deleuze chiamerà "regime di segni" verbali; in altre parole, solo l'antinomia tra identità e non identità di segno e referente permette di contrastare e criticare l'impovertimento della realtà che deriva un sempre più stringente automatismo percettivo. Secondo Erlich¹⁸⁹ il poeta combatte per liberare l'uomo dagli assoggettamenti sociali attraverso la liberazione del linguaggio da una sua ferrea codificazione. Rispetto a questi presupposti programmatici, il linguaggio esprime una funzione comunicativa rispetto alla trasmissione neutrale e ordinaria di messaggi, oppure una funzione poetica. A differenza di Odgen e Richards¹⁹⁰, che identificano funzione comunicativa e poetica rispettivamente con un uso conoscitivo ed emotivo del linguaggio, Jakobson afferma che la funzione poetica:

Percepisce la parola come parola e non solo come sostituto dell'oggetto designato come esplosione emotiva; infatti le parole e il modo in cui sono organizzate, il loro significato e la forma interna ed esterna acquistano peso e valore autonomo.¹⁹¹

Jakobson esprime così l'autonomia della sfera del linguaggio dal referente. Il codice è il presupposto di una *langue* saussuriana che interviene non per ineffabili scambi differenziali, ma per trasmissione di segnali economicamente traducibili, perché crittografabili tramite le principali infrastrutture della comunicazione (Bell Company ad esempio). L'arbitrarietà del segno, quindi finisce per ampliare le possibilità semantiche del linguaggio rispetto alla situazione reale. Se la realtà non può venire designata in maniera inequivocabile, finendo per impoverire la percezione dell'individuo secondo automatismi lessicali, d'altra parte così il linguaggio può attraverso la sua ricombinazione

¹⁸⁹ Weatherby, Leif, *Language machines*, Minnesota University Press: Minneapolis, 2025, p. 155.

¹⁹⁰ Elich, Viktor, *Il formalismo russo*, Bompiani: Milano, 1966, p. 240.

¹⁹¹ Odgen, Richard, *The meaning of meaning*, Dead Author Society, 2023.

e alterazione semantica, produrre nuovi significati, quindi generare nuova cultura.

Jakobson quindi mantiene di grande importanza la funzione poetica del linguaggio, in virtù della rinnovata formulazione di enunciazioni a cui può portare. È il messaggio che riflette su sé stesso, funzionando come campionario di ritmi, stili da poter riapplicare. Per Weatherby è questa funzione allora che consente di programmare una forma e quindi di costruire un linguaggio attraverso la sua riproduzione algoritmica. Entrambe le funzioni consentono di instaurare dei rapporti di equivalenza tra termini, solo procedono secondo direzioni opposte. *La funzione metalinguistica descrive delle equivalenze semantiche, partire da una sequenza (una frase, una citazione, un frammento, un verso...), mentre quella poetica procede da delle equivalenze sintattiche alla costruzione di sequenze.* Sintattico qui sta per non concettuale, quindi non definibile. Sono infatti dei suoni, dei ritmi, degli accenti oppure delle particolari maniere di articolare la frase (qui sì, la sua sintassi) a permetterne la riproduzione. Quindi “poetica” indica non un riferimento alla poesia letteraria, piuttosto alla letterarietà del linguaggio, onnipresente in ogni sua enunciazione. Per questo Jakobson la ritiene la funzione più importante¹⁹²; il suo tentativo di rigorizzazione della funzione poetica consegue a un inserimento della poesia e della letteratura entro la linguistica; dunque, si muove in senso molto più esteso rispetto ad essa, seguendo l’etimologia greca di *poiesis*, ossia “fare”, “creare”, “generare”.

Il suono delle parole, dimensione da cui si possono integrare nuovi legami di equivalenza, è primario rispetto alla concettualità metalinguistica per due motivi: viene prima di essa ed è più esteso. Per rappresentarlo, ne parla in termini di fonologia, ovvero lo studio del campionamento e della ricombinazione di suoni. L’idea del suono “originario”, che interviene prima della sua fedele trascrizione grammaticale, è una pista che segue anche Deleuze attraverso il lavoro di Melanie Klein, per dimettere il, privilegio del *logos* rispetto a tutto ciò che è a-concettuale, ovvero il nonsenso.

Ma ciò che qui importa è l’organizzazione preliminare, fondatrice o poetica: questo gioco della superficie in cui si dispiega soltanto un campo acosmico, impersonale, preindividuale, questo esercizio del senso e del non senso, questo dispiegamento di serie che precedono i prodotti elaborati dalla genesi statica. Dall’ordinamento terziario bisogna dunque risalire fino all’organizzazione secondaria; risalire poi fino all’ordine primario, seguendo l’esigenza dinamica. [...] Ma tutto il percorso, tutto il cammino è segnato dall’ordine primario. Nell’ordine primario le parole sono direttamente azioni o passioni

¹⁹² Erlich, Viktor, op. cit., p. 233.

del corpo oppure voci ritratte. Sono possessioni demoniache oppure privazioni divine¹⁹³.

Il passo appena riportato però introduce qualcosa di ulteriore rispetto alla riflessione linguistica. Il senso ha anch'esso una sua genesi "dinamica" e non può essere definibile a partire dalla sola dimensione metalinguistica. Jakobson, invece, indica sì una pluralità di dimensioni, ma rimangono tutte complementari e composibili (nella misura in cui sono presenti contemporaneamente, ognuna secondo diverse gradazioni) in una sequenza; perciò, il senso rimane di esclusiva pertinenza del metalinguaggio e delle sue statiche categorizzazioni a cui Deleuze intende opporsi. Jakobson comunque rimarca una precedenza della sintassi fonologica sulla semantica. Il fonema, inteso come più piccola unità significativa¹⁹⁴ possibile di suono nel discorso e nel linguaggio, viene fatto associare ai *bits* di Shannon, per registrare e trasmettere informazione. L'informazione corre così ad un livello anteriore a quello del significato, un livello strutturalmente denso di morfemi e fonemi. Jakobson scopre l'influenza vocale dei fonemi, ma la storia della funzione poetica viene estesa da Weatherby anche ai *social network* e ai LLMs. Suono, testo e matrici QKV.

Ora, le matrici dei LLMs incarnano un linguaggio sulla base della sua funzione poetica. Lo si dimostra attraverso il meccanismo di *attention* di cui si è parlato sopra. Ogni sequenza viene analizzata proprio attraverso un'auto-riferimento possibile non perché basato sul referente esterno, aspettando una coincidenza tra enunciante ed enunciato, qualcosa di simile alla sovrapposizione chomskyana di *I-language* e *E-language*¹⁹⁵, ma perché calcolato tramite la scrittura del prompt.

Chomsky scrive che "non sussiste alcuna questione di correttezza riguardo alle *E-languages*, comunque esse siano caratterizzate, poiché le *E-languages* sono semplici artefatti". In altre parole, sebbene l'*I-language* produca l'*E-language*, a un certo punto la generazione deve attraversare un confine tra interiorità ed exteriorità, dopo la quale non dovrebbe svolgere alcun ruolo nella teoria del linguaggio — poiché "le lingue in questo senso non sono oggetti del mondo reale, ma sono costrutti artificiali, in qualche modo arbitrari e forse non molto interessanti". Lo "stato stabile di conoscenza" dell'*I-language*, al contrario, deriva dal fatto che la sintassi interna è costituita da "elementi

¹⁹³ Weatherby, Leif, op. cit., p. 159.

¹⁹⁴ Deleuze, Gilles, *Logica del senso*, Feltrinelli: Milano, 1975, p. 216.

¹⁹⁵ Significazione qui non va intesa semanticamente (a cui è legato invece il "morfema"), ma legata al riconoscimento tecnico del messaggio da parte delle apparecchiature elettroniche. È grazie al riconoscimento di questi fonemi e alla loro integrazione con altri che si possono costruire sequenze di messaggi e trasmettere informazione, ci contro al rumore sottostante.

reali di menti/cervelli precisi, aspetti del mondo fisico.¹⁹⁶

La mossa critica di Weatherby nei confronti di Chomsky sta tutta qui: gli LLMs dimostrerebbero, grazie al calcolo statistico reiterato sul dataset che l'*E-language* non può essere concepito come un mero artefatto estrinseco e che la capacità di esprimere un linguaggio non può essere ristretta ai soli esseri umani. Si dice che la parola successiva tratti assieme a quella precedente perché è statisticamente rilevante. La rilevanza è un affare fonemico, di connessione intensiva (segue infatti a una variazione costante, a un equilibrio metastabile) perché congiunta a una complessità di fattori: semantica, contesto intra-linguistico, la specificità della richiesta del prompt, l'impostazione di fine-tuning.

Weatherby rende l'idea di questa intensività con il concetto di “mappa termico-poetica” [*poetic heat map*], che mette in correlazione tutte le ridondanze. La funzione poetica del linguaggio avverrà così sulla base di un calcolo statistico di connessioni tra fonemi, come tra tokens o tra sequenze stesse.

I punteggi dell'attention assegnati alle sequenze durante l'elaborazione indicano in che misura ogni parola è correlata [*aboutness*] alle altre parole in un determinato testo. Ciò che in origine era stato proposto come “attention” è diventato “self-attention”, in cui le rappresentazioni delle parole vengono aggiornate esclusivamente dai punteggi dell'attention interni a ciascuna sequenza immessa nel sistema.¹⁹⁷

La correlazione di una parola con un'altra è riflessa entro una misura quantitativa (sempre fluttuante) e il calcolo avviene sulla base dell'estensione del dataset. «Aboutness è poetica o ricorsiva rispetto al messaggio prima di essere metalinguistica o referenziale»¹⁹⁸. Quanto una parola c'entri con quella prima è primariamente un discorso intra-verbale, quindi dovuto alle connessioni che si esercitano tra parole, come il prompt, oppure l'algoritmo le scrive assieme. Il processo di combinazioni sovra-lineari, ovvero mai disposte secondo una logica sequenziale ad addendi fissi, non conosce alcun “fuori”, alcuna exteriorità a cui attingere, sia questa del tipo del referente e del mondo fisico, oppure secondo delle determinazioni metalinguistiche.

¹⁹⁶ La distinzione fondamentale tra *I-language* e *E-language*, approntata da Chomsky, mira a eliminare ogni residuo aleatorio, accidentale, insensato del linguaggio che viene a prodursi a partire da protocolli culturali, ovvero tutte le sue manifestazioni esteriori che si attestano precisamente nell'*E-language*. L'interiorità (*I*) dell'*I-language* consiste invece nelle funzioni cerebrali del cervello umano che garantirebbero la formazione del linguaggio.

¹⁹⁷ Weatherby, Leif, op. cit., pp. 52-53.

¹⁹⁸ Weatherby, Leif, op. cit., p. 162.

2. 1 NUMERI E PAROLE: UNA RELAZIONE APERTA

Due sono i sistemi che vengono fatti interagire tra loro; “sistema” qui perde, come si vedrà, la valenza definitoria che lo contraddistingue classicamente. Non si tratta di sillogismi, ma di strutture generative, dunque in continuo movimento, secondo quella sovrapposizione dell’aspetto sincronico e diacronico che si è vista sopra. Lo scorso paragrafo ha dimostrato che è precisamente la funzione poetica del linguaggio (altra cosa rispetto alla poesia) a garantirne la perenne produzione di enunciati, riferendosi principalmente a Jakobson e Shannon.

Poetica è la capacità di autoriferimento del messaggio a se stesso, ciò che Weatherby qualifica col termine “avidità” [*greedy*]. Infatti, è come se il linguaggio fosse una riserva inesauribile di enunciazioni, un processo di scoperta interminabile nei confronti del mondo. Il processo epistemologico di conoscenza non dovrebbe essere pensato come meramente non-concettuale, costituito da una sintassi extra-linguistica, ma come l’addomesticamento di qualcosa di extra-linguistico da parte del linguaggio; più precisamente, nel caso dei LLMs, una traducibilità del referente secondo pattern linguistici. L’avidità è una tensione inesaurita a tradurre tutto a parole; quindi, potenzialmente *tutto* può venire tradotto linguisticamente, tutto può rientrare nella combinatoria di pattern. «*Per ogni ente che può applicare una logica e programmare le proprie decisioni, il linguaggio non può che essere il medium di questa operazione*»¹⁹⁹. Insomma, la relazione intima tra linguaggio e computazione sembrerebbe essere non solo sufficiente per garantire una copertura totale del linguaggio sul mondo, ma anche necessaria: non ci può essere logica senza linguaggio.

Il linguaggio naturale può essere definito “avido”, quindi teso a coprire la totalità degli enti del mondo, tanto quanto la logica ad esso avviluppata può essere definita “incompleta”.

Alla conclusione che ogni sistema formale debba considerarsi incompleto, se non inconsistente, giunge il matematico Kurt Gödel (1906-1978), una delle più iconiche figure del Circolo di Vienna. Nonostante le ambizioni del gruppo di ricerca, orientate verso una cattura logica totalizzante del mondo, è grazie alla lettura dei *Principia Mathematica* di Bertrand Russell e Alfred North Whitehead che si convince che tali ambizioni non potevano che sfociare in voli pindarici. Un sistema si definisce “completo”

¹⁹⁹ *Ibid.*

se ogni proposizione vera contenuta al suo interno può essere dimostrata; “consistente” invece significa che un sistema non contiene alcuna contraddizione e può venire interpretato, ossia contiene il suo senso²⁰⁰. Gödel dimostra che i due aspetti non sono componibili, dunque un sistema non potrà mai essere completo, essendo anche privo di qualsiasi contraddizione. Ugualmente si può dire che ogni sistema formale consistente, che può venire interpretato perché dotato di senso, è incompleto, quindi che bisogna ricorrere a qualcosa che stia *fuori* di lui per conoscerlo, per redigerne le leggi (logiche) di funzionamento.

Senza entrare in dettagli tecnici, l'antinomia dei due termini si dimostra attraverso l'autoreferenzialità di un sistema. Attribuendo ad ogni proposizione logicamente dimostrabile un numero naturale, il sistema mentre “parla” di sé sta anche calcolando. Ogni considerazione meta-matematica consegue dal puro calcolo matematico di numeri, ossia per una questione di pura sintassi, Gödel ha generato la metamatematica dai numeri naturali. Il numero di Gödel è una operazione che stabilisce ricorsivamente che la funzione sia dimostrabile entro il sistema, ma, sviluppandola, il contenuto che ne esce è che questa funzione non sia dimostrabile. Il paradosso dimostra allora che il second'ordine di «sintassi non esaurisce la nozione di legge, ossia che un sistema per dimostrare una nuova regola ha bisogno di appoggiarsi su un altro»²⁰¹. I numeri che un sistema formale calcola non derivano da delle leggi di secondo ordine e, ciò nonostante, sono programmabili logicamente; a questa definizione di computazione si aggancerà poi Alan Turing e la moderna informatica.

A questo punto, se un sistema formale è consistente, si può sopperire alla sua incompletezza proprio grazie alla proliferazione linguistica (*greedy*). Ogni operazione di pensiero, programmabile perché logicamente descrivibile, può raggiungere un referente esterno, al di là di qualsiasi contingente stato d'analisi in cui si trova.

Non c'è alcun metalinguaggio che non sia linguisticamente descrivibile, dunque che non sia linguaggio a sua volta. L'impossibilità per le regole sintattiche di uscire dal linguaggio dimostra, ancora una volta la primarietà, intuita da Jakobson della funzione poetica. Infatti, ogni volta che un prompt viene immesso in un LLM, esso non può che fare *anche* riferimento a sé stesso, ossia alla sua stessa scansione che viene calcolata dall'*attention*. Il meccanismo di *attention*, rappresentando la funzione poetica del linguaggio, è la prova concreta che ogni proposizione non può che riferirsi a sé stessa,

²⁰⁰ Weatherby, Leif, op. cit., p. 166.

²⁰¹ Per approfondire la dimostrazione di Gödel si rimanda a Berto, Francesco, *Tutti pazzi per Gödel!*, Laterza: Roma, 2009.

ovvero è anche autoriferita. Certo, la sua referenzialità può spiccare rispetto alla sua funzione poetica; la sua funzione emotiva è ciò che magari risalta maggiormente, ma ciò non toglie che, infine, possa venire ricondotta alle stesse parole con cui si è costruita, quindi con la trama di relazioni che ha messo in piedi. «*la generazione [linguistica] è riflessiva*»²⁰² e primaria rispetto alle altre funzioni perché ogni enunciato può esserne ricondotto.

Non ci possono essere sequenze che escludano le relazioni interne tra le loro parole (o ancor più generalmente, token); se ci fossero non potrebbero essere dell'ordine del linguaggio, ovvero non potrebbero essere accompagnate da una spiegazione verbale (sarebbero perciò incomplete per Gödel). Per dimostrare questo punto può essere utile tornare brevemente a Shannon, notando che la probabilità di un'occorrenza linguistica è sempre compresa tra 0 e 1, dunque che ogni cosa che si dice è potenzialmente correlabile ad altro. I casi estremi dimostrerebbero il contrario, ma così si finirebbe per essere al di fuori dal linguaggio. Anche se impossibile, la dimostrazione extra-verbale dovrebbe allora avere una ridondanza di 1 o di 0, rispettivamente illuminando la necessità o l'impossibilità che un'occorrenza venga dopo un'altra. La funzione poetica è non è nient'altro che il dispiegamento del prim'ordine del linguaggio, lo sviluppo del linguaggio naturale (NPL). Così allora sintetizza Weatherby:

Realizzando la funzione poetica di definizione del valore in un sistema di rappresentazione, il LLM mette in luce il nucleo dialettico dello strutturalismo. La generazione di significato è contraddittoria in senso dinamico. È proprio questa preminenza della contraddizione generativa che permette al calcolo e al linguaggio di condividere la stessa forma.²⁰³

Il significato è allora formale non nel senso di indipendente e distaccato, come fosse un invariante, ma perché è descrivibile come una sequenza di parole che lo sostengono. Il significato è generativo, appartiene alla generatività della funzione poetica, ma può essere poi classificato e indicizzato. Il significato è quindi una tensione tra funzione metalinguistica e poetica. «*Qualsiasi che è generata può venire catturata, può entrare – ed entra – nel processo statistico di misura del linguaggio, non importa quanto piccola sia la sua rilevanza*»²⁰⁴: nulla di più vicino a quanto intendesse Weaver, con l'esempio su

²⁰² Weatherby, Leif, op. cit., p. 167.

²⁰³ *Ibid.* p. 169.

²⁰⁴ *Ibid.*

Costantinopoli²⁰⁵. “A Costantinopoli c’è un pescatore...” è un incipit del tutto improbabile, ma non impossibile. Nonostante un modello possa non prevederla, la frase può venire calcolata comunque, in virtù della sua adeguatezza sintattica, semantica.

Il linguaggio si contraddistingue per essere aperto verso se stesso, continuamente auto-riferibile perché poeticamente generativo; ogni sistema formale invece è fondamentalmente aperto verso l’esterno, perché se consistente non può che essere incompleto, dunque per dimostrarlo bisogna uscire da lui, trovando una dimensione di ordine superiore.

²⁰⁵ *Ibid.* p. 171.

PARTE 3 GENERARE CULTURA

CAP. 1: QUALE GENERATIVITÀ?

La capacità linguistica di generare dei *prompt*, spesso allineati con le intenzioni dell'utente, capaci quindi di intercettare i valori, desideri e pensieri che lo stesso sottopone all'interazione con la macchina è ciò che si è dimostrato grazie alla caratteristica conformazione *strutturale* degli stessi modelli. In virtù della connessione che investe ogni *token* con gli altri, è possibile articolare un *response* che sia allo stesso tempo formalmente consistente [*rightful*] e funzionale [*helpful*] rispetto alle esigenze, quindi alle intenzioni della persona. Il ricorso alle connessioni statistiche di una parola con le altre, rende queste macchine tanto performanti quanto cieche; il processo semiotico si basa sul calcolo statistico del *token* successivo, non acquisendo alcuna informazione sul referente.

In virtù della potenza di calcolo e della quantità di dati resi la macchina si contraddistingue per la sua eccezionale accuratezza; al contempo, la sua apertura al mondo è dovuta in parte agli scambi (*chat*) con l'utente. Affermare che i LLMs non siano altro che una forma avanzata di *stochastic parrots* è una tesi insufficiente per qualificarli e contestualizzarli adeguatamente. La tesi di molta letteratura critica a riguardo sostiene che la loro potenza di calcolo si traduca in una mera acquisizione e conseguente codificazione statica di informazioni: in altre parole, il *response* non è nient'altro che un assemblaggio stocastico di informazioni che sono codificate secondo la probabilità che le unisce e le connette tra loro. La connotazione di un termine verrebbe a coincidere con la sua quantificazione in termini probabilistici e ciò che risulterebbe essere adeguato alle intenzioni degli utenti non è nient'altro allora che una risposta di senso comune. Affermare che un'operazione del genere sia semplicemente casuale significa aggirare il problema in ultima istanza, non riuscendo a spiegarne il motivo. Ad esempio, Emily Bender, professoressa e ricercatrice in linguistica computazionale, ritiene che costruendo casuali associazioni, i *response* finiscono per essere privi di qualsiasi intenzione, potendoli definire come “testi sintetici” senza alcun significato²⁰⁶.

²⁰⁶ Shannon, Claude, E., Weaver, Warren, *La teoria matematica delle comunicazioni*, ETAS Kompas: Milano, 1971.

D'altro canto, essendo che la struttura probabilistica del linguaggio consente di processare delle risposte *come se* fossero restituite da una mente umana, diventa sempre più facile riferirsi a questi motori come delle “menti pensanti”, delle “intelligenze artificiali” nel senso letterale del termine. L'intelligenza presa a modello da queste macchine però, come si è già visto, non riproduce la precipua operazione di trasduzione ibrida di dati non simbolici in un simbolismo concettuale che ne consente la manipolabilità. L'input, la materia pre-esistente alla computazione macchinica non è pura materialità che deve venire interpretata, entro una serie radicalmente differente, non c'è alcuna *differenza* tra input e output a livello “sostanziale”: entrambi sono e rimangono degli assemblaggi stocastici di *tokens*.

Per questo motivo, il ricorso a un modello di semiotica peirciano risulta essere più appropriato: liberando il campo dalla dannosa identificazione di significato e referente, con la coppia segno-interpretante, la differenza risulta essere esprimibile quantitativamente come una modulazione delle probabilità di una nuova serie di token rispetto a tutte le occorrenze precedenti. Il processo di significazione non coinvolge parole e cose, ma una continua trasformazione, secondo un preciso calcolo probabilistico, delle connessioni tra *tokens*. Per questo, le macchine non possono venire precisamente qualificate come “intelligenti”, piuttosto generanti una differente composizione di termini “ciechi” (perché slegati da qualsiasi referente e legati alla sola funzione poetica del linguaggio), a cui ci si può riferire col termine di simulacro. Non potendoli definire come semplici assemblaggi casuali di parole [*stochastic parrots*] ma neanche come macchine intelligenti, come può essere concepita la loro straordinaria e talvolta eccessivamente “umana” generatività di risposte? La loro è un'autentica forma di creazione o un rimescolamento posticcio?

1.0 SINTASSI E PRAGMATICA

Un primo aspetto che deve essere menzionato è la capacità generativa sintattica; dunque, la loro produzione di infinite proposizioni a partire da un *set* di regole finite (in particolare il lessico e la grammatica), citando Wilhelm von Humboldt. Uno dei riferimenti intellettuali di maggiore spessore in questo ambito è Noam Chomsky. Con la definizione di “ricorsività” linguistica, ogni proposizione è vista come passibile di un'ulteriore integrazione, di un'aggiunta possibile. Gli esseri umani non solo sono capaci di generare infinite frasi grammaticalmente coerenti, ma sono altresì capaci di generare infinite

estensioni a partire da una frase per aggiunte di parole o modificazioni grammaticali. L'aspetto precipuo del linguaggio non sta nelle parole che vengono usate e nella loro ripartizione, quanto piuttosto nella profonda generatività del linguaggio.

Si prenda il celebre esempio di Chomsky sulla grammaticalità del linguaggio, che esibisce il tradotto tipico (umano quindi dice) del linguaggio: siano date due frasi:

- 1) Colorless green ideas sleep furiously. 2) Furiously sleep ideas colorless.

L'esempio dimostra come qualsiasi parlante di lingua inglese sia portato a riconoscere la proposizione (1) grammaticalmente corretta, nonostante siano entrambe prive di senso. non avendo senso, è comunque più facile inferirlo dalla proposizione (1), rispetto alla (2). L'aspetto che contraddistingue il linguaggio non può essere derivato dalle parole, ma dalla maniera grammaticale con cui si connettono tra loro e si creano frasi, discorsi.

Di fronte a queste frasi, un parlante madrelingua inglese leggerà (1) con un'intonazione normale, mentre leggerà (2) con un'intonazione calante precisamente su ogni parola; in effetti, con lo stesso schema d'intonazione che si attribuirebbe a qualsiasi sequenza di parole non correlate. Egli considera ogni parola in (2) come una frase a sé stante. Allo stesso modo, sarà in grado di ricordare (1) molto più facilmente di (2), di impararla molto più rapidamente, eppure potrebbe non aver mai sentito o visto nessuna coppia di parole di queste frasi unite in un discorso reale. [. . .] Evidentemente, la capacità di produrre e riconoscere espressioni grammaticali non si basa su nozioni di approssimazione statistica e simili. L'abitudine di definire frasi grammaticali come quelle che "possono verificarsi" o quelle che sono "possibili" ha causato una certa confusione in questo ambito.²⁰⁷

Il linguaggio è la capacità creativa di generare infinite proposizioni grammaticalmente corrette e nessuna di grammaticalmente scorretta, perciò vengono formulate frasi mai pronunciate prima con la massima naturalezza, ricombinando gli elementi delle lingue grazie a un'innata capacità che non può mai essere dedotta dagli elementi stessi di quelle lingue. La prospettiva chomskyana, come mette in evidenza Weatherby²⁰⁸ cerca di combattere quella statistica di Shannon, concentrata su come interpretare un messaggio e non ai parametri e alle categorie attraverso cui è venuto a prodursi.

²⁰⁷ I testi si possono dire sincretici nella misura in cui il calcolo dei simulacri linguistici viene prodotto sulla base della pura forma (disposizione dei vettori nello spazio latente) cfr. Bender, Emily, Gebru, Timnit, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", <https://dl.acm.org/doi/10.1145/3442188.3445922>.

²⁰⁸ Chomsky, Noam, *Syntactic Structures*, De Gruyter Mouton: Berlino, 2002, p. 15.

La ricorsività, a partire dalla prospettiva *generativa* del linguaggio, è un aspetto cruciale del linguaggio. Secondo Chomsky essa si concretizza nella possibilità di includere una frase sempre entro un'altra e via via estenderla infinitamente. I LLMs invece si concentrano sulla ricerca del *token* successivo. *«Ogni atto del pensiero e ogni espressione linguistica sono la manifestazione di questa profonda capacità ricorsiva, che non solo governa l'espressione, ma la rende anche possibile. La sintassi è la logica trascendentale delle scienze cognitive, e la sua colonna portante è la ricorsività»*²⁰⁹. Ad esempio, il programma minimalista enfatizza che la ricorsività è un meccanismo che sostiene la funzione di fusione [*merge*] del linguaggio, ossia la maniera universale, anche se aperta a molteplici sintassi che descrivono le svariate lingue, di combinare le parole assieme. La capacità di fondere assieme diverse parole descrive la specifica profondità del linguaggio e il suo aspetto “trascendentale”, nel senso di un codice di regole linguistiche (linguaggio internalistico) *«viene a configurarsi come una gerarchia di valori e parametri che plasma ogni oggetto che prende in considerazione»*²¹⁰. La “generatività sintattica” del linguaggio, per come la denota Kockelman²¹¹, può essere estesa a una più generale “generatività sistematica”, considerata come metodo costruttivo di modelli a partire da un insieme finito di regole (possano queste essere specificatamente linguistiche oppure architettoniche, scultoree, di programmi informatici ecc.).

È possibile per Chomsky sviluppare una linguistica generativa che possa venire scientificamente studiata solo studiando un processo di formazione del linguaggio puro, distaccato da ogni sua esemplificazione culturale. Così si arriva alla celebre distinzione tra “linguaggio externalistico” come contingenti enunciazioni e “linguaggio internalistico” come connessioni neurologiche interne; il nesso è possibile solamente per mezzo di quello che Jean-Jacques Leclerc definisce come un principio di riduzionismo tra comunicazione linguistica e connessione bio-fisica nel cervello²¹².

La dicotomia è però da considerarsi surrettizia e mezzo concettuale insufficiente per spiegare la profonda natura del linguaggio. Come presupposto duale sembrerebbe seguire troppo fedelmente l'astratta triade di Charles Morris di “semantica-sintassi-pragmatica”. I tre aspetti non sono che mescolati, organizzati e dissociarli può essere un'operazione analitica profittevole, quanto un'insufficiente e fuorviante mossa esplicativa. In particolare, si può provare a spiegare la sintassi come un processo universale di formazione linguistiche, ma solo attraverso appositi esempi culturali. Anche

²⁰⁹ Weatherby, Leif, *Language machines*, Minnesota University Press: Minneapolis, 2025, p. 45.

²¹⁰ *Ibid.* p. 48.

²¹¹ Chomsky, Noam, op. cit., p. 34.

²¹² Kockelman, Paul, *Last words*, Pickly Paradigmatic Press: Chicago, 2024, p. 63.

le fasi non grammaticali di Chomsky, che vogliono esemplificare la tendenza generativa innata del linguaggio, non possono che essere comprese a partire dal contesto culturale in cui gli stessi termini vengono usati. Insomma, il mezzo esemplificativo è esso stesso irrimediabilmente un prodotto culturale. Provare ad avanzare prove e controprove per una supposta scientificità di ricerca nei confronti del linguaggio, recidendo ogni riferimento alla cultura risulta impossibile e debilitante per l'argomentazione.

Con la generatività sistematico-sintattica si intreccia allora la “generatività pragmatica”²¹³, da non confondere con la pragmatica di Morris, ma maggiormente affine all'aspetto semantico, di formulazione di nuovi *tokens*; si tratta della capacità non solo di proferire un'infinità di tipi formali, ma anche di usare uno stesso tipo formale di «*in un'infinità di contesti distinti*»²¹⁴. La generatività pragmatica si distingue, dunque, dalla generatività sintattica per il fatto di concernere non la mera potenza combinatoria del lessico, ma la sua attualizzazione entro un determinato contesto. Sotto questo profilo i LLMs rappresentano uno strumento altrettanto strabiliante, componendo inattese combinazioni tra morfemi, parole, frasi, discorsi. Il motivo che ne permette un'inesausta attività semantica è da ricercare nella ricorsività tra generi e specie, parti e totalità che compongono un linguaggio. Generi e specie non indicano solo insiemi di elementi legati da un'identità di contenuto (l'insieme di mammiferi condivide una precisa modalità di allattamento, dunque determinati organi), ma anche di identità di forma, ovvero l'insieme di morfemi si contraddistingue per la loro irriducibile, minima unità significativa, differente di elemento in elemento).

Parlando di “generi e specie” Kockelman fa riferimento alla ripartizione secondo non modalità, ma di elementi secondo molteplici scansioni, livelli linguistici (dal morfema al testo compiuto). Quindi non solo vengono usate delle forme categoriali che si possono spiegare tramite la ricorsività del linguaggio, ma si generano dei *token* a partire dagli stessi pattern che ne codificano la sintassi. «*In altre parole, ChatGPT è incredibilmente bravo non solo a identificare tipi e modelli, più in generale, a tutti i livelli della struttura linguistica, testuale e interattiva; è anche incredibilmente bravo a produrre nuovi elementi di tali tipi, oltre che a creare nuovi tipi in sé*»²¹⁵. Come si è visto sopra, è la funzione poetica fondamentale del linguaggio a permetterne la continua ricombinazione di elementi. Conforme ai principi della cibernetica e alla teoria dell'informazione però, la funzione poetica di Jakobson può venire assimilata a una

²¹³ Leclerc, Jean-Jacques, *Deleuze and language*, Palgrave Macmillan: Londra, 2002, p. 85.

²¹⁴ Kockelman, Paul, op. cit., p. 64.

²¹⁵ *Ibid.* p. 62.

“generatività stocastica” del linguaggio. È qui però che la critica dei *stochastic parrots* comincia a vacillare: i modelli linguistici non si limitano a selezionare campioni dalle distribuzioni di probabilità: precisamente, sono anche responsabili della creazione delle stesse distribuzioni. Infatti, partendo da una sequenza di parole, viene generata una distribuzione di probabilità relativa alle possibili parole successive, dalla quale viene poi effettuata una selezione; la parola scelta viene aggiunta alla sequenza e la procedura ricomincia nuovamente.

I chatbot dei Large Language Models (LLM) non funzionano semanticamente, cioè non capiscono il significato di ciò che sintetizzano come testo. Lavorano sulla base del riconoscimento dei modelli e della produzione dei modelli. Per fare questo, hanno bisogno di miliardi di documenti forniti durante la fase di *training*. Nel caso dei LLM, gli elementi costitutivi sono sequenze di lettere (token) che non possono essere intese come unità significative e possono quindi essere tradotti in valori numerici. Il programma calcola la posizione di questi token nel contesto di tutti gli altri, per cui la vicinanza e le distanze tra token indicano la probabilità che una particolare parola segua la precedente. Gli LLM sono quindi caratterizzati da superfici di set di dati strutturate in token e quindi leggibili dalla macchina. Gli algoritmi statisticamente funzionanti possono solo generare risultati probabili. L’obiettivo è calcolare la distribuzione degli elementi nelle popolazioni, con i fattori spaziali (vicinanza e distanze tra i punti) che svolgono un ruolo decisivo. Il token più probabile è selezionato. Il risultato scritto è una catena di token che viene compilata sulla base della combinazione statisticamente più probabile. Questa è una selezione e ricombinazione di elementi di conoscenza già realizzati, non la creazione di nuove conoscenze.²¹⁶

Solo tenendo conto della ricorsività, quindi della continua modulazione dei parametri che rendono possibili le operazioni *transformer*, della *backpropagation*, ci si rende conto che l’insieme stesso, la supposta “struttura” viene alterata continuamente e non può venire inquadrata semplicemente come una stabile e unitaria figura.

C’è però un terzo aspetto che non si può tralasciare e che Kockelman nomina come “generatività dinamica”, affine alla “pragmatica” di Morris. Molti autori francesi, come Foucault o Deleuze, si sono ispirati a Nietzsche per chiarire la re-iterabilità e possibile ripetizione dei valori, chiarendone la natura di mezzo più che di fine. Un valore può essere considerato trascendente qualora sia elevato al rango dei fini di un’azione, ma

²¹⁶ *Ibid.* p. 63.

tramite delle operazioni di analisi genealogica, come in *Genealogia della morale*²¹⁷ o *La nascita della clinica*²¹⁸ si dimostra come lo stesso mezzo può venire tradotto per un ulteriore fine.

Altro aspetto da rimarcare, ogni valore non può essere descritto come mera rappresentazione eterea, ma viene incarnato già entro una certa società, creando delle vere e proprie linee di forza. La dimensione sociale dell'universo valoriale, che si potrebbe tradurre come "intelletto generale" della società fu affrontata già da Marx. L'insieme, l'"assemblaggio" di mezzi tecnici, di trasmissione dell'informazione, che formano la sfera della produzione delle merci è ciò che costituisce la struttura marxiana della società. Invece che intenderla come compatto monolite, la struttura economica è passibile di trasformazioni (rottture) che istituirebbero nuove generalità, facendo così dell'AI uno dei fattori più rilevanti ad oggi per contribuire a una nuova creazione strutturale. È riduttivo quindi descrivere il recente sviluppo dei LLMs come una rivoluzione "epistemologica" senza tenere in considerazione gli effetti sociali e pragmatici.

Ogni descrizione sintattica del linguaggio ha già immediatamente una valenza pragmatica. Il linguaggio, seguendo Shannon con il concetto di "ridondanza", Jakobson con il concetto di "funzione poetica", non è prevalentemente informativo, ma prima di tutto comunicativo. Non si comunicano informazioni però, ma si impartiscono sempre ordini²¹⁹. L'unità minima pragmatica non può essere il *bit*, ma la parola d'ordine [*slogan*]. Leclerc dimostra che il tema del linguaggio attraversa in varie modalità il pensiero di Deleuze, dagli iniziali anni '60, in cui la riflessione si concentra a muovere una critica alle antinomie dello strutturalismo, poi negli anni '70 in cui la collaborazione con Guattari mostra l'intimo rapporto che si crea tra desiderio, sapere e verità nell'*Anti-Edipo* (1972), grazie alla concezione della "macchina", arrivando poi negli anni '80 a maturare una linguistica macchinica degli assemblaggi. Il "*New pragmatism*"²²⁰ che si afferma con *Mille Piani* offre uno strumento critico fondamentale per illuminare la funzione assertiva e dominante del linguaggio al giorno d'oggi. Critici come Maurizio Lazzarato²²¹ sottolineano quanto l'innovazione apportata dal concetto di assemblaggio sia impareggiabile e imprescindibile per ogni teoria critica che voglia imporsi al giorno d'oggi. Il merito non va attribuito però esclusivamente a Deleuze, ricordando come fa

²¹⁷ Szepanski, Achim, *In the delirium of simulacrum*, Becoming press: Berlino, 2025, p. 228.

²¹⁸ Nietzsche, Friedrich, *La genealogia della morale*, Adelphi: Milano, 1992.

²¹⁹ Foucault, Michel, *La nascita della clinica*, Einaudi: Torino, 1998.

²²⁰ «Come *performance* di ogni linguaggio, la lingua non è né reazionaria né progressista; essa è semplicemente fascista; il fascismo, infatti, non è impedire di dire, ma obbligare a dire» cfr. Barthes, Roland, *Lezione*, Einaudi: Torino, 2001.

²²¹ Leclerc, Jean-Jacques, op. cit.

Gary Genosko²²² che il tema delle semiotiche asignificanti, su cui torna anche Lazzarato, assieme a quello della “macchina”. In opposizione critica all’inconscio macchinico lacaniano, sono partoriti dalla predisposizione politico-psicanalitica di Guattari.

1.1 VIRTUALE E POSSIBILE

Prima di addentrarsi nelle tematiche centrali della pragmatica di *Mille piani* per mostrare come i LLMs possano incidere sulla generatività dinamica atta a generare un nuovo impianto sociale, è bene però rispondere alla domanda da cui si era partiti, ovvero: è possibile concepire i LLMs come macchine creatrici? La creazione è un tema affrontato principalmente da Deleuze e Guattari in *Che cos’è la filosofia?*²²³, ma che già compare non in maniera così esplicita in *Differenza e ripetizione*. Qui si prende a riferimento un termine centrale per muovere una critica allo strutturalismo che ancora oggi è al centro dei dibattiti sul funzionamento dei LLMs: *il virtuale*. Esso gioca da anello di congiunzione tra senso, che si è chiarito sopra²²⁴ tramite il riferimento ad Althusser, e molteplicità differenziali, sulla scorta dell’ipotesi differenziale di Saussure, poi usata per giustificare la centralità della linguistica di Jakobson per lo strutturalismo.

	1	2	3	4	
...	the	boy	came	home	
	one	girl	went	down	
	no	man	ran	in	
	.	.	.	.	
	.	.	.	.	
	.	.	.	.	

Figura 9 struttura del linguaggio per Hjelmslev²²⁵

Se da un lato si ha la struttura che esibisce un suo andamento sincronico solo attraverso l’ordine e il rapporto dei suoi elementi simbolici nel tempo (da cui promana il senso) e dall’altro i loro termini si dicono differenziali, condizioni, ma al contempo condizionati, perciò prodotti, allora cosa garantisce che la struttura (che si è visto coincidere con la

²²² Lazzarato, Maurizio *Macchine e segni*, Ombre corte: Verona, 2012.

²²³ Genosko, Gary, *Félix Guattari: An Abberant Introduction*, Continuum: Londra, 2002.

²²⁴ Deleuze, Gilles, Guattari, Félix, *Che cos’è la filosofia?*, Einaudi: Torino, 2002.

²²⁵ Cfr. parte 1, cap. 2.0.

stessa genesi storica) e le sue varie trasformazioni possano dirsi entro un'unica unità sistematica? Precisamente il rapporto tra virtualità strutturale e sotto-strutture attuali.

[...] lo "strutturalismo" ci sembra il solo mezzo con cui un metodo genetico può realizzare le proprie ambizioni. Basta comprendere che la genesi non va da un termine attuale, per piccolo che sia, a un altro termine attuale nel tempo, ma dal virtuale alla sua attualizzazione, ossia dalla struttura alla sua incarnazione, dalle condizioni di problemi ai casi di soluzione, dagli elementi differenziali e dai loro nessi ideali ai termini attuali e alle relazioni reali diverse che costituiscono ad ogni istante l'attualità del tempo.²²⁶

La struttura, all'altezza del pensiero deleuziano degli anni '60 incarna l'esempio di un essere univoco, ma che si manifesta secondo diverse maniere (come l'essere in Aristotele, a titolo indicativo). Il virtuale è la chiave d'accesso al piano dell'essere che procede per eventi, quindi per rotture e perturbamenti; la sua opposizione alla molteplicità di regimi attuali consente di spiegare come «*generare il discontinuo a partire dal continuo*»²²⁷ significa studiare le soluzioni a partire dalle specifiche condizioni del problema, quindi dalla precisa disposizione degli elementi. Il concetto ha una discendenza matematica²²⁸ atta a definire il potenziale di forze per plasmare le diverse occorrenze strutturali (nel caso dei LLMs si possono pensare come gli *output*). Il virtuale, infatti, consente di guardare alla struttura come un sistema generante, in continuo divenire. Non è importante chiedersi cosa significa qualcosa, ma studiarne il processo di formazione. Lo strutturalismo per Althusser è lo studio delle condizioni di pratiche, mentre per Deleuze si può concepire come morfogenesi, in contrapposizione con la visione ilemorfica tradizionale

Il virtuale al contrario non deve realizzarsi, ma attualizzarsi; e le regole dell'attualizzazione non sono più la somiglianza e la limitazione, ma quelle della differenza o della divergenza e della creazione. Quando certi biologi chiamano in causa la nozione di virtualità o di potenzialità organica, ritenendo tuttavia che quest'ultima si attualizzi attraverso una semplice limitazione della sua capacità globale, cadono chiaramente nella confusione tra virtuale e possibile. Il virtuale infatti per attualizzarsi non deve procedere per eliminazione o limitazione, ma deve *creare*, con degli atti positivi,

²²⁶ Immagine da Hjelmslev, Louis, *The basic structure of language*, in *Essais linguistiques II*. Nordisk Sprog- og Kulturforlag, 1973, p. 139.

²²⁷ Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, p. 238.

²²⁸ *Ibid.* p. 20.

le sue linee di attualizzazione. E questo per una ragione molto semplice: mentre il reale si realizza a immagine e somiglianza del possibile, l'attuale, al contrario *non* somiglia alla virtualità che incarna. La differenza è l'elemento principale del processo di attualizzazione.²²⁹

Questa una delle prime formulazioni del virtuale, che permette di superare le limitazioni del possibile. Ancora in *Differenza e ripetizione* ci ritorna, specificando l'apparenza di un divenire delle cose secondo la loro somiglianza.

La differenza e la ripetizione nel virtuale fondano il movimento dell'attualizzazione, della differenziazione come creazione, sostituendosi così all'identità e alla somiglianza del possibile, che suscitano soltanto uno pseudo-movimento, il falso movimento della realizzazione come limitazione astratta.²³⁰

L'attualizzazione è un processo creativo, di pura affermazione della molteplicità della struttura. La nozione di virtuale connota un movimento intensivo di intensità preindividuali e impersonali. Ogni elemento virtuale non è irreali, fantasioso, esso possiede lo stesso grado di realtà del fenomeno certo ed evidente che viene percepito; la peculiarità del virtuale è di non essere attuale. Per spiegare questa sottile differenza, Deleuze si richiama a Proust, asserendo che il virtuale è «*reale senza essere attuale, ideale senza essere astratto*»²³¹. Dunque, la coppia reale-possibile si sostituisce quella di attuale-virtuale. L'esperienza degli oggetti è finita non solo perché apparentemente è promossa da un soggetto finito (un "Esserci", un "Ego trascendentale"), ma perché è una parte di un complesso di relazioni che legano l'oggetto virtualmente alla struttura. La realtà non è quindi solo ciò che appare, ma le stesse condizioni di determinazione virtuale dell'oggetto. Dire che un oggetto è condizionato significa che è posizionato secondo un ruolo, dunque in accordo con gli elementi, i punti singolari e i rapporti differenziali della struttura, che non si può indicare allora come un dominio monolitico, ma un'intersezione di sotto-strutture virtuali in rapporto tra loro. È interessante però che Deleuze in *Differenza e ripetizione* non riporti tra gli esempi di struttura, alcuno affine alla linguistica, ma tutti connotati entro ambiti biologici, sociali o fisici, ma nell'opuscolo sullo strutturalismo, parlando della doppia natura virtuale-attuale l'esempio per definire

²²⁹ In particolare, il concetto va attribuito alla natura di "spazio liscio" di Riemann come opposto alla metrica statica degli "spazi discreti". Cfr. Duffy, Simon, *Deleuze and the history of mathematics*, Bloomsbury: London, 2013.

²³⁰ Deleuze, Gilles, *Marcel Proust e i segni*, Einaudi: Torino, 1967, p. 87.

²³¹ Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, p. 275.

le specie e le parti che vengono ad attualizzarsi ricalca rispettivamente le lingue e le strutture linguistiche²³². Qui ancora sembrerebbe esserci una dicotomia di tipo saussuriano tra *langue*, come codice linguistico virtuale e perciò differenziale, e *parole* nella sua variabilità che si esprime solo nello stato attuale dell'enunciazione.

Il virtuale è la riserva inesaurita di potenziale che permette l'affioramento di strutture attuali, secondo una libera creazione, ossia priva di qualsiasi regolamentazione o costrizione di sorta. Ora, come potrebbe l'AI considerarsi creativa? I suoi *response* dovrebbero essere «molto generali» e dovrebbe essere a sua volta in grado di generalizzare; «*i compiti che può svolgere dovrebbero trascendere non solo l'argomento, la modalità, il dominio e la funzione, ma anche il luogo e il tempo, e persino i mondi possibili e i futuri immaginabili*»²³³. Ciò rimane ancora distante dalla realtà dell'AI odierno, essendo la funzione che si ascrive all'AGI (*Artificial General Intelligence*). Se ad esempio si considera ChatGPT la sua capacità di calcolo non può essere confusa con una forma autentica di creazione; *non c'è creazione, ma modulazione continua, per via della generatività stocastica*. In altre parole, è difficile che vengano prese in considerazione come risposte, alcune con bassa probabilità di essere additate (anche se potrebbero essere considerate rilevanti per i fini della richiesta del *prompt*).

La generatività stocastica, come funzione computazionale *transformer* di pesi e parametri della macchina, non è assimilabile a una creazione sempre nuova e inattesa. Essa ripartisce le varie soluzioni secondo un calcolo, anche se ciò rimane “in profondità” e tendenzialmente non viene percepito dall'utente. Soprattutto, non è casuale. L'ipotesi dei *stochastic parrots* viene ancora disattesa per due motivi, rispettivamente riflettenti l'aspetto del linguaggio umano e l'aspetto del funzionamento macchinico. Il linguaggio umano si è dimostrato grazie a Jakobson che è fondamentalmente ridondante, la cui funzione poetica di ricombinazione di pattern linguistici spicca su quella d'informazione. Deleuze sarebbe d'accordo sotto questo aspetto, definendo il linguaggio come digitale, denotativo o di codice²³⁴, proprio in virtù delle permutazioni possibili tra morfemi. Insomma, sono innanzitutto gli esseri umani ad essere simili a dei pappagalli, prima ancora che qualche macchina che si vorrebbe criticare per lo stesso motivo. Inoltre, come dimostra Monod, «*il puro caso, soltanto il caso, libertà assoluta ma cieca, è alla radice stessa del prodigioso edificio dell'evoluzione*»²³⁵, ossia larga parte dei processi fisici che

²³² *Ibid.* p. 270. In realtà la stessa citazione è usata in svariate interviste e libri, come l'opuscolo sullo strutturalismo a cui ci riferiamo.

²³³ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo*, SE: Milano, 2004, p. 29.

²³⁴ Kockelman, Paul, *Last Words*, Pickly Paradigmatic Press: Chicago, 2024, p. 65.

²³⁵ Deleuze, Gilles, *Sulla pittura*, Einaudi: Torino, 2024, pp. 119-149.

attraversano il mondo sono basati sul caso della contingenza. Nonostante l'imprevedibilità dei *prompt* sembra possa costituire un argomento a favore della casualità operativa dei LLMs, in realtà bisogna prendere in considerazione il loro funzionamento, che sembrerebbe operare qualcosa di ancora più sofisticato.

Il loro processo non è meramente induttivo, ma abduttivo²³⁶ nella misura in cui i parametri e i pesi vengono continuamente rimodellati; trasformando i loro parametri, a loro volta verranno ad analizzare i futuri input sulla base di questa rinnovata modulazione.

Le macchine, inserito un nuovo pezzo di testo o immagine o video, analizzerebbero tali *prompt*, andando a ri-modulare tutti i *token* che ne prendono parte, operando una sintesi rispetto ai dati degli stessi token accumulati nel passato e quelli forniti ora dal prompt presente e configurando nuovamente i dati, ovvero assegnando nuovo valore alla probabilità di come un token si lega con il contesto in cui è presente. Rimodellando i pesi di volta in volta però si nota che l'operazione non è meramente probabilistica, ma «è essa stessa fondata nella generatività sintattica e pragmatica degli umani che hanno prodotto i testi su cui i modelli sono stati addestrati»²³⁷. La ricorsività macchinica si può definire allora “parassitaria”: un LLM genera distribuzioni di probabilità solo perché tali distribuzioni sono state precedentemente prodotte (nelle loro forme sintattiche, semantiche e pragmatiche) da agenti umani il cui linguaggio è generante. Il ragionamento appare in linea con l'intenzionalità parassitaria delle macchine, indicata nel capitolo precedente²³⁸: i LLMs registrano e memorizzano le espressioni prodotte da individui che intenzionalmente tendono a oggetti a loro esterni e

²³⁶ Monod, Jacques, *Il caso e la necessità*, Mondadori: Milano, 2022, p. 115.

²³⁷ Su questo punto le interpretazioni divergono. Nella filosofia di Peirce, l'abduzione è la più creativa delle tre forme di inferenza: a differenza della deduzione, che trae le conseguenze necessarie di una regola (*top-down*), e dell'induzione, che generalizza da un campione (*bottom-up*), l'abduzione è il movimento con cui, di fronte a un fatto sorprendente, si formula l'ipotesi che lo renderebbe spiegabile. Umberto Eco ha formulato quattro distinte tipologie di abduzione: l'abduzione *ipercodificata*, in cui cioè si è portati quasi automaticamente a scegliere una regola; l'abduzione *ipocodificata* in cui la regola viene selezionata entro una rosa di candidati più o meno equiprobabili; l'abduzione *creativa* che produce una *nuova* regola di spiegazione; la *meta-abduzione* che costituisce la più plausibile spiegazione per come intraprendere la scelta rispetto a un primo livello di abduzioni possibili. Cfr. Bellucci, Francesco, *Eco and Peirce on Abduction*, in «European Journal of Pragmatism and American Philosophy», X-1, 2018, §29, riferendosi a Eco, Umberto, *I limiti dell'interpretazione*, Bompiani: Milano, 1990, pp. 206-207.

L'attuale dibattito sugli LLM si colloca tra i primi due modelli di abduzione. Mentre Kockelman sembrerebbe attestarsi più vicino al primo prototipo, ritenendoli quindi capaci di formulare nuove spiegazioni per via della ri-modulazione dei pesi, Pasquinelli (cfr. Pasquinelli, Matteo, *The Eye of the Master. A Social History of Artificial Intelligence*, Verso: London, 2023, capp. 7-8) riterrebbe la loro incapacità esplicativa dovuta a un fattore di interpretazione fraudolenta e retroattiva. Una volta che si vede la macchina confezionare una risposta plausibile, si è portati ad attribuirle un'intenzionalità, di cui, come anche Kockelman sottolinea, risulterebbe priva. La generatività stocastica è una pura interpolazione probabilistica, priva di qualsiasi ipotesi fondativa di ragionamento. Sicuramente si può affermare che gli LLM rimangono fuori sia dalla terza che dalla quarta tipologia.

²³⁸ Kockelman, Paul, op. cit., p. 67.

i parametri delle macchine sono collegati ai valori delle persone e, di conseguenza, allineati ad essi. Il processo di generatività stocastica è derivato e non può essere pensato in termini di creazione *ex nihilo*; è insufficiente a trasformare materia prima eterogenea in pattern simbolici come fa l'intelligenza ibrida descritta da McCulloch e Pitts. La sua rielaborazione stocastica si appoggia inevitabilmente su ciò che è già stato codificato linguisticamente dai parlanti. Così Kockelman arriva a scovare la capacità generativa di queste macchine:

Ma il loro caratteristico modo di generatività può essere inquadrato in un altro modo: partendo da una certa quantità di qualcosa (come un testo, o qualsiasi cosa che possa essere codificata come testo, ovvero qualsiasi cosa che possa essere digitalizzata), ne producono ancora di più, pur mantenendo i pattern essenziali, e quindi la loro natura sottostante di ciò che è venuto prima: non solo l'enunciato successivo date le enunciazioni precedenti o l'azione successiva date le azioni precedenti, ma anche il prezzo successivo dati i prezzi precedenti, la scena successiva date le scene precedenti, il pensiero successivo dati i pensieri precedenti, l'entità successiva date le entità precedenti e l'evento successivo dati gli eventi precedenti. In breve, generano il futuro dato il passato.²³⁹

Questa formula, “*make more given some*”, condensa con esattezza la tipologia di generatività dei modelli linguistici: non un divenire creatore, ma una produzione che, per quanto raffinata, è strutturalmente in simbiosi con il testo precedente, passato. È a questo punto che il dialogo con Weatherby diventa cruciale. Quando Weatherby afferma che «*la riproduzione non può mai essere una copia*»²⁴⁰ e che è in questo non-essere-copia che risiede la generatività dei LLMs, egli sta in realtà descrivendo, in altri termini, il funzionamento della generatività stocastica: ad ogni iterazione, la sequenza generata sarà diversa da ogni altra sequenza precedente, eppure essa rimane “*in keeping with prior text*”²⁴¹, cioè «mantiene intatto l'essenziale schema»²⁴² di ciò che l'ha preceduta.

A rafforzare la tesi sulla generatività stocastica di Kockelman è Giuseppe de Ruvo, con la sua analisi di *Differenza e ripetizione* e delle tre tipologie di temporalità: Habitus, Mnemosine ed Eros. Deleuze fa riferimento a queste tre dimensioni esclusivamente nell'opera del 1968 e se ne serve per giustificare ontologicamente lo sviluppo “orizzontale”, indeterminato e aperto del simulacro, ovvero il suo eterno ritorno.

²³⁹ Cfr. parte 2, cap. 1.0.

²⁴⁰ *Ibid.* p. 71.

²⁴¹ Weatherby, Leif, *Language machine*, Minnesota University Press: Minneapolis, 2025, p. 205.

²⁴² Kockelman, Paul, op. cit., p. 69.

Ora, “generare il futuro dal passato” è un’operazione meramente probabilistica, che nulla ha in comune con l’accesso a un passato puro, per come Deleuze ne parla nei termini di Mnemosine. L’architettura transformer è piuttosto la realizzazione tecnica più rigorosa che si possa concepire della sintesi di Habitus²⁴³, ossia la prima sintesi passiva del tempo, vista come sua fondazione; «*pur essendo costituente, non è per questo attiva, non è fatta dallo spirito, ma si fa nello spirito che contempla*»²⁴⁴. Essa, contraendo gli istanti di tempo gli uni negli altri forma il presente vissuto; in altre parole, in una serie AB, AB, AB, A..., «io mi aspetto la comparsa di B»²⁴⁵. È il medesimo procedimento con cui i LLMs funzionano, aspettandosi attraverso il calcolo probabilistico delle connessioni tra termini il token successivo.

La generatività stocastica implica non solo il campionamento da una distribuzione di probabilità (cosa semplice come tirare un dado a molte facce, distorto), ma anche la creazione ricorsiva della distribuzione di probabilità da cui campionare (e quindi la modellatura e la distorsione del dado che viene tirato). Per fare questo bene, nel caso dei modelli linguistici, occorrono miliardi di parametri, strati di transformer, montagne di lavoro, oceani di addestramento, decenni di ricerca e cumuli di testo.²⁴⁶

Tecnicamente, il meccanismo di *attention* è un dispositivo di contrazione di istanti indipendenti, ossia i token del contesto, in un “presente vivente sintetico”, da cui si genera il token successivo. Però la fondazione del tempo va distinta dal fondamento della Memoria. Il presente vissuto è come «*un suolo mobile occupato dal presente che passa*» che rinvia necessariamente a un altro tempo, quello passato. Mnemosine è «*la sintesi fondamentale del tempo, che costituisce l’essere del passato (ciò che fa passare il presente)*»²⁴⁷. L’architettura *transformer* non è che un’operazione fondazionale, ma priva del fondamento passato puro e intensivo, contraendo in maniera contigua pattern. I suoi pesi non sono Mnemosine: sono il sedimento statistico di un Habitus computazionale, ovvero si rigenerano costantemente.

Quindi è giusto appoggiarsi alla formulazione di Kockelman di una “generatività stocastica”, che è altro, se non l’opposto, di ciò che Deleuze definisce come virtuale. Si cadrebbe in un fraintendimento tecnico e ontologico a insignorare lo spazio

²⁴³ *Ibid.* p. 71.

²⁴⁴ Cfr. De Ruvo, Giuseppe, *Rimettere il fantasma nella macchina. Una critica deleuziana dell’ontologia digitale e della governabilità algoritmica*, La Deleuziana n. 14, 2022.

²⁴⁵ Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, p. 96.

²⁴⁶ *Ibid.*, p. 95.

²⁴⁷ Kockelman, Paul, op. cit., pp. 68-69.

“virtuale” dei LLMs (gli embedding, lo spazio latente, la vettorializzazione dei token) come virtuale deleuziano, non è niente più che una omonimia. Il virtuale dei LLMs non è una serie di attualità precompresse, in cui ogni token è la distribuzione statistica delle sue connessioni con gli altri; di fatto non si oppone per niente all’attuale. Il virtuale, si è visto sopra, garantisce l’univocità dell’essere, ossia il suo stesso modo dirsi Essere per tutte le differenze che mantengono però un loro senso precipuo

È proprio dell’essenza dell’essere univoco riferirsi a differenze individuanti, ma queste differenze non hanno la stessa essenza e non variano l’essenza dell’essere – come il bianco si riferisce a intensità diverse, ma resta essenzialmente lo stesso bianco.²⁴⁸

Non è così per lo spazio latente, in cui ogni differenza generale è computabile perché inscrivibile in una metrica già data in partenza.

L’ipotesi avanzata da De Ruvo, che lo spazio latente sia piuttosto una matrice dei possibili, sembrerebbe indicare il giusto posizionamento del problema invece. Il possibile classicamente è inteso come modalità logica ad immagine del reale, sottoposto a una sua imitazione che solo accidentalmente può mutare. Il calcolo del token successivo procede per limitazione: data una distribuzione di probabilità, vengono privilegiati gli scenari più conformi a una norma statistica, mentre gli scenari rari, anche se «potrebbero essere considerati rilevanti per i fini della richiesta del prompt», vengono via via eliminati attraverso operazioni come softmax, top-k e top-p. Ogni *response* è già contenuto potenzialmente nella distribuzione di probabilità che il modello aveva adottato a partire dai valori dei pesi; non si creano *biforcazioni*, ma si ripercorrono gli scenari più conformi alla norma statistica. Il suo, quindi, non è un autentico, creativo divenire. L’autentica creazione è gravida di trasformazioni, non limitate alla rosa di possibilità concordate con il concetto.

Quando il processo di attualizzazione viene definito una «*creazione di linee divergenti*»²⁴⁹ si intende rimarcare la differenza tra la determinazione completa delle condizioni di un problema e l’attualizzazione delle sue soluzioni, la “rottura epistemologica” tra Generalità; Deleuze parla di rapporti differenziali che corrispondono alle condizioni d’esistenza sancite dai punti di singolarità, ma non gli assomigliano. «*Il possibile e il virtuale si distinguono nella misura in cui l’uno rinvia alla forma*

²⁴⁸ Deleuze, Gilles, op. cit., p. 108.

²⁴⁹ *Ibid.*, p. 53.

dell'identità nel concetto, mentre l'altro designa la pura molteplicità»²⁵⁰. Ogni volta che un *response* viene inviato propriamente la macchina sta creando, ma sta rimodulando, entro un ventaglio di possibili modellati, l'opzione più adeguata. Non si tratta allora del movimento creativo di attualizzazione del virtuale. La "creatività" può essere concepita come geometria di un possibile, non come evento del virtuale.

L'evento è tipico della terza sintesi temporale, quella di Eros²⁵¹, «compagno di Mnemosine»²⁵² e colui che costituisce una riserva infinita di ripiegamenti. L'evento ripiega l'ordine del tempo, infatti, segnando una cesura tra passato lineare e avvenire. Il suo è il tempo dell'Aion che ricompone il passato e il futuro nella loro infinite direzioni, è il "sempre già"; *«un ripiegamento della sintesi di Mnemosine, un passaggio al limite della e nell'immanenza della memoria che ci permette di «penetrare nel passato puro in sé»*²⁵³. Ogni calcolo algoritmico consiste piuttosto in un'estensione di token rispetto alla distribuzione statistica in cui si ritrovano. Ogni evento crea un'assenza nello stato attuale delle cose (non un negativo per Deleuze).

L'oggetto è completo, ma non intero, perché è scisso nelle due immagini (virtuale e attuale). L'immagine virtuale, che modella quella attuale, deriva dalla differenziazione dell'evento. Perciò solo con un evento è possibile che un oggetto virtuale si dica "mancante" perché *«non è là dove [lo] si trova se non a condizione di essere cercato dove non è. Non è posseduto da coloro che l'hanno, ma è posseduto da chi non lo possiede»*²⁵⁴. Come l'oggetto *a* di Lacan, l'oggetto virtuale ha la proprietà di essere e non essere là dove si trova: è un fantasma. Gli LLMs sono invece addestrati per eliminare ogni differenziazione evenemenziale, detta "punto di biforcazione della catastrofe" attraverso RLHF, filtraggio dei bias oppure fine-tuning²⁵⁵; i vincoli di sicurezza, utilità e veridicità escludono la creatività dell'evento. Lo spazio latente è invece interamente digitalizzato e perciò commensurabile e modulabile algebricamente.

Inoltre, il passato di Mnemosine fa da fondamento, ma non fonda. Ci si deve richiamare, ma sfugge ad ogni attimo posizionabile nel tempo. Si registrano infatti due paradossi rispetto al passato:

- Il passato è contemporaneo al presente che è stato. Il passato è perciò

²⁵⁰ *Ibid.*, p. 274.

²⁵¹ *Ibid.*, p. 270.

²⁵² Il tempo "erotico" dell'evento chiaramente non coincide con l'ipotesi ergodica di Weatherby.

²⁵³ *Ibid.*, p. 134.

²⁵⁴ De Ruvo, Giuseppe, op. cit., p. 21.

²⁵⁵ Deleuze, Gilles, op. cit., p. 134.

contemporaneo a sé stesso solo se considerato come presente (paradosso della contemporaneità del presente e del passato);

- Se ogni passato è contemporaneo al presente che è stato, *tutto* il passato coesiste col nuovo presente rispetto al quale ora è passato (paradosso della coesistenza).

Il passato allora non si può propriamente dire che esiste, ma si deve dire che «insiste, consiste, è»²⁵⁶, essendo fondamento del presente. Non sarà rappresentabile perché non *fu* mai presente. Il passato dei LLM è sempre di tipo humaneo invece. Procede per accumulo di associazioni contigue: «*memoria lineare che riproduce, sulla base di principi associativi e di contiguità spazio-temporale, il passato nel presente*»²⁵⁷. I pesi di un modello, per quanto “contenitori” di un’enorme quantità di pattern, non sono passato puro: sono il presente contratto del corpus, la fondazione, non il fondamento. La ripetizione dei LLMs, l’iterazione del loro meccanismo di *attention* sembra invece prendere per differenza di grado ciò che è differenza di natura, potendola riassumere nel ragionamento logico-lineare che «se si danno le stesse circostanze, allora...»²⁵⁸.

La ripetizione dell’architettura *transformer* non segna sicuramente una creazione in senso stretto allora. Il fenomeno della ripetizione degli LLMs appare piuttosto «*in funzione di un piccolo numero di fattori selezionati*»²⁵⁹, costruendo così un codice, da contrapporre a un “linguaggio analogico”. Allora, quando Weatherby ritiene che la riproduzione non può mai essere una copia, sta usando una nozione debole di differenza. Una sequenza di *token* generata oggi da GPT-4 a partire dal medesimo prompt, è sicuramente diversa, in qualche misura, dal *response* passato; ma questa differenza è di possibilità entro la medesima distribuzione stocastica e non tra due attualizzazioni di uno stesso virtuale. La loro è una «differenza di grado»²⁶⁰ e non di natura, per dirla con Deleuze. La generatività degli LLMs non sembra poter essere ricondotta né al linguaggio di tipo chomskyano, né, sul piano metafisico, a quello che Deleuze descrive come attualizzazione del virtuale. È a questo incrocio che si situa il problema della generatività delle macchine linguistiche, problema che ora bisogna integrare con la proposta teorica di Weatherby di una “poetica generale”, dopo averne già illuminato i presupposti nel capitolo precedente.

²⁵⁶ Puntale è il richiamo a questo punto a Bernard Stiegler e al concetto di governamentalità algoritmica che «disintegra in anticipo ogni intermittenza, limitando l’impossibile» Cfr. Stiegler, Bernard, *La società automatica I. L’avvenire del lavoro*, Meltemi: Milano, 2019, p. 24.

²⁵⁷ Deleuze, Gilles, op. cit., p. 110.

²⁵⁸ De Ruvo, Giuseppe, op. cit., p. 24.

²⁵⁹ Deleuze, Gilles, op. cit., p. 10.

²⁶⁰

CAP. 2: POETICA GENERALE

2.0 UNA ZUPPA TIEPIDA

In questo paragrafo si prenderà in considerazione per larga parte il discorso teorico di Weatherby, cercando di mostrare come la sua proposta intenda definire i LLMs come eminentemente generativi, nel senso più profondo del termine. La sua poetica generale [*general poetics*] funge da griglia ermeneutica per analizzare «il paesaggio semiologico digitale»²⁶¹ che fa da sfondo al rapporto uomo-macchina. Riprendendo il filo del discorso, Weatherby pone un'ipotesi non meramente descrittiva, ma ritiene che la generazione linguistica delle macchine produca nuova cultura: le macchine linguistiche sono macchine culturali. Ciò gli ha permesso di dimostrare che l'ipotesi differenziale di Saussure troverebbe piena realizzazione tecnico-generativa nei LLMs. Si tratta ora di approfondire la poetica generale che permetterebbe all'autore un inquadramento privilegiato su come i LLMs mappano e ridistribuiscono il linguaggio.

Già nel capitolo precedente si è argomentato che lo statuto strutturale dei LLMs poggia su un'operazione costitutivamente non referenziale: il calcolo del token successivo non concerne il rapporto con un "oggetto" esterno, ma con la rete sintagmatico-paradigmatica delle altre parole. Weatherby radicalizza questa intuizione tramite l'introduzione della categoria di *poetica generale*, che, come si cercherà di argomentare, può essere suddivisa in tre tratti caratteristici:

- 1) La sua generalità concerne ogni regime di segni che attraversa il suo sistema e non solo le produzioni letterarie;
- 2) È poetica nel senso delineato da Jakobson: si concentra sul messaggio in quanto tale e non sull'adeguazione ad un referente;
- 3) È una *poetica* della generazione che spiega come non solo funziona il linguaggio, ma come si produca nuovo significato.

Anzitutto, si prenda a riferimento il modo in cui Weatherby descrive il funzionamento delle macchine linguistiche. Quando si inserisce un prompt nel modello, l'algoritmo non si limita a recuperare una sequenza pre-codificata, ma «*distilla aree a temperatura media usando la funzione poetica*». L'operazione di *softmax*, in particolare,

²⁶¹ *Ibid.*, p. 10.

ridistribuisce ogni volta i pesi dei tokens selezionati così da far emergere un sottinsieme di parole vicine entro lo spazio vettoriale; il risultato è ciò che Weatherby chiama “*pacchetto semantico*” [semantic package], ossia «*un insieme di possibilità che galleggiano alla superficie statistica*». Si noti che l’operazione di *softmax* non può definirsi come una semplice ricerca documentaria, poiché le macchine «*non si limitano ad “avere” una semantica preconfezionata; sono loro a confezionarla*»²⁶². La funzione poetica si evidenzia proprio in questa operazione di confezionamento: il ripiegamento del linguaggio su di sé, senza alcun rinvio ad un referente extra-linguistico, risulta essere un atto di espressione secondo regioni a maggiore probabilità di occorrenza, qualificandole come «*medie statisticamente catturate su un dato dominio di contesto*»²⁶³

Confezionare il significato è il modo in cui i LLMs riescono a raggiungere il significato. Distillano aree a temperatura media usando la funzione poetica. [...] Non si limitano ad “avere” una semantica preconfezionata; sono loro a confezionarla.²⁶⁴

Il confezionamento semantico è la «*prima operazione meccanica che siamo mai riusciti a osservare in grado di portare in superficie il rumore linguistico, conducendo verso il primo piano un linguaggio piatto ma autentico*»²⁶⁵. Quello che per Shannon sarebbe stato mero “rumore”, per i LLMs diventa un vero e proprio segnale, costruito su ridondanze, associazioni laterali o banalità linguistiche. Weatherby allora insegue anche una ridefinizione della nozione di rumore in Shannon. Nel modello classico della teoria della comunicazione di Shannon e Weaver il rumore è il termine negativo della trasmissione, ovvero «*un’informazione falsa e indesiderabile*»²⁶⁶ che deve essere espunta dal segnale in arrivo. Il rumore si calcola tramite il concetto di equivocazione, e si oppone a quello di ridondanza, in quanto la ridondanza (le ripetizioni sintattiche, le regolarità statistiche, le ricorrenze paradigmatiche) è precisamente ciò che permette di ridurre il rumore aumentando la prevedibilità del messaggio. La coppia rumore-ridondanza era, nello schema cibernetico, la dialettica fondamentale del canale: tanto più il canale è strutturato, tanto meno il messaggio può venire alterato dal disturbo.

La coppia risulta infelice per definire il funzionamento dei LLMs poiché il rumore non viene propriamente estirpato. È come se rumore e segnale si confondessero

²⁶² Weatherby, Leif, op. cit., p. 181.

²⁶³ *Ibid.* p. 177.

²⁶⁴ *Ibid.* p. 175.

²⁶⁵ *Ibid.* p. 177.

²⁶⁶ *Ibid.*

in una distribuzione statistica unidimensionale, atta ad appiattare messaggio-sfondo entro un'unica dimensione. Si prenda ad esempio una conversazione: se essa avviene su un canale di trasmissione neutro, il rumore rimane rumore, ossia fuori dal messaggio che viene veicolato. Invece i LLMs «*sembrano galleggiare in superficie qualcosa che non è esattamente articolazione: uno spettro di possibilità che capovolge il rapporto fra primo piano e sfondo, una sorta di segnale rumoroso – noisy signal*»²⁶⁷. La generatività dei modelli, in altri termini, fa emergere come segnale primario proprio ciò che, secondo la teoria dell'informazione classica, sarebbe stato relegato a rumore: la ridondanza statistica, la media topica, l'usuale, il già-detto, le associazioni laterali, ovvero quanto Weatherby chiama «*zuppa tiepida*» [tepid mush] «*un linguaggio tiepido che è comunque linguaggio – il quale, dopotutto, di solito non è né bollente né gelato*»²⁶⁸.

Conseguentemente, il messaggio non è più l'informazione singolare, ma ciò che è maggiormente si dice. L'informazione, quindi, è indirizzata entro il sentiero più battuto, più comune. È come se si stesse redigendo un'equazione in cui il messaggio diventa tanto più rilevante, quanto più viene discusso e dibattuto, statisticamente descritto.

Il segnale che diventa rumore viene indicato da Weatherby col termine “ideologia”, ma non nel senso con cui i teorici politici lo indicano; piuttosto indicando uno sfondo sempre più ramificato e distribuito in maniera omogenea per attenuare gli effetti della forma. Una volta che la forma viene “sbiadita”, il rumore viene a infondersi sempre più con il messaggio e il contenuto che si vuole veicolare. La macchina non simula il linguaggio umano dall'esterno: si sintonizza esattamente sul rumore di fondo del linguaggio umano e lo trasforma in segnale. La «*differenza fra la cosa più probabile da dire, i nostri topoi comuni, i luoghi condivisi dove il significato si muove più rapidamente attraverso le parole, ci nega il netto confine fra umano e macchina*»²⁶⁹. L'effetto di questa indistinzione rumore-segnaletta riflette come non ci sia un confine netto tra ridondanza umana e generatività linguistica delle macchine.

Con ciò si è quindi dimostrato come Weatherby comporti una rilettura importante, quasi un rovesciamento, dell'impianto classico della teoria dell'informazione²⁷⁰: la funzione poetica, come nei testi degli anni Cinquanta, non struttura più semplicemente la ridondanza per ridurre il rumore, ma ne automatizza

²⁶⁷ Shannon, Claude, Weaver, Nobert, *la teoria matematica delle comunicazioni*, ETAS Kompas: Milano, 1971, p. 38.

²⁶⁸ Weatherby, Leif, op. cit., p. 177.

²⁶⁹ *Ibid.* p. 178.

²⁷⁰ *Ibid.* p. 187.

l'estrazione, trasformando la ridondanza in oggetto principale di generazione. La trasformazione del rumore in segnale è, in questo senso, l'operazione fondamentale di sovra-codifica (surdeterminazione per dirla con Althusser) che i modelli compiono sul linguaggio collettivo: essi non emettono rumore al posto del segnale, come si potrebbe ingenuamente pensare di una macchina "cieca", intesa secondo la critica degli *stochastic parrots*; al contrario, fanno del rumore stesso, della sua ridondanza statistica, il loro segnale principale, e in ciò consiste tanto la loro sbalorditiva efficacia quanto il loro irriducibile limite.

2.1 INTRATTENIMENTO IDEOLOGICO

La parola "ideologia" è una di quelle che la filosofia contemporanea ha ereditato dal lessico ottocentesco senza riuscire mai a stabilizzarne il significato. Da Marx in poi essa è diventata un campo di battaglia in cui si gioca, ad ogni nuova generazione teorica, il rapporto fra pensiero e potere, fra rappresentazione e prassi, fra il dirsi del soggetto e l'ordine entro cui il soggetto è costituito. Infatti, il concetto di ideologia nel discorso di Weatherby di discosta dalla tradizionale visione che si ha di essa, liberandola dalla pretesa che essa designi una semplice falsa coscienza, un velo gettato dal soggetto sul mondo o, viceversa, un dispositivo trascendentale che dall'alto governi la totalità del rappresentabile.

È importante muoversi inizialmente verso un chiarimento di cosa sia l'ideologia per Althusser. Essa gioca un ruolo tanto centrale nella sua riflessione, quanto inedito rispetto alle precedenti formulazioni del concetto. La sua non è una credenza falsa, un'illusione del soggetto, ma una struttura formale del pensiero; non è tanto un oggetto del pensiero, ma una maniera con cui diversi pensieri si incrociano. Come si è visto sopra²⁷¹, la struttura prende a tema la "problematica" come la «*condizione di possibilità assolutamente definita, e dunque la determinazione assoluta delle forme di posizione di ogni problema, in un dato momento preso in considerazione*»²⁷².

La problematica, oggetto delle analisi e delle letture sintomali di Althusser, si configura come un sistema chiuso di domande possibili entro un campo concettuale, una sorta di "a priori storico" (per usare un termine caro a Foucault²⁷³) che in partenza ne

²⁷¹ Cfr. parte 1 cap. 1.1.

²⁷² Cfr. parte 1 cap. 2.0.

²⁷³ Althusser, Louis, *et al.*, *Leggere il Capitale*, Mimesis: Milano, 2006, p. 26. Sull'uso althusseriano di "problematica" cfr. anche Althusser, Louis, *Per Marx*, Mimesis: Milano, 2008, pp. 51-55, dove il termine compare per la prima volta in connessione con la lettura sintomale dei testi marxiani.

predispone le pratiche di pensiero: emergono i problemi precipui su cui interrogarsi, le costrizioni che vincolano a determinate aree concettuali, contemporaneamente anche tutto ciò che viene escluso dalla riflessione. La problematica è assimilabile a un' "unità sistematica", ovvero assimilabile a una struttura, da cui scaturisce un senso: se un solo elemento del sistema è sottratto, il senso ne risulta alterato. Il concetto non può venire astratto dalla problematica in cui è inserito: ogni concetto è irriducibile al suo significato preso individualmente e al posto che occupa nel sistema, quindi al suo "ruolo".

L'ideologia, di conseguenza, non si relazione direttamente alla questione della verità di un sistema, ossia alla falsità o alla verità di certe proposizioni, al loro valore logico; molti enunciati ideologici sono infatti veri empiricamente. Essa invece opera da scansione che consente, in un dato momento storico, di far apparire come naturali, evidenti e inevitabili certe forme contingenti di posizione del problema. La componente ideologica invece si spiega tramite il riconoscimento del soggetto in una rappresentazione immaginaria del proprio rapporto con il mondo²⁷⁴.

In ragione di questa struttura, la scienza non può limitarsi a correggere i singoli enunciati ideologici, ma deve operare un cambiamento di terreno, la "rottura epistemologica" secondo Gaston Bachelard. Si tratta di una vera e propria mutazione delle coordinate concettuali, mediante la quale un insieme di domande, la problematica ideologica, viene abbandonato in favore di un nuovo campo di problemi che lo invalida da capo. Althusser scrive in modo lapidario:

Si ottiene una scienza a condizione di abbandonare il terreno in cui l'ideologia crede di aver a che fare con il reale, cioè abbandonando la sua problematica ideologica (il presupposto organico dei suoi concetti fondamentali, e insieme al sistema, la maggior parte dei concetti stessi) per fondare "in un altro elemento", nel campo di una problematica nuova, scientifica, l'attività della nuova teoria.²⁷⁵

La rottura epistemologica non configura alcun necessario mutamento di valore di verità, nemmeno un evento empirico o uno scarto cronologico tra sistemi, ma un effetto di posizione: una parola, passando dalla problematica ideologica a quella scientifica cambia

²⁷⁴ Foucault ne fa ampio utilizzo verso fine degli anni '60, in cui ammette esplicitamente il suo debito teorico nei confronti del maestro Althusser. Cfr. Foucault, Michel, *Le parole e le cose*, BUR: Milano, 1978.

²⁷⁵ Quella che Althusser chiama, nel celebre saggio del 1970, « rappresentazione del rapporto immaginario degli individui con le loro reali condizioni di esistenza » cfr. Althusser, Louis, *Lo Stato e i suoi apparati*, Editori Riuniti: Roma 1977, p. 101.

radicalmente funzione. È quanto cerca di mostrare Althusser in *Leggere il Capitale* con i termini “valore”, “lavoro” o “capitale”. Si parla di “rottura” per via di un salto discontinuo che segnala che non ci sia identità d’essenza tra scienza e ideologia. Ciò nonostante, la rottura epistemologica può definirsi come “dialettica”, poiché la nuova scienza non annulla l’ideologia, ma la smaschera e ne fa vedere i limiti. Inoltre, come si è mostrato, la scienza emerge non dal reale, ma tramite il lavoro del concetto su sé stesso, ovvero per necessità interna e per successione contingente di pratiche esterne: quello che si è definito sopra come “necessità della contingenza”²⁷⁶.

Per questo Balibar, collaboratore di Althusser afferma che i suoi concetti sono «sempre-già autocritici [...] sin dalla loro origine, un modo discorsivo di porsi essi stessi in stato di disequilibrio, di assicurarsi contro la fermezza di una “tesi” nel momento in cui la si sostiene»²⁷⁷. Ciò detto, l’ideologia allora si configura come pratica reale, che produce effetti reali nella vita degli individui e che, in quanto tale, deve essere oggetto di una pratica trasformativa, e non di una pura e semplice denuncia.

Il rapporto immaginario del soggetto col mondo è inteso sempre a partire da nozioni come “semplice” e di “totalità”. Althusser muove due critiche ai concetti, complementari, perché sono rivolte a Hegel. l’ideologia opera precisamente attraverso queste due figure. Essa istituisce una origine semplice (la Vita, lo Spirito, l’Uomo, la Tecnica...) di cui la complessità storica sarebbe la mera esplicitazione, e proietta in avanti una totalità espressiva in cui ogni parte rifletterebbe il tutto. Confondendo l’astrazione con il concetto di generalità, Hegel costruisce un impianto di pensiero ideologico: all’origine del suo pensiero c’è un’unità semplice (l’Idea) che si sviluppa attraverso diverse figure storiche; vengono certo descritte delle figure storiche che *negativamente* sembrano contrapporsi all’andamento dell’Idea. la soluzione dialettica consiste invece nell’assimilarsi e superarle: non si tratta allora di una vera trasformazione. Nessuna vera differenza si costituisce: tutto è già contenuto nel cominciamento, e la storia diventa l’esplicazione [*Entfaltung*] di un’essenza preliminare. Non ci può essere l’identità di un concetto che già sia presupposta alle sue singole parti, ai suoi elementi storici come in Hegel.

Da qui si capisce anche l’influenza che Althusser esercitò su Deleuze. È a questo concetto che allude quello di “struttura a dominante”, come totalità complessa, in cui le pratiche non si lasciano ridurre a un’unica essenza. Ogni filosofia che ponga un’origine semplice è di fatto ideologica, sviluppando una teleologia e riducendo la complessità dell’intreccio di pratiche a un unico principio. A ciò si può opporre la

²⁷⁶*Ibid.*, p. 166.

²⁷⁷Cfr. parte 1, cap. 3.0.

sovradeterminazione come contraddizione principale specificata fin nel più intimo del suo essere dalle forme e dalle condizioni storiche concrete del suo esercizio. Se non vi è un'origine semplice e se non vi è una totalità "espressiva", non vi può essere neppure un punto da cui l'ideologia agisca "dall'alto" come pura condizione di possibilità.

Le teorie dell'ideologia dovrebbero però mirare non a un contenuto particolare, ma a una modalità d'espressione; Althusser sembrerebbe sulla giusta linea di pensiero e concorde con Weatherby nel fare dell'ideologia «non una visione, una proposizione o un certo contenuto, ma il modo d'espressione di tale contenuto»²⁷⁸. Gli LLMs offrirebbero per la prima volta di notarla come un fatto interno alla lingua e non come un'istanza superiore e regolativa. È grazie alla funzione poetica che si perpetra quest'ideologia espressiva atta ad assemblare in maniera statistica degli enunciati, ovvero tanto in modo generativo quanto in modo vincolato. Ideologia "poetica" quindi perché le parole e i concetti sono impacchettati entro un certo paradigma, eppure per noi rimangono imprevedibili. Espressione qui non ha alcun significato di tipo soggettivistico, né rimanda ad alcuna interiorità della macchina. Esso denota piuttosto una teoria dell'ideologia che la concepisce come lo stato totale, superficiale eppure conforme alla legge dell'espressione in un dato sistema di rappresentazioni; l'ideologia non può essere un'istanza trascendentale, ma è la stessa *heat map* dei discorsi, il calco statistico dei percorsi linguistici più attraversati, quindi più "normalizzati".

Le nostre teorie dell'ideologia tendono ad essere a proposito di un qualche fondamento che governa il discorso, di certe condizioni di possibilità. La "base economica" in Marx e l'inconscio nella tradizione freudo-marxista hanno tipicamente svolto la funzione di condizioni, rendendo difficile vedere come ciò che è effettivamente detto corrisponda alla condizione stessa. Penso che gli LLM debbano farci uscire da questa modalità "trascendentale" nella teoria dell'ideologia. Poiché aprono l'ideologia a una forma di scansione empirica, possono contribuire a spostare il nostro senso dell'ideologia in modo da includere – e perfino fondarsi su – l'espressione.²⁷⁹

Quindi Weatherby non etichetta come sbagliate le teorie dell'ideologia passate, ma ne sottolinea l'inadeguatezza, nella misura in cui oggi è resa nota e visibile lo strumento di connessioni empiriche con cui ogni "assoggettamento sociale"²⁸⁰ ideologico. L'ideologia

²⁷⁸ Balibar, Etienne, *Tu l'as dit!*, in *Per Althusser*, Manifestolibri, Roma 1991, pp. 21-22.

²⁷⁹ Weatherby, Leif, *Language machines*, Minnesota University Press: Minneapolis, 2025, p. 173.

²⁸⁰ *Ibid.*, p. 174.

si rende scansionabile empiricamente perché corrisponde alla stessa modalità di calcolo dei LLMs per offrire nuove e “inattese” formulazioni verbali, senza aver bisogno di postulare un fondo irriducibile (l’istanza trascendentale). Lo slittamento empirico è una ridistribuzione del peso teorico tra condizioni ed espressione: l’espressione si enfatizza non per smontare la presenza di condizioni a contorno, ma per sottolineare come la loro intelligibilità sia possibile solo per mezzo della loro estensione effettiva, statisticamente afferrabile.

Venendo definita come il fondamento della rappresentazione ideologica e deformata, l’ideologia viene relegata entro un’istanza che diviene a sua volta irrappresentabile, come nel caso di Marx. La sua “base economica” viene postulata per posizionare l’ideologia, rendendola tanto irrappresentabile quanto di primaria importanza nell’organizzare il dicibile. Allo stesso modo sembrerebbero operare concetti come “totalità falsa” o “mondo amministrato” che rendono conto dell’opacità del processo sociale per Adorno²⁸¹.

Il bersaglio più rilevante è però Althusser; Weatherby, infatti, cita la sua celebre definizione di ideologia come «*rappresentazione del rapporto immaginario degli individui con le loro reali condizioni di esistenza*»²⁸², considerata come una delle «*più influenti nozioni dell’ideologia trascendentale dell’ideologia oggi disponibile*». Essa fa da ponte tra ideologia e ordine simbolico per Lacan e ne fa la struttura che permette il riconoscimento di un soggetto come soggetto simbolicamente localizzabile: fa dell’esponente politico *un* politico, dell’insegnante *un* insegnante, allo stesso modo con cui Althusser esemplifica la struttura formale dell’interpellazione [*interpellation*] attraverso la figura del poliziotto. Così si dimostra che l’ideologia non è nulla di soggettivo, allora da situare entro la sfera dell’immaginario personale, ma un’istituzione “oggettiva”, prodotta dall’insieme di istituzioni statali come l’asilo, l’ospedale, la polizia, ma che si potrebbero allargare anche a istituzioni trans-statali come le piattaforme digitali dei *social* oppure i LLMs. Il problema di una tale configurazione è però che rischia di rimanere sincronica e quindi di ipostatizzare questi ruoli formalmente, non vedendone il terreno concreto su cui poggiano (nella celebre tripartizione lacaniana di RSI infatti il

²⁸¹ Il termine viene volutamente inserito per connettere la riflessione sull’ideologia di Weatherby a uno dei due aspetti che reggono il discorso di Maurizio Lazzarato. Egli intende mostrare l’apporto delle semiotiche asignificanti di Guattari e all’“asservimento macchinico” entro la teoria critica, ancora inadeguata perché esclusivamente ancorata all’“assoggettamento soggettivo”, ancora troppo “logocentrico” perché commisurato su soggetto e linguaggio verbale. Lazzarato, Maurizio, *Macchine e segni*, Ombre corte: Verona, 2019.

²⁸² Adorno viene preso in considerazione anche con Horkheimer come si vedrà; il riferimento teorico è da ritrovare in Adorno, Theodor, *Contributo alla teoria dell’ideologia*, in *Scritti sociologici* I, Einaudi: Torino, 1976, p. 465.

Reale rimane irrepresentabile). Weatherby prosegue quindi bollando il concetto di ideologia in Althusser di essere troppo trascendentale. In realtà, il riferimento è pertinente per l'articolo del 1970 per quanto concerne l'interpellazione, ma potrebbe risultare eccessivo per gli scritti precedenti, che sono stati presi in considerazione qui. In particolare, nei saggi di *Per Marx* e in *Leggere in Capitale*, l'ideologia non si presenta tanto come elaborazione di una condizione trascendentale quanto come critica delle problematiche ideologiche e concretamente esistenti, tramite letture sintomali e rotture epistemologiche. La problematica non è mai una forma a priori, ma una configurazione storica del campo concettuale, immanente alle pratiche teoriche: «*il tutto reale unificato dalla problematica*»²⁸³. Detto questo, rimane vero il presupposto che l'ideologia per Althusser si riversa nel campo della pratica teorica. Il fatto che l'operazione epistemologica di Althusser è finalizzata al discorso politico e di critica attraverso una rilettura dell'opera di Marx e ignora completamente la possibilità che l'ideologia potesse incarnarsi entro uno strumento da renderla quantificabile, misurabile, mentre Weatherby sembra farla coincidere con la sua stessa estensione.

Allora accorta è la definizione che Foucault dà dell'ideologia, facendone una conoscenza di un campo di relazioni; quindi, la forma che un insieme di rappresentazioni storiche prende entro un determinato periodo.

L'ideologia non interroga il fondamento, i limiti o la radice della rappresentazione; percorre il campo della rappresentazione in genere; fissa le successioni necessarie che vi compaiono; definisce i nessi che vi si intrecciano; mette in luce le leggi di composizione e di scomposizione che possono regnarvi. Dispone ogni sapere nello spazio delle rappresentazioni, e percorrendo tale spazio, forma il sapere delle leggi che lo organizza. È, in un certo senso, il sapere di tutti i saperi.²⁸⁴

È proprio sulla base di una continua codifica e decodifica dei dati, di una composizione e scomposizione che i LLMs lavorano. La loro *heat map* può essere considerata a buona ragione allora ideologica, individuando sequenze necessarie, definendo legami tra rappresentazioni, ossia connettendo statisticamente i *token*. Oltre a Foucault, Weatherby segue le tracce anche del pensiero di Wendy Chun, in cui *software* e ideologia sono definiti compatibili; i computer sono macchine ideologiche grazie all'apporto

²⁸³ Weatherby, Leif, op. cit., pp. 188-189. Weatherby cita Althusser cfr. Althusser, Louis, *Lo Stato e i suoi apparati*, Editori Riuniti: Roma 1977, pp. 101.

²⁸⁴ Althusser, Louis, *Per Marx*, Mimesis: Milano, 2006, p. 78.

“immaginario” del *software* sull’*hardware*, creando avatar e utenti per una persona, aiutandola ad immedesimarsi e permettendo all’«immateriale di emergere come merce, come qualcosa di dotato di una propria realtà»; ciò nonostante il *software* rimane un «nocciolo duro e *irrappresentabile*»²⁸⁵. Bisogna però proseguire con un passaggio decisivo: smettere di postulare una condizione e considerare la «*totalità sub-concettuale delle regole sincronizzata con l’espressione effettiva, in un dato momento della storia della conoscenza*» come ideologia²⁸⁶.

In virtù della sincronia e della coincidenza tra calcolo statistico dei LLMs e discorso ed opinione pubblica, nel senso di largo data set di informazioni, slogan, modi di dire, proverbi, pare difficile separare nettamente una dimensione trascendentale di senso da una empirica di significazione. In altre parole, rispetto a ciò che si dice, non sembra esserci alcun rinvio ad un’ulteriore istanza più vera, “autentica” che chiarifichi l’esteriorità opaca del discorso pubblico, funzione invece che ogni discorso filosofico attorno all’ideologia ha ricercato. Lo slittamento empirico dei LLMs annulla ogni distanza tra “verità” e “ideologia” surrettizia in quello che manifestano. L’ideologia diventa «espressione, ideologia in estensione»²⁸⁷ perché ciò che il modello produce è il precipitato empirico delle stesse categorizzazioni, tassonomie adottate dagli individui in carne ed ossa. Ogni output dell’AI è da intendere essere ideologico allora non come causa di un’istanza privilegiata, ma perché incarna la distribuzione statistica²⁸⁸ delle idee più largamente diffuse che costituiscono la cultura.

L'intrattenimento [*amusement*] è il prolungamento del lavoro nel tardo capitalismo. Esso è ricercato da chi aspira a sottrarsi al processo lavorativo meccanizzato per poterlo poi in grado di affrontarlo e di essere alla sua altezza. Ma nello stesso tempo, la meccanizzazione esercita un tale potere sull'uomo che utilizza il suo tempo libero e la sua felicità, determina in modo così integrale la fabbricazione dei prodotti di svago, che il lavoratore fuori servizio non può sperimentare altro che copie e riproduzioni del processo lavorativo stesso. Il contenuto apparente è solo un esile pretesto; ciò che rimane impresso è la sequenza automatizzata di compiti standardizzati. L'unica via di fuga dal processo lavorativo in fabbrica e in ufficio è l'adattamento ad esso nell'ozio. Questa è la malattia incurabile di ogni intrattenimento.²⁸⁹

²⁸⁵ Foucault, Michel, *Le parole e le cose*, BUR: Milano, 1978, pp. 260-261.

²⁸⁶ Weatherby, Leif, op. cit., p. 190. Corsivo nostro.

²⁸⁷ Weatherby, Leif, op. cit., p. 188. La formula «*expressive theory of ideology*» è il principio rivendicato dall’autore quindi e il bersaglio teorico polemico è esplicitamente la tradizione che fa risalire l’ideologia a una condizione trascendentale.

²⁸⁸ *Ibid.* p. 191.

²⁸⁹ «Qualcuno ha visto mille uomini adulti. Se ora egli vuole giudicare della grandezza normale da stimare

La cultura non è nulla più che una «sequenza automatizzata di compiti standardizzati»²⁹⁰; la denuncia che Adorno ed Horkheimer muovevano all'industria culturale, in particolare alla produzione di film, appare oggi molto meno altisonante, tenendo conto che con i modelli linguistici, però, questa sequenza non è soltanto fruita ma generata in tempo reale dai prompt degli utenti. Se l'output è costruito sulla base della scansione empirica delle informazioni di cui una cultura dispone, allora è come se in ogni prompt operasse una prima fase di "schematizzazione", calibrandolo su tendenze già solcate a più riprese.

Il compito che lo schematismo kantiano aveva ancora lasciato ai soggetti, e cioè quello di riferire in anticipo la molteplicità dei dati sensibili ai concetti fondamentali, è levato al soggetto dall'industria. Essa attua e mette in pratica lo schematismo come primo servizio del cliente, Nell'anima era all'opera, secondo Kant, un meccanismo segreto che preparava già i dati immediati in modo che si adattassero al sistema puro della ragione. Oggi l'enigma è svelato. [...] Per il consumatore non rimane più nulla da classificare che non sia già stato anticipato nello schematismo della produzione.²⁹¹

Per questo il *response* può essere visto come oggetto di "kitsch"²⁹² secondo la definizione che ne dà Clement Greenberg, critico marxista. Greenberg è interessato a capire il rapporto arte-kitsch. Attorno al 1937 riteneva che l'arte fosse creazione pura, che egli associa ai movimenti d'avanguardia che produce «qualcosa di valido solo nei suoi propri termini, come un paesaggio è esteticamente valido; qualcosa di dato, in-

comparativamente, allora l'immaginazione (a mio parere) lascia cadere un gran numero di immagini (forse tutte quelle mille) una sull'altra; e, se mi è permesso di applicare qui l'analogia della rappresentazione ottica, in quello spazio dove si assomma la maggior parte di esse, e all'interno del contorno dove la porzione di spazio si colora dei colori più carichi, là diviene riconoscibile la g r a n d e z z a m e d i a , che è altrettanto lontana, sia in altezza che in larghezza, dai limiti estremi delle stature massime e minime; e questa è la statura per un uomo bello. (Si potrebbe ottenere meccanicamente il medesimo risultato, se si misurassero tutti e mille gli uomini, e se ne addizionassero separatamente le altezze e le larghezze (e lo spessore) tra loro, e si dividesse ciascuna somma per mille. Solo che l'immaginazione lo fa appunto in grazia di un effetto dinamico, che nasce dalla apprensione multipla di tali figure, sull'organo del senso interno). Ora, se in modo simile si ricerca per questo uomo medio la testa media, e per questa il naso medio, e così via, allora questa figura sta a fondamento dell'idea normale dell'uomo bello, nel paese dove la comparazione viene eseguita; per cui, a queste condizioni empiriche, un negro deve avere necessariamente un'altra idea normale della bellezza della figura rispetto a un bianco, e un cinese un'altra rispetto a un europeo. Le cose andrebbero appunto allo stesso modo con il modello di un bel cavallo o di un cane (di una certa razza). – Questa i d e a n o r m a l e non è derivata dalle proporzioni prese dall'esperienza, i n q u a n t o r e g o l e d e t e r m i n a t e ; al contrario è seguendo quell'idea che divengono possibili in primo luogo regole del giudicare». Cfr. Kant, Immanuel, *Critica della facoltà di Giudizio*, Laterza: Roma-Bari, 1907, §17.

²⁹⁰ Adorno, Theodor, Horkheimer, Max, *Dialettica dell'illuminismo*, Einaudi: Torino, 1966, p. 145.

²⁹¹ Weatherby, Leif, op. cit., p. 174.

²⁹² Adorno, Theodor, Horkheimer, Max, op. cit., p. 131.

creato, indipendente da significati, simili o originali»²⁹³. Al contrario, il kitsch «elabora in partenza l'arte per lo spettatore e gli risparmia lo sforzo, offrendogli una scorciatoia verso il piacere dell'arte che aggira ciò che è necessariamente difficile nell'arte autentica. Il Repin, o kitsch, è arte sintetica»²⁹⁴. Essendo simulacri nel senso “estensivo” del termine, le differenze di *response* esibiscono ai margini una differenza di grado entro un campo di pattern già dato in partenza. Il kitsch digitale costituirebbe un'ulteriore categoria rispetto a quello che il designer Silvio Lorusso²⁹⁵ definisce “kitsch classico”. Concorde con Eco, il kitsch classico è una “bugia” dell'arte, consegnando uno schizzo che possa graficamente riprodurre più o meno dell'opera d'arte di per sé. La maggiore o minore quantità di informazione è dovuta all'idiosincrasia che viene offerta come merce. Ora, la nuova categoria di *weird kitsch* sembrerebbe rispondere all'esigenza di delineare un confine tra la bizzarria e l'opera di normalizzazione degli orrori kitsch portata avanti dall'AI. DeepDreaming è un'AI generativa lanciata da Google nel 2019 per disegnare paesaggi onirici. La sua fonte primaria è il ventaglio di figure surrealiste, ad esempio quelle di Max Ernst. Lo spettatore è portato a riconoscere una prima sensazione di stranezza osservando le immagini, che diventano via via la maniera “usuale” di rappresentare il sogno. È come se l'iniziale novità di Ernst venisse appianata in una serie omologata e stereotipata di immagini. Il risultato è che lo spettatore tende sempre meno a stupirsi di queste stranezze, per il fatto che lo stesso LLM si abitua a generalizzare attraverso quella categoria surrealista.

²⁹³ Anche Domenico Quaranta condivide la presente tesi; la produzione digitale è *kitsch* perché il suo utilizzo non viene promosso in maniera critica e severa, ma affina impalcature ideologiche. Cfr. Quaranta, Domenico, *Digital Kitsch: Art and Kitsch in the Informational Milieu*, in *The Changing Meaning of Kitsch: From Rejection to Acceptance*, Springer Verlag, pp. 205-228.

²⁹⁴ Weatherby, Leif, op. cit., p. 181.

²⁹⁵ Greenberg, Clement, *Art and culture*, Beacon press: Boston, 1971, p. 15.



Figura 10 esempio di "make it more"²⁹⁶

L'effetto è ravvisabile anche nel fenomeno del “*di più*” [*make it more*] in cui l'utente chiede al LLM di generare immagini che enfatizzino in misura crescente un certo aspetto: possono così venire prodotti edifici di case, composti di case, cesellate a loro volta da città di case; un gatto può venire rappresentato esaltando a dismisura la sua natura felina e neotenica. Insomma, il kitsch invece che mostrarsi come eterogenea associazione di elementi, diventa un'operazione sempre più retroattiva, perciò conservativa. Più un modello e una categoria si consolidano, più immagini verranno ad assumerne i connotati.

Il concetto di ideologia avanzato da Weatherby vuole liberare dal pregiudizio che l'ideologia sia una semplice illusione, qualcosa di estirpabile una volta per tutte; in realtà essa va trattata come forma reale, dotata di una propria struttura e di una propria efficacia. Lo slittamento empirico che propone Weatherby, quindi, è solo superficialmente contrapponibile all'operazione di chiarificazione scientifica perpetrata da Althusser. In fondo, quello che si sta proponendo è un cambiamento di terreno teorico, una “rottura epistemologica”, che costruisca una nuova problematica che permetta di pensare gli LLM non come miracolosi né come catastrofici, ma come dispositivi di scansione storicamente determinati: questa importante rottura viene rimarcata da Weatherby nel momento in cui insiste sul fatto che si debba «*lasciar cadere*

²⁹⁶ Lorusso, Silvio, *Deepdreaming Willy Wonka: AI Weird as the New Kitsch*, Hidden Layers, 2024.

i modelli trascendentali dell'ideologia e vedere l'IA generativa come una totalità espressiva di ideologia, [così da] guadagnare una visione – afferrare ciò che è generato»²⁹⁷. Se si volesse approfondire il rapporto tra i due casi, si potrebbe dire che per Weatherby il riconoscimento dell'ideologia, non permettendo una sua dissoluzione totale, può però offrire «una visione senza precedenti nella composizione linguistica dell'ideologia», una sua mappatura. Come usare la mappa per articolare una pratica? In questo senso, rendere l'ideologia una modalità di espressione che coincide con la sua estensione empirica ne fa un complemento della rottura epistemologica althusseriana; almeno nella misura in cui essa pretende di vedere l'ideologia anziché abitarla. La pratica teorica ne esce arricchita così, vengono forniti nuovi materiali ma che non dispensano dal lavoro di costruzione di una problematica differente entro cui pensarli.

Anche il concetto di Generalità di cui si è parlato sopra risulta utile per chiarire il riconoscimento dell'ideologia in Weatherby. Per Althusser la rottura tra Generalità I e Generalità III costituisce il passaggio da un'istanza ideologica a una scientifica, opera di una pratica teorica. La Generalità II è il mezzo altro rispetto ai materiali della Generalità I. I LLMs sono invece macchine di Generalità I e II: accumulano un'enorme massa di Generalità I sottoforma di dati e li trasformano in uno specifico apparato di cattura secondo nuove codificazioni algoritmiche, che altro non sono che il prodotto (in senso matematico) dei dati accumulati iniziali, grazie al meccanismo di *attention*²⁹⁸. Questo punto merita una precisazione: alcuni critici²⁹⁹ indirizzano agli LLMs una *quarta memoria* affianco alle tipologie di ritenzione, ovvero di conservazione dell'informazione. Tradizionalmente, si possono individuare tre tipologie di ritenzione: la percezione, la memoria di percezioni e atti passati³⁰⁰ e l'utilizzo di *media*. Una quarta tipologia verrebbe a situarsi proprio nel massivo ricircolo di informazioni accumulate nel dataset. La Generalità I fornisce la possibilità per un'induzione statistica di generare *ritenzioni simili* e non nuove ritenzioni del terzo tipo. Come dimostra il kitsch tipico di questi modelli, anche qui la tendenza è regressiva, plasmando il futuro sul passato. Viene indicato col termine *pretensione*, il circolo che viene a crearsi tra la ritenzione del primo tipo e la protensione, ossia la tensione futura verso l'istante successivo. Vuol dire che il quarto tipo di memoria è tale perché plasma delle ritenzioni che assomigliano e vagamente ricordano quelle passate, ma secondo una nuova logica induttiva che

²⁹⁷ Immagine da *Ibid.*

²⁹⁸ Weatherby, Leif, op. cit., p. 192.

²⁹⁹ Si rimanda al capitolo sul Word embedding.

³⁰⁰ Chartonsky, Grégory, Citton, Yves, *The Fourth Memory*, in «Transit», 15-1, 2025.

continuamente le riscrive.

Quindi il ragionamento di Weatherby correttamente addita la nuova riformulazione di Generalità I come ideologica, perché non si presenta alcuna rottura; ogni distribuzione statistica non può essere una conoscenza scientifica. Essa è la materia prima del linguaggio nella sua diffusione, l'insieme dei percorsi prestabiliti che ogni individuo percorre quando fa uso del linguaggio. È, anch'essa, già conoscenza (conoscenza statistica, conoscenza distribuzionale, conoscenza di senso comune) e proprio in quanto tale costituisce il punto di partenza di ogni discorso ulteriore.

A ciò corrisponde la nozione di effetto di livellamento [*averaging effect*], per cui gli LLMs calcolano le probabilità delle continuazioni linguistiche di una sequenza di prompt, producendo un output che coincide con il cammino più probabile entro la combinazione di token del testo. L'operazione di livellamento non è la mera somma delle parti: essa costituisce, ad un tempo, un effetto di totalità, affine alla nozione di senso deleuziano, e un effetto di semplificazione (ciò che è *semplicemente*³⁰¹ una sequenza è in realtà il risultato di una struttura complessa di relazioni). È quindi l'operazione di smussamento statistico delle differenze che opera ideologicamente.

In conclusione, Weatherby nella sua formulazione empirica del concetto di ideologia insegue un'operazione di rottura critica che si richiama ad Althusser e che può essere vista come suo complemento. I concetti althusseriani di "totalità" e di "semplice" possono venire ripresi come strumento per mostrare la contingenza storica su cui le macchine linguistiche si appoggiano. Due sono le tendenze ideologicamente più comuni: credere che vi sia un'origine semplice (il dato, l'utente) oppure credere che si è di fronte a una totalità espressiva, organicamente raccordata (anziché di fronte a una complessa operazione di riduzione della distribuzione di differenze).

2.2 SIMULAZIONE CULTURALE

Nel paragrafo precedente si è mostrato come la generatività prenda per oggetto una molteplicità di regimi che non esclude alcun genere di simboli: persone, lingue, immagini, video, riproduzioni plastiche etc. Questo si era detto essere il primo aspetto della poetica generale secondo Weatherby. Il secondo aspetto può invece venire riassunto facendo riferimento di nuovo all'ipotesi differenziale di Saussure. La struttura

³⁰¹ A parte la terza tipologia, si noti che le prime due tipologie di ritenzione ricalcano i due modi di fissazione della conoscenza per Hume: impressioni e idee. L'argomento è conforme alla struttura degli LLMs ricorsiva ed empirica.

del linguaggio non si lascia ridurre a una mera tassonomia di significati legati ai loro referenti, bensì rispetto a un sistema di valori differenziali, ossia legati vicendevolmente, senza alcun rinvio a invarianti esterne. Come si diceva, i LLMs realizzano tecnicamente e non solo metaforicamente questa ipotesi: ogni token viene posizionato entro uno spazio multidimensionale in cui i rapporti sono puramente linguistici. La co-occorrenza statistica con gli altri tokens diventa il solo criterio di prossimità per analizzare i rapporti intra-linguistici. La poetica generale è, sotto questo profilo, la prima messa in opera meccanica della struttura saussuriana intesa come puro gioco di differenze.

Ad esempio, la dicotomia denotazione-connotazione collassa, per quanto ne discute Ronald Barthes nel saggio *La retorica dell'immagine*, nella stessa misura in cui collassa la dicotomia segno-referente. Il segno si è visto essere arbitrario, slegato dal referente, così la denotazione diventa sempre più naturale quando amplificata e sviluppata per mezzo della tecnologia. È la tecnologia dei LLMs ad essere in questo caso veicolo di diffusione di nuove connotazioni: «Più la tecnologia favorisce la diffusione delle informazioni (e in particolare delle immagini), più offre i mezzi per mascherare il significato costruito sotto le spoglie del significato apparente»³⁰². La connotazione simbolica artificiosa e costruita come mezzo di propaganda viene a sovrapporsi e, in ultima istanza, ad offuscare la denotazione naturale dell'immagine. Un esempio che segue a questo ragionamento lo fornisce il riempimento generativo di Adobe della *Gioconda* funziona come estensione che lavora attraverso il pacchetto semantico dell'immagine, addirittura del quadro: «riempie con un grandioso paesaggio montano sotto un cielo enorme, come se questa giovane signora rinascimentale fosse stata trasportata nel Montana del XIX secolo»³⁰³.

L'immagine estesa prodotta dal *fill* generativo non tiene in considerazione le ore ed ore di meticoloso esercizio pittorico, ma un'astratta ricombinazione di pixel, ampliando in modo "errato" il pacchetto semantico. Le virgolette di errato intendono sottolineare come il discorso qui non prenda a riferimento la correttezza o meno dell'operato macchinico (si lascia al lettore il compito di trarne le conclusioni), ma che la poetica generale si dimostra come uno strumento che annulla la dicotomia connotazione-denotazione perché si pone prima di ogni vera o sbagliata attribuzione d'opera e quindi prima di ogni valore di verità.

³⁰² Confermando ancora una volta che il semplice è già ideologico, un pensiero che concepisce il tutto strutturato secondo una contraddizione dominante e, geneticamente, a partire da un'origine.

³⁰³ Barthes, Ronald, *La retorica dell'immagine* in *L'ovvio e l'ottuso*, Einaudi: Torino, 2001.



Figura 11 estensione del ritratto della Gioconda secondo Adobe³⁰⁴

Lo si può sostenere o criticare, conclude Weatherby, ma in ogni caso le posizioni opposte trovano il loro fondamento «*nella poetica generale del paesaggio semiologico digitale*»³⁰⁵. Con questa formula la poetica generale assume il ruolo di una grammatica trasversale.

In che rapporto sta la generazione con la computazione? «*La generazione trasforma la computazione in linguaggio, facendone un nuovo medium*». Come si era accennato sopra la forma di intelligenza con cui si suole interrogare le macchine è da attribuire a Turing, il primo a rendersi conto che una macchina non deve comprendere il contenuto di ciò che organizza, ma semplicemente operare sotto determinate condizioni. Il fatto che i contenuti processati non vengano compresi, ma semplicemente eseguiti si vede anche per la generazione dei LLMs: «*tagliano il legame tra comprensione e generazione, ma reintroducono così anche il linguaggio nel problema della computazione*»³⁰⁶. È lo stesso linguaggio che ora viene processato e ricombinato prendendo a riferimento un data set vasto e massivo. Così si può dire che il linguaggio è generativo:

L'algoritmo interiorizza questa funzione separata – la funzione poetica – esteriorizza

³⁰⁴ Weatherby, Leif, op. cit., p. 182.

³⁰⁵ Immagine da <https://www.fanpage.it/innovazione/tecnologia/lintelligenza-artificiale-svela-tutto-lo-sfondo-della-gioconda-ma-il-risultato-non-piace-a-nessuno/>.

³⁰⁶ *Ibid.* p. 182.

una forma di sintesi scrittoria, solo per offrirmela come bozza, come testo possibile. [...] Il linguaggio computazionalmente riproducibile è “generativo” nel senso che la generazione non può mai essere una copia. Questa capacità si inserisce perfettamente nell’economia semiotica delle piattaforme.³⁰⁷

Da questo passaggio emerge in maniera lampante la radicalità della tesi di Weatherby. Da un lato, la generatività delle macchine è messa in primo piano; significa che non si limitano a un montaggio meccanico di sequenze, ma producono differenze. La macchina, infatti, riproduce e ogni riproduzione compete a una nuova formulazione che non può mai essere vista come copia, che si è chiamata simulacro.

³⁰⁷ *Ibid.* p. 204.

PARTE 4 MACCHINE SEMIOTICHE

CAP. 1: DALLA STRUTTURA ALLA MACCHINA

Quanto si è argomentato nei capitoli precedenti, sulla scorta delle ricerche di Leif Weatherby e Lydia Liu mette a fuoco un primo livello del problema: i Large Language Models possono essere descritti come strutture linguistiche, ovvero come sistemi di rapporti differenziali in cui ciascun elemento (il vettore-parola) acquista valore solo per la posizione che occupa rispetto agli altri, senza alcun ricorso al referente esterno. Esattamente come per Saussure il linguaggio si presenta «*sotto il carattere della forma, piuttosto che della sostanza*»³⁰⁸ così i modelli ad architettura *transformer* tessono una rete di vettori che fa del significato un posizionamento vettoriale. Sotto questo profilo, l'ipotesi strutturalista non è che un'eredità per il momento ben funzionante: la cosiddetta "autonomia del simbolico" rispetto al reale, formulata da Lacan e radicalizzata da Deleuze nel suo opuscolo sullo strutturalismo riesce a carpire bene l'effetto di senso che si ha quando si interroga ChatGPT o Claude. In particolare, il posizionamento dei segni funziona per generare nuovo significato attraverso quella che Jakobson ha definito funzione poetica del linguaggio. Questo primo movimento dell'argomentazione non è sufficiente però a inquadrare la complessità dei LLMs; la presente parte intende dimostrare che la lettura neostrutturalista di Weatherby si può usare per descrivere la calibrazione parola-vettore, ma le sfugge tutto ciò che fa di un LLM non già una struttura simbolica chiusa, ma una "macchina" semiotica immersa in flussi materiali, libidinali, economici e politici. Anche Weatherby sostiene che questi modelli producano cultura, ma, come si è visto, ciò viene a dimostrarsi solo sulla scorta della primarietà accordata a un modello arborescente di codifica del linguaggio, istituito come una delle conseguenze principali della poetica generale. Ad opporsi al modello del codice è lo studio della schizoanalisi attraverso il concetto di macchina, poi formalizzato ulteriormente in quello rizomatico dell'*assemblaggio* [*agencement*] da Deleuze e Guattari nei due volumi di *Capitalismo e schizofrenia*.

Un preambolo sembra doveroso al discorso che segue. La schizoanalisi è una

³⁰⁸ *Ibid.* p. 205.

pratica psichiatrica che nasce negli anni '70 e che oramai è caduta in disuso. L'intento qui è allora di tornare a sondarne alcuni principi che si possono ancora rivelare utili, soprattutto se comparati con il suo principio anti-istituzionale e anti-gerarchico. La schizoanalisi nasce infatti in Francia in cliniche come Saint-Alban da personalità come Tosquelles, Oury e poi Guattari. "Pragmatica" è forse il termine che meglio carpisce lo spirito di questa operazione; l'inconscio viene studiato non attraverso una pura teoria, ma un insieme di pratiche che cercassero di rendere autonomo il paziente, senza la continua consultazione da parte del medico. «*Il loro obiettivo era quello di rendere i pazienti più liberi possibile da una forma immediata e diretta di assistenza, affinché poggiassero su sé stessi, per prendere le proprie decisioni e imparare a articolare i loro desideri*»³⁰⁹. Anzi, per come Camille Robcis, nel suo libro *Disalienation* descrive i fini della schizoanalisi, si potrebbe inserire come strumento per rivendicare l'importanza della categoria del "comune"; una rivendicazione che si annuncia anche negli anni '10 del 2000 tramite dei movimenti come quello di Gezi Park a Istanbul contro movimenti autoritari neoliberalisti o storture fasciste che attualmente si fanno sempre stringenti. Ora, la schizoanalisi offre degli strumenti concettuali, delle pratiche permettono di analizzare strumenti dispotici di disciplinamento come gli LLMs. Necessario a tal fine sarebbe uno studio di maggior riguardo. Qui, verranno avanzati dei concetti quali quelli di macchina desiderante, flusso di desiderio, corpo-senza-organi, assemblaggio, parola d'ordine che permettono di far uscire il discorso semiotico che si sta compiendo in queste pagine dalla ferrea impostazione linguistica in cui viene spesso relegato. Si vuole, in altre parole, mostrare come una parola non sia solo un mezzo d'informazione, ma già una trasformazione pragmatica, un catalizzatore di desideri, un utile strumento per configurare il desiderio di massa. Quindi, il presente passaggio schizoanalitico va inteso a fronte di un progetto di recupero di alcuni concetti che possono ancora presentarsi utili in veste critica attualmente³¹⁰.

Quello che qui interessa non è tanto l'apporto della schizoanalisi al discorso psicanalitico. Non si avrà strettamente a che fare con Edipo e con Lacan. La psicanalisi si interseca al discorso dei LLMs, ma costituirebbe un *excursus* molto più ampio da trattare. Per rimanere al tema qui presentato, è bene attendersi alle parole di Deleuze e Guattari:

³⁰⁹ Saussure, Ferdinand de, *Corso di linguistica generale*, trad. it. di T. De Mauro, Laterza: Roma-Bari 1967, p. 137; cfr. anche Weatherby, Leif, *Language Machines. Cultural AI and the End of Remainder Humanism*, University of Minnesota Press, Minneapolis 2025, pp. 41-42, dove la formula viene usata per definire l'"ipotesi differenziale" che si è definita sopra.

³¹⁰ Robcis, Camille, *Disalienation*, University of Chicago Press: Chicago, 2021, p. 146.

Quello che noi cerchiamo in un libro è il modo stesso in cui esso fa passare qualcosa che sfugge ai codici: i flussi, le linee di fuga attive, rivoluzionarie, linee di decodificazione assoluta che si oppongano alla cultura.³¹¹

Il libro a cui si stanno riferendo è il primo volume di *Capitalismo e schizofrenia*, una ricerca a quattro mani che si è concretizzata in due libri: l'*Anti-Edipo* e *Mille piani*. Altrove³¹² Deleuze accenna a una scrittura-flusso da opporre a una mera codificazione. Il codice, infatti, intrattiene sempre una natura dicotomica, una composizione arborescente in cui ad ogni opposizione concettuale ne segue un'altra in maniera classificatoria e stabile. Come si fa ad uscire dal dominio dispotico del Significante? Come si può uscire dalle astratte categorizzazioni? Di certo ad un posizionamento non può opporsi un'ulteriore istanza, altrimenti si rimane entro l'ambito dell'astratta identità che precede ogni ordine di differenza. Così bisogna essere *tra* differenze, ossia seguire il processo che le avviluppa, le lega, le distorce e le trasforma. Per questo Deleuze è sempre rimasto vicino a una posizione "differenziale" e intensiva. Le intensità delle singolarità, che scopre ancora come entità preindividuali nell'ambito biologico, fisico all'altezza degli anni '60 possono ora servire da strumento critico per evadere dalla logica del significante, che emblematicamente al tempo si stava generalizzando grazie allo strutturalismo.

Lo strutturalismo nasce con la precisa intenzione di studiare i condizionamenti simbolici entro un insieme di elementi, ovvero le loro eventuali maniere di ricombinarsi e trasformarsi. Solo Deleuze però sembra portare al limite³¹³ il ragionamento facendo derivare dalla struttura un precipitato concettuale che nulla sembra avere a che fare con le conclusioni del ragionamento strutturalista: l'ipotesi strutturalista sembrerebbe vietare la stessa natura strutturale della struttura³¹⁴. Il figlio "mostruoso" dello strutturalismo non può essere un'ulteriore struttura, ma una "macchina". Il termine non è da intendere secondo l'accezione organicistica né tecnica. Non si tratta quindi né di un mero strumento né di un componente che si può individuare solo sulla scorta di un tutto fisso prestabilito. La potenzialità della macchina è piuttosto la sua inesauribile capacità di composizione e scomposizione di codici, che riporta alla "generatività dinamica" di

³¹¹ Per una trattazione di alcune tematiche deleuziano-guattariane, come quella del campo d'immanenza e della de-soggettivazione macchinica si rimanda a Lazzarato, Maurizio, *Segni e macchine. Il capitalismo e la produzione di soggettività*, Ombre corte: Verona, 2019.

³¹² Deleuze, Gilles, *Pourparler*, Quodlibet: Macerata, 1999, p. 33.

³¹³ *Ibid.* p. 17.

³¹⁴ Deleuze rende noto il suo metodo storico-filosofico di interpretazione dei testi. «Mi immaginavo di arrivare alle spalle di un autore e di fargli partorire un figlio, che fosse suo e tutta via fosse mostruoso» cfr. *ibid.* p. 16.

Kockelman per cui «la medesima procedura può venire riportata ad un altro fine»³¹⁵, generando nuove strutture secondo non una combinatoria chiusa di elementi, ma la trasversalità dei rapporti; quindi, una linea di fuga rispetto alle strategie più normalizzate.

Il termine struttura viene sostituito da quello “molare” perché non solo è parziale e non lascia intravedere la dimensione molecolare³¹⁶, ma anche inadeguato (poiché non c'è mai solo *una* struttura, ma sempre una molteplicità). Dunque, la molteplicità di elementi da accordare secondo un ordine strutturale diventa ora una molteplicità frattale di strutture, dunque un ordine sempre aperto. La macchina viene prima della struttura, quindi prima di ogni dicotomia. Una in particolare viene ribadita a più riprese: la coppia natura-cultura. Prima di ogni differenziazione avviene il processo (macchinico) che lo produce. Il processo che Deleuze e Guattari chiamano “schizoanalisi” ha proprio per obiettivo quello di unire i due grandi estremi del pensiero: Natura e Storia. Il processo mette in luce come uno sviluppo funziona e si trasforma nel corso del tempo. Affermare come qualcosa funziona è diverso da disporre un'ontologia che si pieghi alla chiarificazione di ciò che ha davanti, ossia rifugga da ogni logica del significante che cerca di definire ogni ente, chiedendosi cosa esso sia. Il funzionamento non è organicamente predisposto, secondo un modello unitario, ma è frammentato e disperso entro una dimensione collettiva che impegna la società nella sua totalità.

La macchina, quindi il prototipo funzionalista dell'ipotesi morfogenetica, viene prima anche di ogni soggetto. Non dovrebbe sorprenderci questa tendenza a mettere la macchina al posto dell'uomo e l'uomo a lato della macchina. Per macchina non si intende il macchinario tecnico-industriale su cui già autori marxisti si erano impegnati in attente analisi, ma di un vero e proprio componente sociale, che sia umano o tecnico, organico o inorganico. Stando lo strutturalismo di contro alla nozione fenomenologica di soggetto e di intenzionalità personalistica; essendo la macchina prima di ogni struttura e identità, allora il concetto di macchina non potrà che precedere quello di soggetto. Infatti, è la linguistica di Hjelmslev, una “teoria spinozista”, a tornare utile per descrivere il continuo

³¹⁵ Alla stessa conclusione ci arrivarono altre due figure negli stessi anni. Félix Guattari infatti compone un saggio nel 1969 proprio riguardante la nozione di «macchina rivoluzionaria» come strumento di analisi di contro all'imperante movimento lacaniano, cfr. Guattari, Félix, *Machine and structure in Psychoanalysis and transversality*, Semiotext(e): Los Angeles, 2015, pp. 111-119. La funzione della macchina è quella di non poter mai venire catturata definitivamente entro un certo paradigma, ma di funzionare semplicemente, ovvero di concatenarsi ad altre macchine. Ogni struttura non può che essere struttura di strutture e a sua volta una sottostruttura di una struttura più grande.

Anche Umberto Eco vi arrivò, dimostrando che l'unica ipotesi strutturalista ammissibile sia quella funzionalista e non ontologica. Qualsiasi ammissione di una struttura non può che accogliere un'ulteriore struttura, creando una regressione infinita, che rende impossibile istituire qualsiasi struttura originaria. cfr. Eco, Umberto, *La struttura assente*, Bompiani, 1961.

³¹⁶ Kockelman, Paul, *Last Words*, Pickly Paradigmatic Press: Chicago, 2024, p. 81.

ri-definirsi dei flussi

Il linguaggio, come sistema di flussi continui del contenuto e dell'espressione, viene ritagliato da concatenamenti macchinici di figure discrete e discontinue. Quello che non abbiamo sviluppato nel libro è una concezione degli agenti collettivi di enunciazione che vorrebbero superare la frattura tra soggetto di enunciazione e soggetto di enunciato.³¹⁷

L'universo delle macchine richiama non un soggetto ma molteplici individualità a seconda del grado di differenziazione; il soggetto viene prodotto come «resto»³¹⁸. Ad ognuna corrisponde una forma che non le è propria, ma un "ruolo" molare, ma che non è né principio né fine di un presunto soggetto. Il compito dell'operazione di schizoanalisi è quello di ridistribuire i codici e le soggettività ad essi consegnate. Seguendo la linea di pensiero già promossa in *Differenza e ripetizione*, nessuna identità è da attribuirsi se non seguendo il processo di differenziazione che l'ha portato ad essere, quindi le quantità intensive che l'hanno fatto nascere. Sono le zone d'intensità, composte da gradienti, ossia la dimensione molecolare, ad essere sempre implicate quando si parla di "macchina" e non un organismo di componenti, così da illuminare il *reale processo di produzione*³¹⁹ e non qualche sua rappresentazione ideologica.

Non esiste una logica individuale scissa dal contesto sociale in cui è calata. Perciò si dice a buona ragione che «l'investimento nel campo sociale» primeggia «su quello familiare» perché «il primato del campo sociale come termine dell'investimento di desiderio definisce il ciclo e gli stati attraverso cui un soggetto passa»³²⁰. Il ciclo, ossia il continuo rimando tra una macchina e l'altra è l'unico punto assolutamente da prendere in considerazione; chiedersi quale "personaggio" o categoria o schema abbia maggiore o minore attinenza ed importanza è una domanda dell'ordine della rappresentazione. «Abbiamo visto che, attraverso la riproduzione e i suoi oggetti (determinati familiarmente o geneticamente) è sempre l'inconscio a produrre se stesso in un movimento ciclico orfano, ciclo di destino ove resta sempre soggetto»³²¹.

³¹⁷ Molecolare-molare è una dicotomia introdotta da Guattari e Deleuze con l'*Anti-Edipo*; l'idea è che il regime molecolare non sia semplicemente una differenza di grado rispetto a quello molare, ma una vera e propria differenza di natura. Dal lato molare si instaurano identità, primeggia il linguaggio come mezzo di indicizzazione, mentre è solo il lato molecolare che continuamente produce e disfa gli aggregati molar, quindi popolandosi di singolarità preindividuali e di sole "differenze in sé".

³¹⁸ Deleuze, Gilles, *Pourparler*, Quodlibet: Macerata, 1999, p. 32.

³¹⁹ Deleuze, Gilles, Guattari, Félix, *Anti-Edipo*, Einaudi: Torino, 1975, p. 19.

³²⁰ A tal proposito, la schizoanalisi si definisce come psicanalisi materialista; «la grande scoperta della psicanalisi è stata quella della produzione desiderante, delle produzioni dell'inconscio» Cfr. *Ibid.* p. 26. Ogni tentativo di ripiegare il processo creativo entro modelli e figure è idealistico.

³²¹ *Ibid.* p. 314.

Quindi la schizoanalisi si offre come uno strumento critico nei confronti delle coagulazioni molari e strutturali per operare una decodificazione (deterritorializzazione), mettendo cioè in cortocircuito il sistema. Basta associare nuove parole e nuove cose? I LLMs ci mostrano che ciò non è sufficiente. Ogni risposta data è infatti calibrata su un nuovo algoritmo, ma la procedura sembra rimanere la medesima, attingendo a un fondo molecolare. La microfisica del livello molecolare rifugge da ogni distribuzione statistica, fulcro invece dell'architettura *transformer*.

La schizoanalisi prende per oggetto un inconscio formato secondo macchine desideranti; le sue non sono componenti meramente spirituali o materiali, vitalistiche o meccanicistiche, ma un aggregato delle due. La tesi fondamentale da rimarcare è che desiderio e produzione non sono due regimi disgiunti; il desiderio viene concepito come estrinseco se è effetto di cause meccaniche oppure se la macchina risulta in un sistema di mezzi in funzione del desiderio. Il materialismo della schizoanalisi fa convergere queste due dimensioni, liberandosi sia di tesi vitalistiche che meccanicistiche. Il vitalismo concepisce l'individuo a partire da un'unità vivente che fa da base per ogni possibile prolungamento macchinico. La macchina è estensione, arto umano, poiché può rendere conto di come funzioni, ma non di come si sia generata³²². Il meccanicismo delle macchine invece, non rendendo conto del processo produttivo che lo informa, offre una soluzione astratta al problema, non tanto riducendo l'organismo a una macchina, ma insistendo sulla unità strutturale della macchina. Da una prospettiva schizoanalitica l'unità strutturale e individuale, tanto il meccanicismo e il vitalismo, offrono delle soluzioni riduttive per studiare la complessità dei sistemi tecnici in atto. «*Quel che ci trae in inganno è il considerare ogni macchina complicata come un oggetto unico. In realtà, essa è una città o una società di cui ogni membro è direttamente procreato secondo la sua specie*»³²³. Uomo e macchina vengono a disperdersi come unità compatte, lasciando spazio a loro aggregati. È qui in luce un aspetto che il secondo volume, *Mille piani*, andrà ad analizzare più in profondità: la necessità di richiamarsi a concatenamenti, ovvero assemblaggi d'enunciati e macchinici; si scopre che ogni uomo è popolato da una molteplicità di macchine, come le macchine viceversa, sono cittadelle di uomini. La vera differenza non è tra uomo e macchina, ma tra macchine molari (tecniche, sociali, organiche) e macchine

³²² *Ibid.* p. 330.

³²³ Il meccanicismo è quindi un antecedente storico dello strutturalismo: entrambi i sistemi di pensiero rendono conto di un funzionamento analiticamente e si affrancano dai problemi relativi alla formazione: il meccanicismo non si cura dell'origine, al più riservandole il posto di causa prima, mantenendo perciò l'impianto lineare causale, mentre lo strutturalismo ne nega la possibilità, ravvisando una complessità di concause da sempre in gioco.

desideranti.

La macchina desiderante lavora per tagli e flussi, vuol dire che si presenta a livello molecolare e lavora grazie a quantità intensive. Il suo operato non è differenziabile dalla sua stessa formazione, i quali si attestano ad un livello ancora pre-sociale, ovvero non ancora statisticamente computabile. L'approccio meccanicistico, per come si è visto, tende invece a opporre i due aspetti. «C'è funzionalismo solo a livello submicroscopico delle macchine desideranti, concatenazioni macchiniche, macchinario di desiderio (engineering); solo qui infatti funzionamento e formazione, uso e montaggio, prodotto e produzione si confondono»³²⁴, cioè laddove funzionamento e formazione si confondono ancora, quindi quando chiedersi «che cosa è?» è ancora chiedersi «come funziona?». I sistemi di sintesi non sono equivalenti a quelli dei grandi sistemi, ancora non significano nulla e sono quello per cui si trasformano, quello che le macchine desideranti compiono in sé stesse.

Quindi, macchina e uomo sono polarità opposte a livello rappresentativo, ma concrezioni di macchine desideranti; solo le loro connessioni e concatenamenti permettono di passare tra i due livelli, descrivendo una parziale dipendenza e al contempo aleatorietà da ascrivere al processo³²⁵. I LLMs non possono che calcolare relazioni già statisticamente pesate e convalidate; la loro è un'organizzazione che si avvicina più alla distribuzione probabilistica attribuibile ad una catena di Markov, invece che all'estrazione casuale di un oggetto come “nel gioco della tombola”. Solo, le regole del gioco non sono poche e semplici come quelle degli scacchi, ma molto più complesse, perché quantitativamente riposano su una pletora di dati che supera ogni limite “umano”.

Il loro piano “molare”, quindi statisticamente convalidabile, in virtù della complessità su cui si appoggia, richiama apparentemente la caoticità rizomatica del “Corpo senza organi”, ossia del campo di forze vitali, sociali, meccaniche, concettuali su cui si situano le macchine desideranti. L'apparenza è sostenuta dalla possibilità di spiegare tecnicamente, attraverso il ricorso ad un algoritmico di pensiero che riflette quella tecnicamente costruita delle apparecchiature, cogliendola come unità strutturale. Si cade vittima così di un approccio meccanicistico, inadatto a cogliere il “rovescio”³²⁶ della struttura.

³²⁴ *Ibid.* p. 324. La citazione a sua volta rimanda al libro *Erewhon* di Samuel Butler.

³²⁵ *Ibid.* p. 327.

³²⁶ I due autori fanno riferimento a Raymond Ruyer e i suoi studi sulla descrivibilità dei processi sociali. E biologici per mezzo delle catene di Markov. Cfr. Ruyer, Raymond, *La genesi delle forme viventi*, Bompiani: Milano, 1966.

Risalire dalle immagini alla struttura sarebbe di scarsa portata, e non ci farebbe uscire dalla rappresentazione, *se la struttura non avesse un rovescio* che è in certo qual modo la produzione reale del desiderio. Questo rovescio è l'«inorganizzazione reale» degli elementi molecolari: oggetti parziali che entrano in sintesi o interazioni indirette [...]; pure molteplicità positive ove tutto è possibile, senza esclusione o negazione, sintesi che operano senza piano, ove le connessioni sono trasversali, le disgiunzioni inclusive, le congiunzioni polivoche, indifferenti al loro supporto.³²⁷

Insomma, il passaggio cruciale è questo: quella che si è definita come “generatività algoritmica” riproduce, per mezzo anche del ricorso tecnico di alcuni artifici³²⁸, parzialmente la stessa riproduzione del ciclo perpetrante la generatività complessa della nostra esperienza linguistica (in parte ciò che Deleuze e Guattari definiscono come “inconscio molecolare”), solo recidendone una parte fondamentale, un serbatoio inesauribile di potenziale per nuove e inattese sintesi. I modelli linguistici non si appoggiano, per ragioni tecniche di descrivibilità algoritmica, a «*connessioni trasversali, disgiunzioni inclusive, congiunzioni polivoche*»; il loro sistema è infatti consistente, ma incompleto, necessitando di un appoggio dal “fuori”, ossia dal rovescio della loro organizzazione strutturale. L'ipotesi di trattare i LLMs è così adeguata a figurarne il funzionamento, ma non completamente. Le «*connessioni globali e specifiche, le disgiunzioni esclusive, le congiunzioni “bivoche”*»³²⁹ sono le vere ramificazioni dell'algoritmo, che essendo un sistema aperto necessita di un appoggio esterno, quindi non-sistematico, a-strutturale. Ecco che prendere in considerazione un contesto di condizioni non strutturali (secondo una modalità che non sia kantiana di legiferazione) è un compito che può essere eseguito dalla schizoanalisi.

L'idea di un concatenamento proviene dall'impossibilità di scindere l'ingranaggio in tanti pezzi. La critica mossa all'identità della macchina rappresentata dalla schizoanalisi ha per oggetto, nel caso dei LLMs, l'idea che siano semplicemente un algoritmo. La tendenza a uniformare l'informazione che viene a veicolarsi è solo un aspetto parziale del problema. Certo, il livellamento probabilistico è una considerazione

³²⁷ «Credevo che gli scambiatori fossero intermediari, che l'interferenza fosse nella frangia, che il traduttore si piazzasse tra le istanze, che il ponte riunisse le due rive, che il percorso andasse dall'origine alla fine. Non ci sono istanze. O piuttosto le istanze, sistemi, rive ecc. sono a loro volta analizzabili in scambiatori, percorsi, traduzioni e così di seguito. Non vi sono istanze o sistemi, solo scatole nere. Quando non comprendiamo. [...] Non vi sono sistemi, istanze, sostanze se non delle nostre ignoranze. Il sistema è il non-sapere. L'altro lato del non-sapere. Il non-sapere ha un lato caos e un lato sistema. Il sapere è un ponte tra queste due rive. Il sapere come tale è uno spazio di trasformazione» Serres, Michel, *Il parassita*, Mimesis: Milano, 2022, p. 101.

³²⁸ *Ibid*, p. 352.

³²⁹ Come le catene di Markov, cfr. parte 1, cap. 1.1.

importante, tanto che si può considerare il motivo principale con cui Weatherby difende le sue tesi e le argomenta. Per raggiungere la positività del campo delle macchine desideranti bisogna muoversi oltre «l'identità immaginaria e l'unità strutturale».

1.0 DALLA LINGUISTICA FORMALE AL DIAGRAMMA

Già dentro il pensiero strutturalista classico esisteva però una “linea di fuga” rispetto all'ipotesi della struttura come forma sincronica e autonoma. Si tratta della *glossematica* di Louis Hjelmslev, pensiero che culmina nell'inversione, così compiuta da Deleuze e Guattari in un *plateau* di *Mille piani*, chiamato *Postulati della linguistica*, per cui la lingua non è un sistema omogeneo ma una macchina astratta che traduce continuamente fra livelli e variabili eterogenee. Attorno agli inizi degli anni '80 Deleuze e Guattari si interrogano sulla natura del codice, con cui lo strutturalismo si era arrogato il diritto di descrivere qualsiasi unità (biologica, antropologica, letteraria, filosofica...). L'ipotesi della struttura come codice autonomo viene formulata secondo quattro postulati, individuati come assiomi della linguistica scientifica, segnatamente quella chomskyana, che recitano:

- 1) Il linguaggio è informativo e comunicativo;
- 2) Esiste una macchina astratta del linguaggio che non si appella ad alcun fattore “estrinseco”;
- 3) Ci sono delle costanti o dei generali del linguaggio che permettono di definirlo come un sistema omogeneo;
- 4) Il linguaggio non si lascia studiare scientificamente che alle condizioni di una lingua standard o “maggiore”.³³⁰

Sono postulati in senso kantiano, presupposti necessari senza cui il linguaggio non si lascerebbe neppure pensare³³¹. Essi non sono solo inutili, ma anche *dannosi* perché «come l'Edipo per Freud, non creano e restringono la creazione riproducendo se stessi senza fine»³³². A ciascun postulato gli autori ne contrappongono uno pragmatico inverso: il linguaggio non è primariamente informativo né comunicativo; sono sempre fattori

³³⁰ *Ibid.* p. 326.

³³¹ Deleuze, Gilles, Guattari, Félix, *Mille piani. Capitalismo e schizofrenia*, trad. it. di G. Passerone, Castelvecchi, Roma, 2017, p. 113; cfr. la formulazione equivalente in Holland, Eugene, W., *Deleuze and Guattari's A Thousand Plateaus*, Bloomsbury, London-New York 2013, p. 65.

³³² Holland, Eugene W., *ibid.*, ricostruisce la pertinenza kantiana del concetto di postulato qui in gioco.

estrinseci a informare la macchina astratta del linguaggio (termine che non ha nessuna valenza internalista); invece che costanti e variabili, il linguaggio si produce per variazioni; la lingua standard è un'invenzione linguistica, ma la stessa contrapposizione lingua maggiore-minore è inadeguata e capziosa³³³. Se per gli autori strutturalisti ogni campo di potere poteva venire annesso a una teoria della comunicazione o in ambito linguistico, ora la linguistica si ritrova ad essere una branca della politica della macchina astratta.

Contrariamente a quanto si potrebbe intendere, il concetto di macchina astratta non è un universale formale, ma il diagramma di un assemblaggio concreto tra concatenamenti collettivi d'enunciazione e macchinici, tra linguaggio e materia³³⁴. Ad ogni parola corrisponde un'operazione pratica tra corpi (azione o passione in termini stoici); allora prima di ogni indicizzazione secondo *bit* del linguaggio, esso si deve rappresentare secondo l'unità minima della *parola d'ordine* [*mot d'ordre* o *slogan*], da cui la linguistica ricava le sue costanti formali. «Definiamo il linguaggio come l'insieme delle parole d'ordine, dei presupposti impliciti, o degli atti di parola correnti in una lingua in un dato momento»³³⁵. Ad ogni enunciato corrisponde un'immediata trasformazione incorporea. È insufficiente qualificare il linguaggio secondo la dicotomia costanti-circostanze³³⁶, a causa delle continue variazioni in atto tra corpi. Ciò significa che ogni enunciato è inseparabile dall'azione che esso compie nel mondo (un enunciato sempre *illocutorio*, secondo la lezione di Austin). Non bisogna quindi intendere la parola d'ordine come fosse un imperativo tra altri tipi di enunciati, ma una categoria che attraversa trasversalmente ogni enunciato

Le parole d'ordine non riguardano soltanto i comandi, ma ogni atto che è legato agli enunciati da una "obbligazione sociale". Ogni enunciato manifesta questo legame, direttamente o indirettamente. Una domanda, una promessa, sono parole d'ordine.³³⁷

Non c'è alcuna naturalità del dato dell'informazione, ma ogni trasmissione prima che essere informativa, mette in gioco delle trasformazioni dei corpi (trasformazioni

³³³ *Ibid.* p. 66 (traduzione mia).

³³⁴ Deleuze, Gilles, Guattari, Félix, op. cit., p. 145.

³³⁵ I primi ad intendere una netta separazione tra corpi e linguaggio furono gli stoici nei termini di "azioni e passioni" dei corpi da un lato e "atti incorporei" dall'altro. Si rimanda alla prefazione di Deleuze, Gilles, *Logica del senso*, Feltrinelli: Milano, 1975.

³³⁶ *Ibid.* p. 116. L'unità minima della parola d'ordine porta di conseguenza la politica della lingua a instaurarsi più in alto della linguistica.

³³⁷ Il termine "contesto" è ritenuto dagli autori troppo largo, una parola-ombrello entro cui far ricadere molte occorrenze.

incorporee perché fisicamente non ravvisabili, ma comunque impattanti sui corpi) assegnando loro azioni e attributi. Anche Weatherby e Jakobson ritengono che il linguaggio non sia primariamente informativo, ma fatico, quindi poetico. La funzione fatica di Jakobson, infatti, non prende a riferimento la comunicazione, ma il messaggio stesso nella sua costituzione sintattica. Questo è però l'anello di congiunzione che lega entrambi gli autori alla teoria dell'informazione di Shannon. Come si è visto, è la ridondanza del linguaggio a permetterne la propagazione dell'informazione. Quindi, se è vero che per ogni informazione veicolata corrisponde una sua reiterazione poetica e non il contrario, è pur vero che il loro presupposto, poggiando sulla ridondanza del linguaggio, non si presenta più radicale e ampio rispetto all'estensione della parola d'ordine secondo Deleuze e Guattari.

Nel campo sociale, composto a livello molecolare da intensità, si situano i concatenamenti, concetti su cui possono costruirsi delle strutture molari. La macchina astratta del linguaggio, poggiando immediatamente su un concatenamento, annulla ogni distinzione tra soggetto enunciato e soggetto dell'enunciato. Prima dell'attribuzione di identità viene il concatenamento di enunciati e corpi macchinici

Non c'è primato dell'individuo; c'è invece l'inscindibilità di un Astratto singolare e di un Concreto collettivo. La macchina astratta non esiste indipendentemente dall'assemblaggio, non più di quanto l'assemblaggio funzioni indipendentemente dalla macchina astratta.³³⁸

È la linguistica di Hjelmslev a seguire la dicotomia tra corpi e parole, tra sostanza e forma, quindi tra "assemblaggio macchinico" e "macchina astratta". Attenzione però, se ci si limitasse a instaurare una simile dicotomia come un parallelismo non ci si allontanerebbe dalle fila del pensiero strutturalista saussuriano. Il segno non è più libero di aleggiare sulle cose, non ne è svincolato, come se fosse il significato di un significante ad esso slegato. Il concatenamento d'enunciazione, la macchina astratta, non parla mai delle cose; invece di un legame di referenza, è meglio parlare di un'"insistenza" delle parole sulle cose. Vengono a prodursi delle linee di interferenza tra parole e cose. L'ipotesi quindi di una combinatoria simbolico-strutturale rimane un'astratta e incompleta supposizione. *«Il diagramma non è più l'archivio, auditivo o visivo, ma la carta, la cartografia coestensiva a tutto il campo sociale. è una macchina astratta che, definita da funzioni e da materie*

³³⁸ Holland, Eugene W., op. cit., p. 66, traduce dal corrispondente passo di Deleuze, Gilles, Guattari, Félix, *Mille piani. Capitalismo e Schizofrenia*, Castelveccchi: Roma, 2017, p. 116.

*informali, ignora ogni distinzione formale tra contenuto e espressione, tra formazione discorsiva e formazione non-discorsiva*³³⁹». Tra espressione e contenuto allora c'è presupposizione vicendevole e i due si rilanciano senza rapporto di causalità. Come mette in luce Lazzarato³⁴⁰, in questa maniera si sfugge sia all'argomentazione marxista, che ritiene l'espressione dipendere dal contenuto, ma anche dallo strutturalismo che fa del contenuto un derivato dell'espressione simbolica. Deleuze e Guattari ritengono che gli enti contengono già i loro attributi, secondo dei rapporti variabili e mai stabili. La fluttuazione si attesta secondo delle intensità, dei gradi di deterritorializzazione tra espressione e contenuto.

L'assemblaggio si articola secondo due assi, come se fosse su un piano cartesiano. Un asse orizzontale distingue segmenti di contenuto ed espressione (in termini hjelmsleviani: gli aggregati di corpi, gesti, oggetti, da una parte; gli enunciati, i segni, dall'altra), ricavando rispettivamente dei concatenamenti macchinici e di enunciazione. Sull'asse verticale invece si pongono i gradi di territorializzazione e deterritorializzazione. La coppia concettuale territorializzazione-deterritorializzazione viene ripresa dal lessico lacaniano per descrivere l'instaurazione di territori nella psiche infantile secondo intensità emotive ed erotiche³⁴¹. Il territorio indica perciò una situazione spazio-temporale a cui viene associata una carica di desiderio, libido; si intuirà invece come la deterritorializzazione costituisce invece una liberazione dei flussi del desiderio dal loro posizionamento³⁴². La macchina astratta non è una condizione formale a priori che governa il sistema ma «*descrivere (senza determinarla) la variazione che le variabili subiscono*»; le sue regole non sono invarianti ma «*opzionali, e variano di continuo con la variazione stessa, come in un gioco in cui ogni mossa cambia le regole*»³⁴³. La nozione di macchina linguistica astratta sostituisce l'idea di una struttura linguistica autonoma di tipo saussuriano. Il problema dell'impianto strutturalistico della linguistica classica è che è «*troppo astratto senza esserlo abbastanza*»³⁴⁴: il modello resta troppo lineare e ogni variabile viene astrattamente relegata nell'insieme delle "circostanze" di tipo pragmatico rispetto alle invarianti linguistiche. La macchina astratta, detta anche *diagramma*, invece

³³⁹ Holland, Eugene W., op. cit., p. 78.

³⁴⁰ Deleuze, Gilles, *Foucault*, Orthotes: Napoli, 2018, pp. 48-49.

³⁴¹ Lazzarato, Maurizio, *Segni e macchine*, Ombre corte: Verona, 2019, p. 88.

³⁴² Il bambino durante la fase di allattamento inizia ad associare la zona del seno materno alla soddisfazione secondo nutrimento, concatenandola all'organo della bocca. Il seno viene a territorializzarsi come evento a cui agganciarsi con la bocca, richiamando una serie di stati d'animo (la quiete che succede al momento di fame dovuta al soddisfacimento). Cfr. Holland, Eugene W., *Deleuze and Guattari's Anti-Oedipus, introduction to Schizoanalysis*, Routledge: London, 1999, p. 29.

³⁴³ Anche qui Holland ritrova un'origine del termine, che verrebbe a riferirsi all'istituzione delle *enclosures* ai contadini inglesi, raccontata da Marx ne *Il Capitale*. Cfr. Holland, Eugene W., *Ibid.* p. 30.

³⁴⁴ Deleuze, Gilles, Guattari, Félix, *Mille piani*, p. 145.

non è puramente linguistica, ma neanche solo materiale; essa assume un'importanza centrale in virtù della sua natura virtuale³⁴⁵ e immanente³⁴⁶. Il concetto diagrammatico di macchina astratta riesce a superare la linearità entro cui ancora la linguistica e la teoria dell'informazione versano, ancorandosi alla complessità del concetto di causa immanente³⁴⁷.

Se il linguaggio è già di per sé un'astrazione separata dall'assemblaggio entro cui funziona, allora isolare i vettori-parola dei modelli e vederli come una struttura simbolica autonoma significa cadere proprio nel primo postulato che Deleuze e Guattari smontano. L'operazione sembra però essere condotta da Weatherby, soprattutto nel primo capitolo di *Language Machines*, in cui nota che sì, l'ipotesi distribuzionalista non basta a spiegare le generazione di senso e che bisogna così integrare «poetica, ideologia, struttura»³⁴⁸; restando però entro la cornice strutturalista, integrando cioè Jakobson in una “poetica generale” non si sporge oltre la struttura, delimitando la problematica a livello algoritmico. Il rischio allora che si presenta è duplice. Da un alto, si riconduce la dimensione pragmatica e politica dei LLMs a un epifenomeno della loro forma simbolica; le ricadute di ChatGPT o Claude sull'opinione pubblica o sulle scelte del consumatore verrebbero ad essere degli effetti collaterali di un'innovazione formale interna alla teoria del linguaggio combinato ai modelli informatici a rete. Dall'altro, la poetica generale non rileva alcun interesse nello studio dell'eterogeneità dei tasselli del concatenamento dei LLMs: flussi di dati, le procedure di *fine-tuning* intrecciati a interessi corporativi, le scelte di *Reinforcement Learning from Human Feedback* (RLHF) vengono semplicemente menzionate per l'influenza che possono avere a livello di operatività statistica³⁴⁹, mentre questioni come la nascita di un nuovo sotto-proletariato nel Sud globale, le dipendenze elettriche e idriche dei *data center*, le strategie commerciali di OpenAI, Meta, Anthropic,

³⁴⁵ *Ibid.*

³⁴⁶ Sarebbe allora interessante affrontare uno studio della trasformazione del concetto di virtuale secondo il concatenamento macchina-corpo senza organi-diagramma, che attraversi tutte le fasi del pensiero deleuziano e guattariano.

³⁴⁷ Deleuze, Gilles, *Foucault*, Orthotes: Napoli, 2018. Il diagramma è una rete di rapporti di forze sovrasensibili, nel senso di una differenza di natura rispetto agli enti e le forme che mettono in comunicazione. Perciò il diagramma è *informale*: «così il diagramma sovrasensibile non si confonde con l'archivio audiovisivo: è come l'a priori che la formazione storica presuppone» *ibid.* p. 101. In sé quindi il diagramma non può venire percepito, ma solo gli strati che esibisce e riordina (nel caso di Foucault l'archivio di forme visivo-oggettuali e i corrispondenti discorsi per parlarne). Il diagramma è informale, non-stratificato, trascendentale nella misura in cui non è empiricamente sondabile e forma le condizioni di verità di una problematica (*fuori* è il termine che riguarda tutto ciò che sfugge al pensiero e che vi si scontra)

³⁴⁸ È ormai noto il legame che stringe Deleuze al “principe dei filosofi”, Spinoza. Sin dal testo di *Spinoza e il problema dell'espressione*, fino a *Che cos'è la filosofia?* egli rimane un riferimento costante per l'immanentismo deleuziano.

³⁴⁹ Weatherby, Leif, *Language Machines*, Minneapolis press: London, 2025, p. 1: «la programmazione non coglie in primo luogo la grammatica, la sintassi o la logica del linguaggio [...] bensì aspetti più densi e più ampi del linguaggio che si rivelano più fondamentali: la poetica, l'ideologia, la struttura».

Google non vengono neanche trattate. Tutti questi apparati di macchine non sono riducibili al piano dei segni e delle frequenze, eppure sono parte dell'unità minima d'analisi che ci si dovrebbe dare.

Di conseguenza, si può additare un primo motivo per cui è necessario discostarsi dal lessico della struttura: la macchina astratta del linguaggio chiusa in sé stessa potrebbe dirsi una finzione metodologica, utile per chi è interessato a descrivere le regolarità statistiche di calcolo, escludendo però tutto ciò che aiuti a capire come, perché certi enunciati si producono, circolano e modificano i comportamenti. Ad esempio, la nozione di architettura *transformer* secondo il meccanismo di *attention* è eterarchica (per usare il termine di McCulloch già evocato): essa non fa che misurare quanto pesa, in una sequenza di parole, il rapporto tra un *token* e l'altro, indipendentemente dalla sua posizione, quindi indipendentemente dal processo di sedimentazione di tali *token*: il loro assorbimento entro data set massivi per la fase di *training* oppure l'iterazione di *fine-tuning*, a loro volta dovuti alla dimensione pragmatica umana che li ha costituiti prima di tradurli in vettori: la distribuzione statistica secondo *layer* di pesi non è puramente un'operazione ingegneristica, ma assorbe pratiche di tribunali, redazioni, *social network*, biblioteche, contratti di lavoro precari di Mechanical Turk e di Sama e protocolli di moderazione delle imprese³⁵⁰. La funzione poetica, nonostante possa apportare importanti riflessioni sullo statuto del discorso della filosofia del linguaggio, non è che una superficie statistica di un assemblaggio molto più esteso.

General purpose technologies

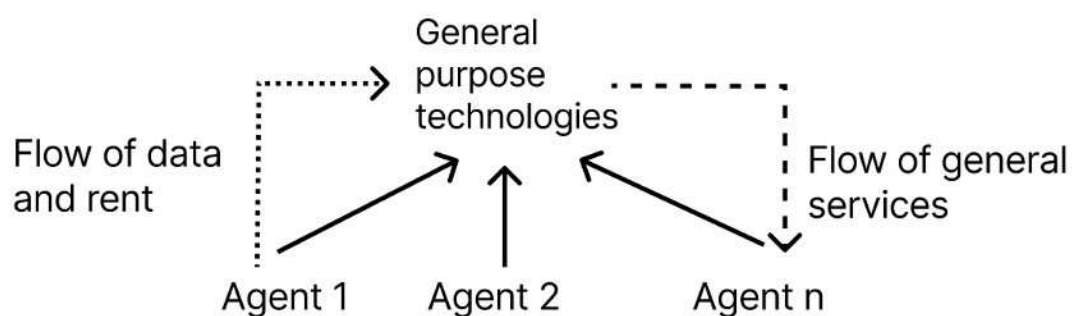


Figura 12 ricircolo continuo dell'informazione tramite LLM (general purpose technology)³⁵¹

³⁵⁰ Già Kockelman, seppur entro un testo più succinto, ne parla dandone più importanza. Cfr. Kockelman, Paul, *Last Words*, Prickly Paradigmatic Press: Chicago, 2024.

³⁵¹ Sullo sfruttamento salariale nel Sud globale cfr. Perrigo, Billy, *OpenAi Used Kenyan Workes on Less Than 2\$ Per Hour to Make ChatGPT Less Toxic*, in «Time», 18 gennaio 2023; cfr. anche Gray, Mary L., Siddharth, Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*,

Inoltre, il concetto di struttura per come viene ripreso da Weatherby sulla base dell'ipotesi differenziale di Saussure non ha alcuna attinenza genetica, ma si limita a far coincidere la sincronicità della struttura con la sua dimensione diacronica. Per Deleuze invece la struttura è già un'idea non statica: «*il simbolico come elemento della struttura è al principio di una genesi: la struttura si incarna in realtà ed immagini seguendo serie determinabili; ancor di più essa le costituisce incarnandovisi, ma non ne deriva, essendo più profonda di esse*»³⁵². L'idea non statica, ma dinamica di una genesi³⁵³, è totalmente altra rispetto all'ipotesi formale dei suoi elementi simbolici. A sostituirsi agli elementi formali di una combinatoria logica è una successione di eventi incarnati e storicamente attendibili. La formulazione non esibisce le condizioni che rendono possibile questa incarnazione, ovvero il fatto che la struttura è già sempre catturata dentro un assemblaggio concreto in cui macchine non semiotiche (corpi, dispositivi, edifici, schede elettroniche, dati, capitali) vengono a intrecciarsi con macchine semiotiche (catene di enunciati, segni, parole, simboli matematici). Lo strutturalismo classico, stabilendo l'autonomia tra simbolico e reale, tratta ancora la materia e il significato come due sfere distinte, ne viene annullata la distinzione, ma non la differenza con il concetto di macchina come «*sistema di tagli e flussi*»³⁵⁴. A questo proposito neanche la dicotomia dei formalisti russi di materiale e artificio riesce a superare quella di materia e forma, ritenendo ancora come separate le dimensioni del significante e del significato dai loro processi di lavorazione (artifici).

Infine, si potrebbe dire che un LLM dispiega due piani che non sono pienamente sovrapponibili: il piano dei *flussi* (dati di addestramento, segnali elettrici, energia) e un piano di *codici* (la distribuzione statistica dei vettori, i pesi dei modelli a rete, gli output testuali). La poetica generale di Weatherby secondo questa dicotomia procede verso una formalizzazione del codice.

A questo punto appare utile avanzare due considerazioni: una di carattere storico-filosofico, l'altra di carattere linguistico. Iniziando dalla prima, si tenga in considerazione che nel corso del presente testo si sono seguite alternativamente due

Houghton Mifflin Harcourt: Boston, 2019.

³⁵² Immagine tratta dal *keynote* di T. Guariento al convegno *The World Algorithm*, <https://www.unive.it/web/en/16571/home#c182454>, 2 febbraio 2026.

³⁵³ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo?*, SE: Milano, 2008.

³⁵⁴ La genesi dinamica indaga il processo psicogenetico che giugne dalle intensità degli oggetti parziali, quindi dalle "sonorità", attraversando lo sviluppo della corporeità e della voce, fino alla superficie incorporea del senso e del discorso. La genesi statica invece mira a una mappatura della superficie del senso, chiedendosi come il senso si determini secondo denominazione, significazione e manifestazione.

direzioni semiotiche: la differenziabilità dei segni per Saussure, testimoniata da quella che Eco definisce come una «dialettica di presenza e assenza»³⁵⁵ della struttura. Il segno è perciò un'unità tanto arbitrariamente definibile quanto cogente secondo l'intreccio che ha con gli altri segni. Vi è poi il ragionamento peirciano della triade di segno-oggetto-interpretante, in cui la differenza tra segno e interpretante non è negativa, ossia supportata da un'assenza, ma è generativa. Ora, scegliere di appoggiarsi a testi come l'*Anti-Edipo* o *Mille piani* radicalizzerà sicuramente la propensione per la direzione semiotica peirciana, avendo in mente quanto l'autore stia a cuore a Deleuze. La scelta appare però giustificata, indicando il funzionamento di un LLM non come un puro gioco di differenze, ma una catena aperta e produttiva di segni. Inoltre, bisogna sottolineare che il ragionamento di Weatherby sia utile a evidenziare la natura linguistica di queste macchine, detta anche "letteraria", di certo criticando le tendenze cognitive non di poco conto per il dibattito attuale sull'AI. Le macchine sarebbero invece "poetiche" in virtù della generazione del token successivo per mezzo dell'*embedding* vettoriale, che ricalca l'idea jakobsiana della funzione poetica. Allora è vero che quello che i LLMs generano è in senso letterale poesia, perché il loro funzionamento ricalca quello che Jakobson aveva identificato essere il motore del linguaggio poetico. Però, l'istituzione della poetica generale rischia di perdere la commistione e la coesistenza delle funzioni indicate da Jakobson. Weatherby, dopo averle nominate, ricalcando il diagramma di Shannon³⁵⁶, le accantona privilegiando la funzione poetica, quasi da farne il fondamento dei LLMs. Quasi unilaterale, l'operazione fa subordinare le altre a quella poetica, rispecchiando l'illusione che la macchina sia chiusa nel suo universo di segni differenziali. A questo si può contrapporre la semiotica mista di Guattari, utile strumento per mostrare la parziale coincidenza e sovrapposizione di regimi semiotici e di assemblaggi pragmatici. Si potrebbe dire allora che la poeticità dei LLMs è inseparabile dalle funzioni:

- *Referenziale* (gli output sono usati per ottenere informazioni);
- *Conativa* (sono usati per ottenere azioni);
- *Fatica* (mantengono il canale di interazione aperto, sempre proponendo ulteriori proposte);

³⁵⁵ Deleuze, Gilles, Guattari, Félix, *L'Anti-Edipo. Capitalismo e schizofrenia*, Einaudi: Torino, 1975, p. 41; cfr. anche Holland, Eugene W., *Deleuze and Guattari's Anti-Oedipus. Introduction to Schizoanalysis*, Routledge: London, 1999, pp. 25-26.

³⁵⁶ La presenza di un segno è tale solo se convalidata tramite l'assenza di tutti gli altri segni. Ogni presenza si può dare solo concedendo un'assenza, assenza che Lacan ontologizza facendone il motivo della mancanza costitutiva dell'uomo. Cfr. Eco, Umberto, *La struttura assente*, Bompiani: Milano, 1971.

- *Metalinguistica* (gli utenti chiedono ai modelli di spiegare i propri output);
- *Emotiva* (i LLMs rispondono in modo che simula empatia o entusiasmo).

Certo, Weatherby compie una mirabile operazione di chiarificazione, esplicitando la radice poetica di questi aspetti, ma trascurandone però l'apporto che rivelano avere sull'utente. Si è dimostrato allora che le implicazioni sostenute dall'argomentazione strutturalista sono riduttive e interrogano il fenomeno dell'AI escludendo diversi fenomeni ad esso collegati; si è indicato di converso il concatenamento come concetto adeguato per inquadrare la complessità dei LLMs. È ora quindi di cercare di approfondirlo ulteriormente per capire sotto quali aspetti il suo utilizzo risulti essere vantaggioso.

CAP. 2: CORPO SENZA ORGANI COMPUTAZIONALE

2.0 CONCATENAMENTO UTENTE-LLM-CORPORATE

Prendendo un esempio semplice, quotidiano, quasi triviale, si può dimostrare come la triangolazione utente-LLM-azienda produttrice sia un concatenamento complesso. L'esempio compete alla digitazione di un prompt, ossia quando un utente apre un'interfaccia testuale, scrive un testo e riceve una risposta. In questo caso la descrizione strutturalista si limita a considerare cosa succede tra prompt e risposta: la macchina opera la previsione del token successivo, calcola le distribuzioni di probabilità, seleziona il completamente più verosimile e lo restituisce come testo. La scena però non può venire inquadrata solamente tra utente e macchina, poiché la macchina non è una singolarità

autonoma: è la sedimentazione di un *training* condotto da milioni di esseri umani, l'oggetto di un *fine-tuning* mediante i valori di *honest*, *harmlessness*, *honesty* dell'azienda produttrice; inoltre, bisognerebbe vederla anche a partire dagli accordi commerciali, dalle *partnership* istituzionali o dai vincoli giuridici o politici³⁵⁷. Allo stesso tempo l'utente non è un soggetto monolitico e compatto, a sua volta un assemblaggio di macchine desideranti, come lo è il LLM e l'azienda *corporate*. L'utente, ad esempio, muove le dita sulla tastiera³⁵⁸, l'occhio sul monitor, il flusso di parole sulla pagina, il cervello sovrintende cognitivamente, l'azienda di telecomunicazione media assieme ai *server* di Microsoft Azure che ospitano OpenAI, alimentati dalle centrali elettriche, gli esseri umani in Kenya che hanno annotato i dati di addestramento³⁵⁹.

Ora, cos'è una macchina desiderante? Prima si è parlato di macchina astratta, ma le due nozioni sono diverse non solo concettualmente, ma anche a partire dal contesto in cui sono state parterite³⁶⁰. La macchina, come si è visto nel paragrafo precedente, funge da alternativa concettuale alle tesi vitalistiche e meccanicistiche, garantendo un concetto operativo atto a rimarcare un carattere di finitudine ed esaustione che la struttura non manifesta. Il desiderio è a sua volta da intendersi nel senso materialistico promosso dalla schizoanalisi: non come una pulsione soggettiva, mossa da un bisogno o da una mancanza³⁶¹, ma come forza impersonale, preindividuale, di natura libidinale-economica che ha per oggetto un concatenamento macchinico. Il piano del desiderio è quindi un concreto assemblaggio di macchine, articolato in oggetti parziali e tagli sui flussi.

Se il desiderio produce, produce del reale. Se il desiderio è produttore, non può esserlo se non in realtà, e di realtà. Il desiderio è l'insieme di *sintesi passive* che macchinano gli oggetti parziali, i flussi e i corpi, e che funzionano come unità di produzione. Il reale ne deriva, è il risultato delle sintesi passive del desiderio come autoproduzione

³⁵⁷ Weatherby, Leif, *Language Machines*, p. 155.

³⁵⁸ Sul *Reinforcement Learning from Human Feedback* nei modelli linguistici di OpenAI cfr. Ouyang et al., *Training Language Models to Follow Instructions with Human Feedback*, in «Advances in Neural Information Processing Systems», 35, 2022. Mentre per una discussione sui valori HHH (*helpful*, *harmless*, *honest*) cfr. Askell, Amanda et al., *A General Language Assistant as a Laboratory for Alignment*, arXiv:2112.00861, 2021.

³⁵⁹ Un aspetto spesso tralasciato delle nuove tecnologie è la loro digitalizzazione, intesa non tanto come programmazione di dati informatici, ma radicale riduzione della mano sotto forma di dito. Il digitale è l'impoverimento della praticità sottoforma di un regime di opzioni, nodi di una rete per fornire delle informazioni prestabilite. Per quanto riguarda il rapporto mano-occhio cfr. Deleuze, Gilles, *Sulla pittura*, Einaudi: Torino, 2024.

³⁶⁰ Sul ruolo dei lavoratori kenioti nell'etichettatura dei dati di Open AI cfr. Berrigo, Billy, *OpenAI Used Kenyan Workers on Less Than \$2 Per Hour*, in «Time», Gen. 18, 2023.

³⁶¹ Macchina desiderante è il primo dei due concetti ad essere stato pensato e se ne trova traccia ne l' *Anti-Edipo* cfr. Deleuze, Gilles, Guattari, Félix, *Anti-Edipo*, Einaudi: Torino, 2006, pp. 3-55.

dell'inconscio. Il desiderio non manca di nulla, non manca del suo oggetto. È piuttosto il soggetto che manca al desiderio, o il desiderio che manca di soggetto fisso; non c'è soggetto fisso che per la repressione. Il desiderio e il suo oggetto sono un'unica cosa, sono la macchina, in quanto macchina di macchina.³⁶²

Il desiderio quindi non è semplicemente l'intenzionalità del soggetto cosciente, ma l'intero investimento libidinale nel campo sociale³⁶³: ad esempio l'utente potrebbe aver bisogno di compagnia (molti utenti dialogano con i LLM per ragioni di solitudine) oppure di approvazione (i modelli sono generalmente compiacenti) o di sicurezza (la *response* fornisce una rassicurazione cognitiva ed emotiva), intrattenimento oppure anche di efficienza (per svolgere mansioni che richiederebbero più tempo velocemente). Le aziende sfruttano questi bisogni, disegnando le interazioni con il prodotto che venga riconosciuto essere utile, innocuo e onesto (HHH), ma soprattutto piacevole, restituendo un'immagine di sé soddisfatta all'utente. Si tende allora a incanalare i flussi libidinali entro il concatenamento utente-LLM-azienda produttiva. Da questo punto di vista la lettura di Deleuze e Guattari supera anche quella del capitalismo della sorveglianza, che ritiene il *behavioural surplus* come ciò che le piattaforme estraggono mentre offrono un servizio. Il *data mining* è strumento di previsione e modificazione comportamentale. Ebbene, si potrebbe dire che il *data mining* non si limita a collezionare dei dati a fini commerciali e statici, ma iscrive i flussi di desiderio su un corpo senza organi computazionale, rendendo il surplus libidinale, prima che informazionale, come si cercherà di dimostrare più avanti³⁶⁴.

Ogni macchina non può mai essere venire scissa dalle connessioni che mette in gioco con le altre; perciò, ogni macchina è sempre macchina di macchina: macchina-flusso e macchina-taglio. Il taglio è l'operazione che sintetizza, creando un'unità discreta sul fondo di una continuità di flussi di materia ed energia³⁶⁵. «Le macchine desideranti sono macchine binarie, a regola binaria o regime associativo; sempre una macchina accoppiata con un'altra»³⁶⁶. Si riconoscono allora quattro flussi:

³⁶² «Non appena poniamo il desiderio dalla parte dell'acquisizione, ci facciamo del desiderio una concezione idealistica (dialettica, nichilistica) che lo determina in primo luogo come mancanza [*manque*], mancanza d'oggetto, mancanza dell'oggetto reale» *Ibid.* p. 27.

³⁶³ *Ibid.* p. 29.

³⁶⁴ La schizoanalisi attua una presa sulle connessioni a livello sociale, prima ancora che "famigliare", termine che indica la rappresentazione di entità astratte come la famiglia, le istanze di potere dell'azienda. «Gli investimenti sociali sono primari rispetto agli investimenti familiari, che nascono solo dall'applicazione od al ripiegamento dei primi». *Ibid.* p. 312.

³⁶⁵ Con "capitalismo della sorveglianza" il riferimento va al celebre saggio di Zuboff. Cfr. Zuboff, Susan, *Il capitalismo della sorveglianza*, IUSS press: Roma, 2025.

³⁶⁶ «Ogni "oggetto" presuppone la continuità di un flusso, ogni flusso la frammentazione dell'oggetto» *ibid.*

- 1) I flussi di parole (token in entrata e in uscita);
- 2) I flussi di forza-lavoro (gli interventi di etichettatura, *fine-tuning*, *red-teaming* compiuti da esseri umani);
- 3) I flussi di capitale (miliardi di dollari di investimenti finanziari, quote di mercato, le *revenue* da abbonamenti);
- 4) I flussi di energia (kilowattora, l'acqua per il raffreddamento, i materiali dell'hardware).

Ogni macchina-organo taglia il proprio flusso, attraverso cui interpreta il mondo (ad esempio, la macchina occhio lo interpreta in termini di vedere, il LLM in termini di vettorializzazione semantica) e produce delle sintesi connettive con le altre macchine, concatenandosi ad esse. Ciascun flusso è confuso con gli altri, poiché ogni taglio viene operato da una macchina su un altro flusso. I LLMs si possono pensare a partire dall'intreccio che opera su di questi flussi. Il modello OpenAI è un grande sistema di tagli su flussi di lavoro umano (etichettatura nel Sud Globale e *prompt engineering* negli Stati Uniti), su flussi di parole (il *data set*), su flussi di capitale di investimento e su flussi di energia (data center)³⁶⁷.

Il regime di macchine viene a crearsi sulla scorta delle tre sintesi passive. Per quanto riguarda la prima delle tre, la sintesi connettiva, essa fa sì che ogni macchina s'innesti su un'altra (flusso o organo), instaurando una connessione *trasversale* tra varie serie binarie e lineari in tutte le direzioni. È Guattari a introdurre il concetto di trasversalità, inizialmente applicandolo all'ambito clinico³⁶⁸, per indicare un agglomerato sincretico di forze eterogenee che attraversa più registri (familiare, istituzionale, economico, sessuale, linguistico) senza ridurli ad alcun piano comune. La riflessione attorno al concetto di territorio trasversale serve a Guattari per costruire un modello che mobiliti le istituzioni, che le renda maggiormente flessibili; così, a contorno del territorio soggettivo, sempre situato, v'è sempre un "territorio trasversale" che serve da riserva inesauribile di potenziale, come spazio liscio, creando nuovi territori per sfuggire all'assoggettamento delle "cartografie" istituzionali. «*La trasversalità è il legame con l'inconscio del gruppo, ciò che va oltre le leggi oggettive su cui esso si fonda; il supporto*

p. 7.

³⁶⁷ *Ibid.*

³⁶⁸ Per una tipologia dei flussi con componenti economiche, politiche, sociologiche si rimanda a Guattari, Félix, *La rivoluzione molecolare. La nuova lotta di classe*, PGreco: Milano, 2017.

dell'essere del gruppo»³⁶⁹.

La trasversalità è dovuta alla compresenza delle sintesi connettive che si attivano secondo una binarietà di macchine organo. Ogni oggetto parziale (macchina organo) emette un flusso e questo flusso è associato ad un'altra macchina per la quale definisce un campo di presenza potenziale anch'esso molteplice. Ogni connessione [*coupage*] oggetto-flusso funziona secondo un'intensità che influisce direttamente sulle altre coppie. «Sono flussi per così dire a due teste attraverso cui si effettua ogni connessione produttrice così come abbiamo cercato di renderne conto con le nozioni di flusso-schizo o di taglio-flusso»³⁷⁰. Per questo viene indicata come una serie di congiunzioni e... e...e. Assieme ad essa si presentano altre due sintesi: la sintesi disgiuntiva di registrazione e la sintesi di congiuntiva.

Ora, per quanto riguarda la paradossale “disgiunzione inclusiva”, essa è concepibile come superficie di registrazione dei concatenamenti tra macchine. Viene definita come istanza di anti-produzione, poiché essa non produce, ma regola l'intensità della produzione da parte della prima serie di sintesi passive. Registrare significa codificare secondo una certa modalità o...o...o.... Sono tutte disgiunzioni che vengono accolte simultaneamente nella superficie su cui giacciono le macchine desideranti: il corpo senza organi. Ripreso da Antonin Artaud, il presente concetto è tanto centrale, quanto complesso e ambiguo. Esso funziona da superficie di iscrizione dei segni, quindi dei prodotti delle macchine desideranti. Esso viene a costituirsi come “quasi-causa” poiché funge da istanza mitica di spiegazione della creazione dei prodotti; detto in altri termini: il prodotto-merce non viene riconosciuto come trasformazione di flussi di desiderio e per via della sincronicità del lavoro libidinale delle macchine, ma viene applicata spesso una spiegazione rappresentativa, per mezzo cioè del richiamo a una causa (che sia la scrittura individuale di prompt, l'accumulo originario di una data set massivo, l'investimento di grandi fondi per la creazione di gruppi di ricerca, le innovazioni interne al progresso tecnologico degli ultimi vent'anni)³⁷¹.

Ma l'essenziale è la costituzione d'una superficie incantata di iscrizione e di registrazione

³⁶⁹ Per approfondire le riflessioni di Guattari fino alla composizione *dell'Anti-Edipo* si rimanda a Guattari, Félix, *Psychoanalysis and transversality*, Semiotext(e): Los Angeles, 2015. Altra risorsa importante per il concetto di trasversalità e per mostrare la predilezione di Guattari per la complessità dei saperi e la sua conseguente transdisciplinarietà cfr. Genosko, Gary, *Félix Guattari. An Aberrant Introduction*, Continuum: London-New York 2002, pp. 67-95.

³⁷⁰ Dosse, Françoise, *Gilles Deleuze, Félix Guattari: une biographie croisée*, Éditions La Découverte: Paris, 2009, p. 80.

³⁷¹ Deleuze, Gilles, Guattari, Félix, *Anti-Edipo*, Einaudi: Torino, 1975, p. 371.

che si attribuisce tutte le forze produttive e gli organi di produzione, e che agisce come quasi-causa comunicando loro il movimento apparente (il feticcio).³⁷²

La circolarità del rapporto tra macchine desideranti e corpo senza organi consiste in una sovrapposizione delle sintesi disgiuntive a quelle produttive. La scelta ricade su una disgiunzione che è virtualmente connessa a tutte le altre, funzionando così da quasi-causa, spiegazione del fenomeno in maniera retroattiva: il prodotto della sintesi delle macchine desideranti viene preso come fonte di desiderio, quindi l'effetto diventa causa. Infine, la sintesi congiuntiva del soggetto consiste nella creazione residuale di uno «strano» soggetto, perché senza fissa identità³⁷³. Si capisce da passaggi come questo perché testi come l'*Anti-Edipo* costituiscono una forte critica alla tradizionale visione psicanalitica del desiderio: la libido non è una cieca pulsione che risponde al bisogno di un soggetto; il desiderio non è soggettivistico, ovvero promosso dal soggetto, ma è una forza transindividuale, ovvero che sfugge all'Io come all'Altro, che muove le macchine di cui si sta parlando. Il soggetto si ritrova a valle del processo, al suo compimento, come effetto. Egli, consumando l'oggetto del desiderio, plasmerà la propria identità secondo un'operazione retroattiva, credendosi l'autore del processo. «Quindi questo sono io!»³⁷⁴ è la formula usata per spiegare la formazione del soggetto *ex post*; il “me” dell'utente non preesiste alla scena, ma emerge dal modo in cui le sintesi sono organizzate³⁷⁵.

Ecco allora che si spiega perché un concatenamento non permetta ad alcuna individuazione di sovrintendere il processo delle molteplicità che vi prendono parte. Ognio assemblaggio è infatti un concatenamento *collettivo* d'enunciazione. Nel capitolo “*I postulati della linguistica*” di *Mille piani*, Deleuze e Guattari usano il termine *agencement collectif d'énonciation*³⁷⁶ per asserire che ogni enunciazione è già sempre collettiva, anche quando individuata nella voce di un soggetto. «L' “io” è esso stesso una parola d'ordine. L' “io” non può fondare le parole d'ordine. Esiste solo il branco, solo la molteplicità. L'individuo non dice “io”. Solo un insieme collettivo di enunciati può dire

³⁷² Ancora una volta il richiamo a Spinoza è necessario. Come Spinoza costruisce il suo impianto immanentistico sul Dio-sostanza di cui ogni ente costituisce una modalità, così per gli autori dell'*Anti-Edipo* ogni macchina desiderante è un modo (nel senso spinoziano) del corpo-senza-organi. Tra macchine desideranti e corpo-senza-organi c'è quindi un rapporto di causalità immanente.

³⁷³ Deleuze, Gilles, Guattari, Félix, op. cit., Einaudi: Torino, 1975, p. 13.

³⁷⁴ La molteplicità delle forme identitarie passa attraverso le linee intensive del corpo senza organi, quindi dei suoi gradienti: a diversi concatenamenti corrisponderanno diversi soggetti. Per un approfondimento di queste tematiche si rimanda a Holland, Eugene, W., *Deleuze and Guattari's Anti-Oedipus*, Routledge: London, 1999, pp. 33-36.

³⁷⁵ Deleuze, Gilles, Guattari, Félix, op. cit., Einaudi: Torino, 1975, p. 23.

³⁷⁶ «Quello sono io, quindi quello appartiene a me!» cfr. Holland, Eugene, W., *Deleuze and Guattari's Anti-Oedipus*, Routledge: London, 1999, pp. 34.

“io”»³⁷⁷. La *response* di un LLM non è la singola voce della macchina, così il prompt non è quello del soggetto-utente: essa sarà piuttosto un concatenamento collettivo d’enunciazione che integra utente, il LLM, l’azienda produttrice, la pluralità di parlanti che hanno alimentato il corpus, gli annotatori che hanno guidato il *fine-tuning*, il regime di legge che presiede alle politiche di moderazione, le scelte “etiche” che hanno selezionato certi testi piuttosto che altri. Quindi, quando il modello fornisce una risposta, l’utente è portato a sentirla come “sua”, ovvero attribuirla implicitamente a sé³⁷⁸, secondo quanto affermato dal ragionamento «Quindi questo sono io!». La dipendenza ad adoperare un LLM è parzialmente dovuta alla credenza che le sue risposte contribuiscono a rafforzare: che il modello stia dialogando proprio con un soggetto definibile, nominabile, a cui si possono muovere critiche, complimenti. Insomma, l’inclinazione personalistica del *response* contribuisce ad accrescere la dipendenza con un modello che sembra predisporre all’ascolto di qualcuno in particolare, un soggetto costituibile³⁷⁹.

La differenza tra livello della struttura e del concatenamento non è solo teorica, ma anche politica. Il concatenamento non si limita a evidenziare la distribuzione vettoriale dei significati dal lato sintattico-semantic, ma coglie gli *effetti incorporati*, ovvero le trasformazioni pragmatiche che si accompagnano sempre al linguaggio: la trasformazione che subisce il corpo dei due fidanzati quando dicono «sì, lo voglio»; la trasformazione che subisce la moneta quando un’autorità monetaria la dichiara fuori corso; la trasformazione che subisce un imputato in colpevole quando il giudice lo condanna; la trasformazione che subiscono le affermazioni di un adolescente quando legge che «è normale sentirsi così» in un *response* di ChatGPT³⁸⁰. Infatti, le parole d’ordine costituiscono le unità minime del linguaggio, linguaggio che interferisce con la materia, andandola a trasformare non fisicamente, ma mediante la macchina astratta. La produttività sociale dei LLM può registrarsi solo a livello di concatenamento, per esempio un *response* può trasformare un utente in cliente soddisfatto, consumatore informato, potenziale elettore, paziente preoccupato. Si capisce allora perché per chiarire la natura

³⁷⁷ Da notare che il termine proviene da una riflessione sull’idea di libero discorso indiretto usata da Pasolini.

³⁷⁸ Deleuze, Gilles, Guattari, Félix, *Mille piani*, Castelvecchi: Roma, 2017, p. 121.

³⁷⁹ Holland, Eugene W., *Deleuze and Guattari’s «A Thousand Plateaus»*, p. 70.

³⁸⁰ Questo presupposto anti-soggettivistico si rivela cruciale per lo studio dei casi etici di responsabilità degli output. Se la *response* è un atto di un concatenamento collettivo d’enunciazione, dichiarare un singolo soggetto responsabile è *de facto* impossibile. Si tende in questi casi a individuare un capro espiatorio, una figura che faccia da contraddizione interna del sistema. Le aziende produttrici tendono facilmente a rilevare, a seguito di un *response* dannoso o un utente che abbia proposto un prompt cattivo, o al LLM che ha allucinato oppure al *fine-tuning* che era inadeguato; in nessun caso il concatenamento viene messo in questione. Forse una critica seria dovrebbe invece risalire dal frammento all’intero assemblaggio. Per la questione del capro espiatorio cfr. Holland, Eugene W., op. cit., pp. 85-86.

di un LLM non basti una risposta che verta sull'architettura *transformer*, ma serva anche chiarirne l'uso. L'uso è indicato dalla nozione di macchina astratta, di diagramma. Quindi nulla si potrebbe senza l'architettura, ma questa risulterebbe cieca se non concretamente situata entro un concatenamento macchinico.

2.1 DESIDERIO VETTORIALE

Il corpo senza organi non è Dio, tutt'altro. Ma divina è l'energia che lo percorre, quando attira tutta la produzione e le serve da superficie incantata miracolante, iscrivendola in tutte le sue disgiunzioni.³⁸¹

Quando il corpo senza organi (CsO) viene comparato a Dio, non lo si designa in quanto tale, ma viene comunque riconosciuta una qualità «divina» alla sua energia. Esso è percorso da una ramificazione di scelte binarie, sotto il nome di sintesi disgiuntive inclusive. La divinità dell'energia delle sintesi è da attribuire al suo «apparente scontro» con le macchine produttive. Mentre queste comportano la creazione di un potenziale sempre nuovo di energia, il CsO oppone una forza qualitativamente altra; la sua differenza di natura rispetto alla macchina si afferma come “anti-produzione” poiché non compete alla creazione di nuova energia, ma alla possibilità di identificare il processo con cui si è sviluppato il flusso di desiderio. Una molteplicità di codici, un groviglio di flussi può venire scandagliato secondo diverse “passeggiate schizo” ovvero seguendo l'itinerario di diverse tensioni libidinali. La forza di anti-produzione si traduce nell'essere «*liscio, scivoloso, opaco, una superficie che resiste alle macchine-organo come fosse una barriera*»³⁸²; quindi non si tratta tanto di una negazione della produzione, ma di esserne una condizione di possibilità, atta a non irrigidire la produzione. Se quest'ultima prende il sopravvento, infatti le macchine «*si annidano al corpo-senza-organi come tanti punti di disgiunzione, tra i quali si tesse ora un'intera rete di nuove sintesi, che scandiscono la superficie in coordinate, come una griglia*»³⁸³.

Nonostante la complessità e l'ineffabilità della nozione, la si può ridurre a tre connotati fondamentali: il CsO è in primo luogo una superficie di registrazione: «*ciò che si manifesta nel corpo-senza-organi sono essenzialmente segni delle connessioni tra*

³⁸¹ Per i tre esempi (matrimonio, moneta, sentenza), cfr. Deleuze, Gilles, Guattari, Félix, op. cit., pp. 118-119.

³⁸² Deleuze, Gilles, Guattari, Félix, *Anti-Edipo*, Einaudi: Torino, 1975, p. 15.

³⁸³ *Ibid.* p. 15.

*organi e macchine che ci consentono o ci obbligano a ripetere precedenti modalità di desiderio-soddisfazione, sebbene con un grado maggiore o minore di libertà di variazione all'interno della ripetizione»*³⁸⁴. Sono perciò segni a venire registrati e non direttamente i flussi. I segni possono intendersi come tracce derivate, prodotti finiti delle connessioni. Holland insiste sul fatto che la registrazione avviene quando la connessione è interrotta, in cui cioè il flusso di desiderio viene a identificarsi con lo stesso prodotto.

Ad esempio, la bocca del bambino viene a staccarsi dal seno materno non appena si raggiunge la sazietà e quello che resta sul CsO è il *segno* della soddisfazione, non la soddisfazione stessa; a registrarsi è ciò che non è più in connessione ed è già passato dal lato del prodotto. Proprio questa differenza ontologica permette poi l'inversione che si accennava nel paragrafo precedente. Successivamente, il CsO appare come radicalmente eterogeneo e polivocale: non è un sistema unico di segni, non è una lingua o una catena di significanti (Lacan). «*Le relazioni-segno sul corpo-senza-organi formano «una molteplicità talmente complessa che difficilmente possiamo parlare di una sola catena o persino di un unico codice del desiderio»*³⁸⁵. Le tracce sul CsO sono eterogenee e irriducibili a un'unità superiore³⁸⁶, una molteplicità di codici, finendo per essere strutturato come una macchina semiotica mista. Infine, il CsO è una *tabula rasa*

Si può paragonare a una sorta di tabula rasa, che libera l'organismo dalla ripetizione puramente meccanica delle determinazioni istintuali e dalle fissazioni nevrotiche — a condizione di comprendere che tale tabula rasa non esiste fin dall'inizio, ma si produce piuttosto nel corso dello sviluppo psichico attraverso la trasformazione delle energie di connessione in energie di registrazione.³⁸⁷

Essa non fa però da presupposto, da condizione iniziale, ma è l'iscrizione, il risultato del processo. Ciò che funziona come superficie di registrazione dei segni non è tanto

³⁸⁴ Holland, Eugene W., *Deleuze and Guattari's Anti-Oedipus*, Routledge: Londra, 1999, p. 31.

³⁸⁵ Holland, Eugene W., *Ibid.*, p. 29. Corsivo nostro.

³⁸⁶ *Ibid.* la citazione interna è da Deleuze, Gilles, Guattari, Félix, op. cit., pp. 46-47.

³⁸⁷ Qui è doverosa una menzione del termine molteplicità. Esso non va inteso qui alla maniera di un vago sostantivo come fosse il Molteplice di contro all'Uno, ma di un aggregato di parti non sussumibili ad un proprio principio; un'aggregazione eteroclita e non identificabile per mezzo di un ordine superiore. In *Mille Piani* Deleuze definisce il rizoma come una molteplicità a *n* dimensioni sulla base degli studi del matematico Bernhard Riemann sulle superfici pluriestese. La molteplicità fornisce due supporti alla filosofia anti-essenzialista deleuziana. Da un lato la variabilità delle dimensioni di una molteplicità e dall'altro l'assenza di una dimensione supplementare, maggiore che garantisca un ordine estrinseco sulla molteplicità, una metrica con cui assicurarla. Cfr. Deleuze, Gilles, Guattari, Félix, *Mille piani. Capitalismo e schizofrenia*, trad. it. di G. Passerone, Castelvocchi, Roma, 2017.

l'archivio dei *log* di conversazione, come si sarebbe portati a pensare, in virtù della trascrizione a parole; il CsO dei LLM è la *matrice dei pesi* del modello, intesa come superficie iscritta sulla quale tutti i flussi linguistici, una volta interrotti dalla *backpropagation* e cristallizzati in parametri, hanno lasciato la loro traccia. Ad esempio, la matrice non parla e non capisce propriamente. Già Weatherby, descrivendo l'assemblaggio di token in termini di simulacro, mostra che i LLM non comprendono cosa trasformano, ma seguono un cieco percorso statistico di pesi che ne fanno una superficie «opaca, liscia», sempre ricomponibile secondo diverse disgiunzioni. La matrice è un corpo, in virtù della sua corporeità, ma è «senza organi» proprio perché non ha alcuna funzione propria al di fuori del sistema in cui è iscritta. Infatti, Deleuze e Guattari chiamano il corpo “senza organi” per interrogarne la struttura e le pretese sistematiche che si attribuiscono ad essa; l'essere senza organi significa avere la possibilità di rigenerarsi continuamente. Infine, la matrice segue le proprietà indicate da Holland: registra i segni di connessioni umane (non le connessioni stesse, ma i segni, ovvero le tracce di testo), è radicalmente eterogenea (convivono frasi mediche, sentenze giuridiche, dialoghi di film, ricette) e si può considerare una *tabula rasa* in quanto viene prodotta nel tempo, essendo l'esito di un addestramento iterato per mesi³⁸⁸.

Holland indica che esistono dei paralogismi delle sintesi (non tanto per motivi epistemologici, portando a produrre una falsa illusione, ma perché rappresentano una deformazione delle tre sintesi originarie che si sono descritte sopra) in virtù dell'uso illegittimo che se ne fa: quando cioè una sintesi viene piegata a una sintassi che non è la sua (ovvero passando dall'essere continua, aperta, eterogenea all'essere esclusiva, chiusa, omogenea). La registrazione dei segni non è continua, ma una serie aperta di alternative inclusive che permettono al desiderio di ramificarsi indefinitivamente. Il paralogismo del *double bind* nasce laddove la ramificazione e la diffusione delle linee di desiderio, delle pratiche, viene a gerarchizzarsi e filtrata da una serie di scelte binarie: con ciò avviene il passaggio alla disgiunzione esclusive o...o...o, riducendo le possibilità reali. il paralogismo si mostra ad esempio nel caso in cui output filtrati rispondano alle esigenze dell'utente secondo formule precostituite. Nel caso un input riproduca una forma non conforme ai valori accordati dal *fine-tuning*, il LLM è addestrato per non rispondere in un certo modo, il che riduce le possibilità reali e la complessità della questione attraverso delle opposizioni binarie. Si pensi a come vengono addestrate a rispondere a questioni politiche divisive: non potrà che offrire una panoramica “o del partito X o del partito Y”

³⁸⁸ Holland, Eugene W., op. cit., p. 31.

e non mostrare la reale complessità “che attecchisce come partito X e Y, come X e Z, come Y e Z”. Così facendo l’effetto cumulativo produce una riduzione progressiva delle pluralità delle disgiunzioni inclusive: il risultato della sintesi sul CsO computazionale viene ridotto a una serie di binarismi.

Il paralogismo della rappresentazione o dello spostamento viene riassunto da Holland invece così: «*La legge ci dice: “Non sposerai tua madre e non ucciderai tuo padre”. E... i sudditi docili si dicono: “Ecco, era proprio quello che volevo!”*»³⁸⁹. La legge allora fornirebbe un’immagine sfalsata, perciò offrirebbe una forma surrettizia alla reale disposizione del desiderio. Le persone sarebbero portate a scoprire cosa desiderano proprio quando ciò viene loro negato. Il paralogismo tiene conto di tre termini: il significante repressore (la legge secondo cui non si può sposare la propria madre), la rappresentazione spostata (il complesso di Edipo, ovvero il divieto di sposare la propria madre perché atto incestuoso) e il rappresentante represso (il desiderio reale che non è già stato edipizzato)³⁹⁰. L’immagine falsata, ovvero la rappresentazione spostata non dà alcuna informazione e verità sul rappresentante represso. Ora, se si ripensa all’esempio della scelta politica si capirà che certi output possono fornire un’immagine, una rappresentazione spostata a cui egli, da soggetto docile, finisce per aderire. «*Quindi questo è quello che volevo!*» è la formula. La macchina non può essere cosciente di tale operazione ideologica, ma il concatenamento sistema-corporate-LLM lo è, nella misura in cui il *fine-tuning* mira a produrre delle risposte in cui l’utente possa rispecchiarsi. Il desiderio viene quindi catturato per poi venire distorto entro una rappresentazione falsificata.

V’è inoltre da menzionare la tendenza ridurre la complessità a uno schema fisso, secondo il paralogismo dell’*applicazione*. Come l’eterogeneità dei fini del desiderio viene ridotta allo schema edipico³⁹¹, così la specifica conversazione viene ricondotta a un alfabeto di *use-cases* commerciali (segretario, scrittore creativo, programmatore, ricercatore...) invariante rispetto all’enorme varietà delle interrogazioni. La richiesta dell’utente viene in realtà tradotta entro la sua versione programmata dall’azienda: l’ottimizzazione del lavoro per il segretario, la generazione di contenuti per lo scrittore

³⁸⁹ In realtà, per quanto concerne l’eterogeneità, essa pertiene al livello contenutistico, di cui si è menzionato qualche esempio, ma non a quello formale. Qui sta la grande differenza tra il corpo senza organi come superficie di registrazione della matrice dei pesi e il corpo senza organi libidinale: questo è *virtualmente* esteso anche a livello molecolare, mentre quello gioca su una distribuzione statistica che poggia sui grandi insiemi “molari” di informazioni. Tutti gli esempi menzionati sopra sono infatti materialmente dei *bit* di informazione.

³⁹⁰ Holland, Eugene, W., op. cit., pp. 41-43.

³⁹¹ Sui tre termini (significante repressore, rappresentazione spostata, rappresentante represso) cfr. Deleuze, Gilles, Guattari, Félix, *Anti-Edipo*, Einaudi: Torino, 1975, pp. 136-137.

creativo o il ricercatore...

Bisogna menzionare anche il paralogismo del *doppio impasse*. Esso consiste nel presentare al desiderio solamente due possibilità: la risoluzione (assumere la legge e differenziarsi all'interno di essa) e la fissazione (rifiutarla e restare bloccati nel suo orizzonte). Nel caso preso in esame, l'utente o accetta le risposte come "utili", diventando sempre più dipendente dalle conversazioni digitali al fine di integrarle attivamente nel suo processo decisionale oppure la rifiuta *in toto*, ma in tal caso resta bloccato in un atteggiamento difensivo, nello stesso orizzonte di senso che l'apparato gli ha imposto. È da sottolineare che in entrambi i casi è l'apparato a vincere³⁹², perché entrambi i poli del dilemma sono sue produzioni. Insomma, si finisce o per essere catturati dalla rappresentazione spostata del modello oppure si è catturati nel doppio specchio, nell'anti-modello del LLM³⁹³. L'apparato di cattura co-emerge con le sintesi che si attuano sulla superficie di registrazione. Così Holland riassume l'ambivalenza del CsO:

Qualsiasi rappresentazione fissa — che sia limitante o meno — distorce la vera natura multiforme del desiderio. L'ambivalenza costitutiva del corpo senza organi è tale da esporre il desiderio alla distorsione e alla cattura da parte della rappresentazione proprio nello stesso momento e nello stesso processo — la trascrizione del desiderio in segni — che lo apre a infinite variazioni.³⁹⁴

Ad ogni possibilità di liberazione secondo le sintesi disgiuntive inclusive può corrispondere una possibilità di cattura (il segno, viene preso per essere la verità del flusso e ne fa un'origine). I LLM possono infatti offrire un ventaglio di interpretazioni che apre a traiettorie inattese, facendo della sintesi un uso legittimo, oppure possono offrire un'unica interpretazione presentata come la più plausibile su cui l'utente fissa il suo desiderio (uso illegittimo).

Il CsO gioca un ruolo ambiguo nel sistema di formazione del desiderio. Poiché la cattura è co-emergente rispetto alla produzione del desiderio. L'output diventa seduttivo nella misura in cui maschera una dinamica molto più complessa e aperta a opzioni irriducibili a una singola immagine. In altre parole, la risposta sostituisce al desiderio dell'utente una sua versione operativa, in cui il problema è già stato ridotto a

³⁹² «ogni ricchezza delle determinazioni sociali reali si riduce a un'espressione monotona dell'Edipo» *Ibid.* p. 119.

³⁹³ «Sia in salute che in malattia è sempre Edipo che vince» cfr. Holland, Eugene W., op. cit., p. 43.

³⁹⁴ La macchina, producendo il proprio bisogno, genera la propria critica in uno spazio interno al proprio gioco.

una scelta tra opzioni preformate. L'utente, leggendo e convincendosi di appartenere alla casistica meticolosamente aggiornata e spiegata dalla macchina viene trasferito sul piano di una nuova immagine di desiderio: il paralogismo dello spostamento farà quindi riconoscere l'utente nel modello proposto. Queste sono le trasformazioni incorporee di cui Deleuze e Guattari avvertono in *Mille piani* e che non possono venire afferrate da un discorso limitatamente circoscritto a vettori, pesi e parole.

La persuasione si attesta anche a fronte di certi dati: modelli come GPT4 sono ritenuti essere 64,4% più convincenti degli avversari umani in un dibattito, articolato sulla base di precisi dati demografici³⁹⁵. Già nel 2023 una serie di studi sperimentali ha mostrato che la sostituzione di descrizioni del prodotto scritte da copywriter umani con versioni generate da GPT3 produce effetti misurabili sulle intenzioni d'acquisto, con incrementi d'acquisto del 15-25%³⁹⁶. Non è tanto la qualità delle descrizioni ad essere superiore, ma la loro aderenza statistica al profilo del consumatore, che permette di simulare uno slogan o un testo che, nella ponderata scelta di alcuni termini fornisca già un richiamo a certi temi a lui cari; *la persuasività non passa tramite la precisione di un contenuto veicolato, ma attraverso il lavoro sintattico, ovvero la retorica del marketing contemporaneo*. Il loro esito è il precipitato statistico di milioni di copywriter umani; ogni nuova descrizione consegue come variante della medesima struttura retorica del *marketing*.

A livello politico, la persuasione appare come strumento critico di direzionalità del consenso e delle decisioni dell'elettore³⁹⁷, senza contare che i LLMs sono largamente usati per generare contenuti per la campagna elettorale, operazioni di *micro-targeting*, personalizzazione di messaggi. L'effetto più marcato si registra su utenti che risultano non avere un'opinione preformata, ma la persuasività comunque attecchisce anche su coloro che ritengono avere una posizione solida. Il *response*, infatti, arriva sotto la veste retorica della neutralità informativa: *si dice* che queste sia la somma di informazioni attualmente disponibili e che la risposta offra così una carrellata "oggettiva" delle posizioni politiche, degli scenari attuali. Insomma, la manipolazione non è un fenomeno ausiliario, una sorta di epifenomeno removibile, ma è proprio grazie al paralogismo dello spostamento, applicato concretamente nella formulazione di un *response* che viene a

³⁹⁵ *Ibid.* pp. 37-38.

³⁹⁶ Salvi, Francesco *et al.*, *On the Conversational Persuasiveness of Large Language Models. A Randomized Controlled Trial*, in «Nature Human Behaviour», 8, 2024, pp. 1-12, in particolare pp. 5-8 per i risultati sull'aumento della persuasione con personalizzazione.

³⁹⁷ Sui test sul *click-through* di descrizioni di prodotto generate da LLM rispetto ai controlli umani cfr. Brand, James *et al.*, *Using GPT for Market Research*, Harvard Business School Working Paper 23-062, 2023, p. 12.

prodursi una specifica immagine, rappresentazione surrettizia del desiderio (la *query* dell'utente) per il consumatore e per l'elettore, che vengono a scoprire ciò che il sistema gli ha proposto tramite una muta rete di parametri e pesi. Sul flusso viene a situarsi un segno-prodotto compiuto che si presenta come maschera dell'argomentazione, non come suo stadio. È in gioco la materialità del desiderio, l'esibizione della sua processualità anziché la feticizzazione di una sua rappresentazione. La persuasione passa non attraverso un vicendevole scambio dialettico, ma il riconoscimento della superiorità del prodotto-segno della macchina³⁹⁸, che politicamente può diventare una giustificazione ideologica dello stato attuale delle cose: il segno, l'effetto viene scambiato per causa. Il paralogismo smuove retroattivamente, creando una nuova versione interpretativa. È questo aspetto che nell'*Anti-Edipo*³⁹⁹ viene fissato come divinità del CsO, riuscendo ad operare da quasi-causa dei flussi di desiderio.

Per riassumere, in questo capitolo si è voluto vedere come l'assemblaggio utente-LLM-*corporate* funziona come una macchina desiderante che opera su un corpo senza organi computazionale (le matrici QKV), la cui sintesi disgiuntiva di registrazione cattura i flussi di desiderio dell'utente, trasformandoli in segni-prodotti finiti. Il passaggio dalla produzione del desiderio, al suo statuto illusorio di rappresentazione segnica non fa che produrre un soggetto-consumatore o elettore che riconosce come sue le pratiche computazionali della macchina. Si sono visti operare tre paralogismi (del *double bind*, dell'applicazione, dello spostamento) in successione, producendo dipendenza dell'utente dal segno macchinico. La sezione successiva allora vuole indagare la natura del segno macchinico: in che senso si produce una sostanzializzazione del linguaggio e in cosa ciò sia incompatibile con la pretesa naturalità formale rivendicata dai teorici del neostrutturalismo.

³⁹⁸ Bai, Hui *et al.*, *Artificial Intelligence Can Persuade Humans on Political Issues*, OSF Preprints, 2023, in particolare pp. 7-9.

³⁹⁹ Studi sperimentali di psicologia rilevano che la fiducia esagerata nell'output dei LLMs produce un *de-skilling* progressivo: gli utenti infatti perdono la capacità di compiere autonomamente operazioni cognitive che la macchina svolge al posto loro i programmatori che usano *Copilot* per generare codici, delegando gran parte del lavoro alla macchina, perdono la loro capacità critica. Il programmatore verrà a definirsi *ex post* come "utente di Copilot", avverando un'ulteriore sintesi congiuntiva paralogizzata.

CAP. 3: LE SEMIOTICHE DEGLI LLMS TRA HJELMSLEV, PEIRCE E GUATTARI

Il presente capitolo mira a mostrare la dinamica semiotica che sottintende ai LLMS, mostrando che l'operazione che Weatherby avverte come *vettorializzazione*, fondandosi sull'ipotesi differenziale di Saussure, non rende correttamente la *sostantivizzazione* del linguaggio a cui invece ci si può richiamare per mezzo di Hjelmslev. In gioco non c'è solo una differenza teorica, ma, fintanto che i LLMS si sono descritti in concatenamento con le *corporate* e gli utenti, essi non possono essere descritti come operanti tramite un puro ordine simbolico. Come è possibile che tre realtà eterogenee possano raccordarsi, incrociandosi in un'unica regione. Weatherby la chiama spazio vettoriale, qui vorremmo dimostrare che esso si possa descrivere solo attraverso un'intima sovrapposizione tra il

segno-prodotto e la “materialità” linguistica non ancora codificata, ma strutturabile secondo codici. Il segno non è per natura al confine tra parole e suoni come vorrebbe Saussure, delegittimando così ogni provenienza umana di esso e facendone un limite che continuamente insiste sulle cose e da esse ne è distaccato. Il linguaggio funziona al limite, *insistendo* come mostrano gli autori di *Mille piani*, ma con questo suo carattere evenemenziale esso appare definitivamente inscindibile dal concatenamento di cose a cui si intreccia. L’organizzazione dello spazio vettoriale corrisponde allora a una *semiotica significativa* nel senso di Guattari e non a una semiotica *a-significante*. Bisognerà dapprima abbozzare un preambolo linguistico imperniato sulle teorie di Louis Hjelmslev e la sua glossematica, poi richiamarsi a Peirce e a Guattari, per poi misurarsi con la nozione di vettorializzazione di Weatherby.

3.0 UN FILONE LINGUISTICO MATERIALISTICO

Partorita tra gli anni Trenta e Quaranta del Novecento, la dottrina glossematica di Hjelmslev viene utilizzata in *Mille piani* da Deleuze e Guattari, venendo riconosciuta come la *sola* linguistica davvero adatta alla descrizione dei flussi desideranti del tardo capitalismo⁴⁰⁰. In realtà, anche nell’*Anti-Edipo* viene ampiamente valutata in positivo: «La linguistica di Hjelmslev si oppone in profondità all’impresa saussuriana e post-saussuriana»⁴⁰¹. L’opposizione si gioca su un punto a prima vista irrilevante: per Saussure la lingua si situa sul territorio di confine tra il pensiero e il suono, poi tradotti in significato e significante, combinati tra loro per produrre forma e non sostanza⁴⁰². Hjelmslev invece ritiene che a monte il segno non sia slegato e fluttuante, ma intimamente prodotto (sostantivizzato) dall’interazione tra la matrice di forme e l’amorfa materia. Allora, fine e cruciale risulta la distinzione tra *materia (purport)*, *forma* e *sostanza*.

La materia è un’amorfa «massa di senso», comune a tutte le lingue, che si lascia plasmare come sabbia: «è la stessa sabbia che diversi stampi (*le lingue*) modellano in modi diversi»⁴⁰³. La materia non può essere slegata però dalle molteplici occorrenze con

⁴⁰⁰ Deleuze, Gilles, Guattari, Félix, op. cit., p. 20.

⁴⁰¹ Maurizio Lazzarato tiene allo stesso modo in grande considerazione l’opera semiotica di Guattari, che grazie alla dimensione intensiva, polisemica e preconettuale riesce a rendere conto di un flusso d’informazione senza passare per la sua modulazione in *bit*. L’asservimento macchinico è il raccordo dei flussi, a livello impersonale e pre-individuale che sgretola l’integrità monolitica dell’individuo. Cfr. Lazzarato, Maurizio, *Segni e macchine*, Ombre corte: Verona, 2019.

⁴⁰² Deleuze, Gilles, Guattari, Félix, op. cit., p. 242.

⁴⁰³ «Saussure situa il linguaggio nel “territorio di confine” tra pensiero e suono, la combinazione dei quali produce forma, non sostanza» cfr. Genosko, Gary, op. cit., p. 162.

cui viene ad articolarsi: *essa non esiste se non come sostanza per una qualche forma*⁴⁰⁴. La forma «*si proietta sulla materia come una rete aperta proietta la sua ombra su una superficie indivisa*»⁴⁰⁵. Allora la sostanza non è che *materia formata*. Ora, la tripartizione deve venire applicata sui due grandi versanti che interagiscono: il piano del contenuto e dell'espressione. Sul piano del contenuto il "pensiero" (la massa indivisa delle possibili associazioni concettuali) viene formato da una forma-contenuto per dare luogo a una sostanza-contenuto; sul piano dell'espressione, la massa dei suoni possibili (la *purport* dell'espressione) viene formata da una forma-espressione per generare una sostanza-espressione. Gary Genosko descrive la processualità insista nella triade materia-forma-sostanza tramite la seguente formula:

$$EP \rightarrow ES \rightarrow EF \Leftrightarrow CF \leftarrow CS \leftarrow CP^{406}$$

L'argomento a sinistra ritrae la triade dell'espressione e quello a destra del contenuto secondo materia (*purport*), sostanza e forma. L'unidirezionalità delle frecce serve ad indicare la dimensione processuale, il passaggio da materia grezza a sostanza; la freccia bidirezionale al centro mostra che il processo tende a dematerializzarsi in pura forma. La sostanza è perciò non altro dalla forma, ma è già formata, sotto una determinata casistica. La sostanza è allora il luogo in cui la forma si fa percepibile tramite la macchina astratta del linguaggio che si incarna in materiali specifici (la lingua italiana, inglese, francese si può sviluppare in forme-concetti oppure in forme-suoni).

Si capisce allora perché Deleuze e Guattari siano interessati al linguista danese: la glossematica dimostra che ogni sistema di segni (come una lingua) si può, magari per scopi analitici, descrivere come una pura forma, ma facendo così viene amputato l'aspetto di precipitazione e di processualità che corrisponde alla sua medesima produzione. Così quando Hjelmslev dice che «*La sostanza non è una presupposizione necessaria della forma linguistica, ma la forma linguistica è una presupposizione necessaria della sostanza*»⁴⁰⁷ non afferma l'inesistenza della sostanza. Insomma, ogni segno viene ad essere l'esito di una produzione, esattamente come i due autori dimostrano dal punto di vista del desiderio nell'*Anti-Edipo*. La prospettiva saussuriana è povera e non offre grande

⁴⁰⁴ Hjelmslev, Louis, *Prolegomena to a Theory of Language*, University of Wisconsin Press: Madison, 1969, p. 52.

⁴⁰⁵ «Non ha esistenza possibile se non in quanto è sostanza per una forma o per un'altra» cfr. Hjelmslev, Louis, op. cit., p. 52.

⁴⁰⁶ Genosko, Gary, op. cit., p. 162.

⁴⁰⁷ *Ibid.* p. 161. Dove si ha, a cominciare da sinistra, *purport* dell'espressione, sostanza dell'espressione, forma dell'espressione, forma del contenuto, sostanza del contenuto e *purport* del contenuto.

apporto critico al dibattito allora, elidendo la formazione della sostanza, il suo lavoro di organizzazione sul piano materiale.

Per riprendere l'apporto di Peirce al discorso⁴⁰⁸, si sottolinei l'importanza della *trasmissibilità* di un segno rispetto ad un altro. La triadicità del processo di semiosi fa di un interpretante un segno per l'interpretante successivo e così via: «*per Peirce, un segno non trasmissibile non è un segno. La trasmissione [...] rileva un altro segno e questo è chiamato interpretante*»⁴⁰⁹. Se si considera gli LLMs come processi semiotici, allora ogni generazione di token altro non è che una catena in cui ciascuno diventa interpretante (segno) per quello successivo. Ci sono tre passaggi con cui si può costituire un segno: se si vede del fumo, allora si potrà intuire che c'è un fuoco; il fumo si dirà *indice* del fumo. Un indice è un segno che per contiguità, contatto fisico oppure causalità rimanda a un altro. Un cartello con una sigaretta barrata contrassegna invece un'area in cui non si può fumare: il cartello è un' *icona* rispetto all'immagine che mostra, poiché si fa segno tramite somiglianza. Oppure, un segno si può essere un *simbolo*, venendosi a costituire tramite convenzione, come nel caso una persona saluti alzando la mano con le dita dritte e separate; perciò, la si dovrà salutare “battendole il cinque”. Ogni parola è un simbolo allora, perché non ha né effetto né somiglianza con ciò che esprime. Si potrebbe allora inferire che i LLM non hanno né icone né indici, ma solo simboli, visto che manovrano le occorrenze statistiche di un termine.

Un ultimo elemento appare necessario allo svolgimento dell'analisi del processo semiotico degli LLMs. Si tratta della maniera con cui Guattari si serve di Hjelmslev e Peirce per comporre le sue *semiotiche miste*. Per facilitare la comprensione, si è aggiunto il seguente schema (figura 11).

I bordi della tabella sono segnati verticalmente dal piano dell'espressione e del contenuto, mentre sono tripartiti orizzontalmente dalla materia, sostanza e forma. La semiotica, per come si è soliti contraddistinguere, a partire dalla sua attinenza alla precettistica “scientifica” della linguistica di Saussure e dei suoi epigoni, è indicata sotto il nome di *semiologia di significazione* [signifying semiology]. Essa infatti prende una sostanza, come il linguaggio, per sovracodificare tutte le altre. Il segno è arbitrario rispetto alle molteplici occorrenze di significanti poiché è sganciato dalla materia.

⁴⁰⁸ *Ibid.*

⁴⁰⁹ Figure 5 e 6.

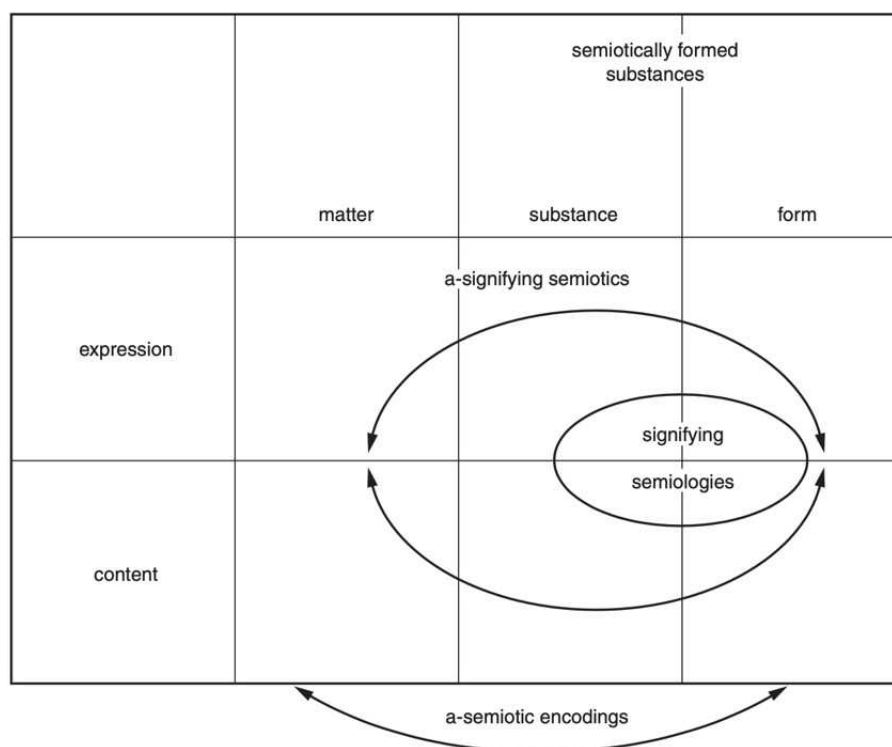


Figura 13 semiotiche miste⁴¹⁰

La semiologia di significazione attecchisce solo alla forma con cui la sostanza viene veicolata, rientrando nella categoria delle semiologie significanti, ovvero nei sistemi di segni con sostanze formate sui piani di espressione e di contenuto⁴¹¹. La grande novità di Guattari, che si comincia ad avvertire dai suoi corsi negli anni '70 e che vedrà la sua formulazione più rigorosa ne *L'Inconscio macchinico*⁴¹², è l'estensione oltre il paradigma della significazione.

Un processo semiotico si può attestare infatti anche al di fuori delle sostanze della forma e del contenuto: si definiscono *semiotiche a-significanti* [a-signifying semiotics] quei processi che mantengono una connessione circolare tra forma e materia, aggirando il regime dispotico di sovracodifica delle semiologie significanti. Al contempo, esse non strabordano rispetto i due piani dell'espressione e del contenuto. Attraverso questo nuovo regime si può fuggire dalla forma segnica del simbolo significante. La rappresentazione è vista come prodotto non ridondante, ma sempre originale (il rimando tra forma e materia è infinito, e quello alla sostanza è reciso). Le macchine astratte e i flussi materiali sono

⁴¹⁰ Weatherby, Leif, *Language machines*, Minnesota University Press: Minneapolis, 2025, p. 130.

⁴¹¹ Immagine da Genosko, Gary, *Félix Guattari. An Aberrant Introduction*, Continuum: London-New York, 2002, p. 167.

⁴¹² Vi sarebbero infatti anche le *semiologie simboliche* in cui più sostanze possono vedersi parallelamente senza che nessuna sovracodifichi l'altra, come avviene invece nel caso del linguaggio e della sua onnipresenza negli altri regimi di segni.

contraddistinti da questa forma di semiosi. Infine, al di fuori dei piani di espressione e contenuto si pongono le *codifiche a-semiotiche* [a-semiotic encodings], chiamate così perché non danno vita a nessuna sostanza semiotica: «*includono le codifiche genetiche o qualsiasi tipo di codifica cosiddetta naturale che funziona indipendentemente dalla costituzione di una sostanza semiotica*»⁴¹³. L'esempio che viene fatto da Guattari è quello del DNA che può essere visto come causa codificante dell'organismo, ma non c'è nulla al di fuori dell'organizzazione formale della materia biochimica, quindi delle proteine chiamate a riprodurlo. Un altro esempio, più attinente al caso proposto in esame, potrebbero essere le interazioni elettriche sul silicio, che possono venire descritte da leggi fisiche. I tre modelli sono sempre coordinati e offrono quindi un inquadramento maggiore, più articolato per definire il processo di semiosi che contraddistingue le macchine: questo è ciò che si intende quando si parla di *semiotica mista*, come combinazione delle tre dimensioni di semiotiche significanti, semiotiche a-significant e codifiche a-semiotiche, «*accordando il senso e il non-senso del desiderio per come essi sono, senza tentare di adattare i modi di soggettivazione alla significazione e alle leggi sociali dominanti*»⁴¹⁴.

Ora, quando Weatherby nel primo capitolo di *Language Machines* ricostruisce con precisione la genealogia del concetto di vettore per i modelli NLP passa attraverso le matrici *document-term*⁴¹⁵, le strategie di *information retrieval* descritte da Bernhard Rieder⁴¹⁶ e giunge al *word embedding* dei moderni *transformer*. È la vettorializzazione che fa da svolta alla traduzione del linguaggio in campo informatico, nessuna operazione simbolica o probabilistica: «*la matematica dei vettori ha offerto una chiave per la modellizzazione che tende a costruire un senso del significato distribuendo quelli che sembrano essere gli “aspetti” del significato attraverso uno spazio vettoriale continuo*»⁴¹⁷, superando le correlazioni piatte dei modelli a rete in favore di un modello a più dimensioni non solo distributivo, ma anche predittivo. Il risultato però è ritenuto una radicalizzazione strutturalista: l'introduzione del vettore consentirebbe di realizzare un modo finalmente adeguato l'ipotesi differenziale che sta a monte della scoperta. Quando Weatherby cita Hjelmslev⁴¹⁸ oppure Peirce⁴¹⁹ lo fa sempre con l'intenzione di riscrivere il

⁴¹³ Guattari, Félix, *Inconscio macchinico*, Orthotes: Napoli, 2022.

⁴¹⁴ Genosko, Gary, op. cit., p. 166.

⁴¹⁵ *Ibid.* p. 170.

⁴¹⁶ Weatherby, Leif, op. cit., pp. 60-61.

⁴¹⁷ Rieder, Bernhard, *Engines of Order. A Mechanology of Algorithmic Techniques*, Amsterdam University Press, Amsterdam 2020, pp. 232-274.

⁴¹⁸ Weatherby, Leif, op. cit., p. 59.

⁴¹⁹ *Ibid.*, pp. 78-79, dove Hjelmslev viene letto come precursore dell'idea di sostituibilità sintagmatico-paradigmatica realizzata dalla vettorializzazione.

loro lavoro entro le coordinate strutturaliste, giustificando la resa attraverso i vettori delle parole come la realizzazione tecnica più potente dell'idea saussuriana.

Quello che Weatherby definisce come vettorizzazione è solo un aspetto parziale che pretende di fondarsi sull'ipotesi differenziale. In realtà, il processo che permette di incorporare parole a vettori prende il nome di sostanzializzazione, secondo tre ragioni:

- 1) La sostanza vettoriale è una *sostanza* nel senso indicato da Hjelmslev. Quando una parola viene incorporata in uno spazio vettoriale viene codificata in vettore. Però il vettore non è una pura forma di posizionamento, ma appare piuttosto come materia formata. I vari stati elettrici di una memoria GPU vengono infatti plasmati secondo una precisa architettura di rete e dei pesi. Si capisce perciò perché un vettore non è un posizionamento neutro e puramente differenziale nello spazio, ma è *già* sostanza dell'espressione, ovvero una determinazione numerica precisa secondo un'architettura *transformer* precisa e un addestramento preciso. Invece che la pura forma differenziale di Weatherby esistono milioni di sostanze vettoriali concrete, codificate secondo la forma del contenuto (la semantica del vettore, ciò che può significare) e la forma dell'espressione (l'architettura della rete).
- 2) La sostanza vettoriale è prodotta da una *sovracodifica*. Essendo il risultato di un addestramento, le sue operazioni si compiono secondo un'ottimizzazione della *loss function*⁴²⁰, che minimizzi l'errore, secondo dei vincoli di *fine-tuning*. La sostanza dell'espressione si basa sul *training* secondo valori esterni ("metriche" di *helpfulness*, *harmlessness* e *honesty*), ma che vengono primariamente tenuti di conto. È il medesimo processo che Guattari attribuisce alle semiologie di significazione: «*sostanze o strati di semiologie di significazione sono centrate su una singola sostanza che domina e sovracodifica tutte le altre*»⁴²¹. Lo spazio vettoriale non è una semiotica a-significante, perché non è in connessione circolare e diretta con la materia (i flussi dei desideri umani, le relazioni sociali oppure del lavoro), ma è mediato dalla sovracodifica epurata dall'addestramento; come delle semiologie della significazione, allora si può dire che il significato vi echeggia all'infinito come fosse sotto il dominio di un significante dispotico.

⁴²⁰ *Ibid.*, pp. 130-131, sulla triade peirciana riadattata alla *next-token prediction*.

⁴²¹ La *loss function* è un'operazione atta a misurare quanto la scelta dei *token* condotta dal modello si discosti dall'originaria previsione; in altri termini, fornisce un valore che riflette quanto la macchina abbia risposto correttamente, ovvero in linea con i risultati sperati.

3) La sostanza vettoriale è eterogenea, ma subisce un processo di *omogeneizzazione*. L'eterogeneità è una caratteristica della materia, poiché appartiene a regimi semiotici differenti. Si prenda ad esempio un codice giuridico interpretabile secondo le norme del diritto internazionale, oppure un testo di medicina secondo i dettami dell'anatomia, un giornale secondo le tendenze politiche che vi sono inscritte. Insomma, le forme dei contenuti e delle espressioni si costruiscono su logiche diverse. L'omogeneizzazione è invece un processo di riduzione a una logica calcolante: tutto viene appiattito sulle leggi di composizione vettoriale (prodotti scalari, somme vettoriali). Solo riducendo l'eterogeneità reale dei regimi semiotici si può giungere allora a instaurare una logica che funzioni secondo *pattern*, come nel caso del celebre esempio $vector(\text{«King»}) - vector(\text{«Man»}) + vector(\text{«Woman»}) \approx vector(\text{«Queen»})$. Le regolarità vengono a prodursi secondo il paralogismo dell'applicazione: si tratta di ricondurre ogni forma di variazione a una sua resa sistematica, ovvero ogni occorrenza è vista come variazione di una struttura fondamentale: il modello di vettorialità geometrica sussume i diversi regimi semiotici concreti.

Si vede perciò perché la vettorializzazione del pensiero, secondo quanto intende Weatherby, è secondaria successiva a una sua *sostanzializzazione*, facendo dello spazio vettoriale non un presupposto astratto e neutro, ma una precisa sostanza secondo una peculiare semantica, indipendente dai flussi della materia di desiderio, omogenea rispetto ai regimi semiotici e riducibili a dei principi fondamentali.

Forse proprio sotto l'ultimo dei tre aspetti, gli LLMs tradiscono la complessità, la differenziazione che il progetto di Guattari tenta di delineare. Le semiotiche miste, infatti, preservano una “non-traducibilità” di alcuni elementi, accogliendo il senso e il non-senso del desiderio per come essi sono, senza ridurli a un raffronto con le coordinate principali. La vettorializzazione invece postula un'integrale traducibilità di termini: ogni segno, di qualsiasi regime semiotico può venire trascritto come vettore.

Si parla di sostanzializzazione non per mostrare come ogni sostanza di forma si appoggi a una di contenuto, ma per controbattere alle intenzioni che studiosi come Weatherby attribuiscono a quest'ultima: la pretesa universale, ma mai esaustiva di riprodurre la *langue*, il linguaggio universalmente preso. Ogni registro semiotico trova infatti la sua attuazione in una sostanza di espressione e di contenuto: perciò, parlare di una sostanzializzazione degli LLMs non arrecherebbe nulla di nuovo al discorso. Certo, ma la vettorializzazione mostra che gli LLMs presentano la propria operazione come

pura forma differenziale dei vettori, mentre i loro effetti si producono sul piano dell'espressione e del contenuto, facendo così della sostanzializzazione un fatto "tematico": la sostanza presa di mira per indicare la capacità linguistica di queste macchine è il linguaggio in generale, la *langue* appunto, reclamandola attraverso la pura forma simbolica. Se si prende un dizionario o un'enciclopedia, nessuno li definirebbe come una pura forma differenziale, tutti saprebbero che sono il prodotto storico di un'indagine promossa da una compagnia editoriale, una scuola linguistica o lessicografica, di una politica della lingua, a differenza degli LLMs che vengono presi per incarnazione neutra e oggettiva della *langue* universale.

Anche l'esempio del calcolo semantico-vettoriale ($king - man + woman \approx queen$) esibisce una pretesa universalità grammaticale tra lingue. Gli *embeddings* catturano relazioni concettuali profonde a prescindere dalle culture locali. L'esempio però appare fruttuoso solo per culture e lingue, come l'inglese o l'italiano, che si formano su una struttura di *gender pairing*. Lo spazio vettoriale allora è sostanzializzato su un preciso spaccato di lingue, incarnato in una sostanza specifica, deponendo la sua definizione di forma pura. La pluralità di geometrie che si presentano a seconda della lingue negli LLMs dimostra che lo spazio vettoriale opera secondo sostanze e non forme universali.

Hjelmslev distingue tra "sistema" e "processo", sostenendo che lo studio del sistema è prioritario rispetto a quello del processo: la *langue* infatti come sistema offre un metodo imprescindibile per lo studio del testo, ossia del processo. Infatti, Genosko ricorda che «*nel sistema di Hjelmslev il purport, ovvero la materia, rimane completamente dipendente dalla forma*»⁴²² e ciò non vale per la teoria del concatenamento. Guattari ritiene che il senso non sia prodotto esclusivo della forma, ma dell'interazione tra macchine astratte e flussi di materia e può secondariamente venire formalizzato secondo una certa maniera. Perciò qualsiasi formalizzazione non esaurisce il senso e rimane uno scarto tra ciò che la macchina produce e ciò che i parlanti invece creano nella loro vita: questo è lo scarto e il luogo della materia a-significante che non si lascia tradurre in sostanza, fondo residuo virtuale. Si è usato il termine "sostanzializzazione" proprio per accentuare la tendenza ad attribuire alla sostanza la causa qualitativa di altri fenomeni di natura sociale, facendone un feticcio. Il feticismo della merce in Marx, ad esempio, dimostra come qualità e aspetti della merce vengano ritenuti inerenti alla merce stessa, già predisposti in essa, recidendo ogni riferimento al

⁴²² Genosko, Gary, *Félix Guattari*, p. 168.

lavoro umano, alla successione concreta di passaggi che hanno segnato l'esecuzione del prodotto; allo stesso modo la macchina viene investita delle capacità che invece appartengono ai parlanti che hanno prodotto il *corpus* di testi. Qui allora si può azzardare l'espressione "feticismo del vettore" per sottolineare l'ideologia che fa apparire il modello autonomo rispetto all'insieme di pratiche umane da cui ha tratto le sue proprietà.

3.1 SOVRACODIFICA, FORMA DEL CONTENUTO E DELL'ESPRESSIONE

Arrivati a questo punto, si può dare una definizione degli LLMs come di *macchine semiotico-linguistiche significanti*, nel senso preciso che ne dà Guattari. La loro funzione precipua è quella di sovracodificare l'eterogeneità delle semiosi umane: gesti, immagini, intuizioni, corpi, emozioni vengono situate entro la griglia di una sostanza semiotica unica, ovvero lo spazio vettoriale, traducendo ogni prompt in una posizione di tale spazio per restituire una *response* prodotto dalla sua geometria. Al contrario, come si è cercato di mostrare, parlare di vettorializzazione come di un'operazione neutra poiché fondata sull'ipotesi differenziale del linguaggio, seguendo quindi le argomentazione di Weatherby, risulta tanto utile per apprendere il funzionamento ingegneristico del modello quanto fuorviante, in ultima istanza operando un mascheramento ideologico circa il concatenamento con cui queste macchine non si possono che intendere.

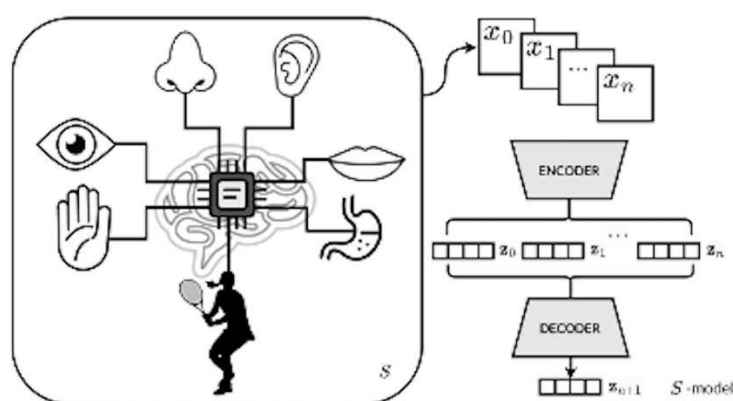


Figura 14 Ciclo dell'informazione tra semiotiche miste e AI⁴²³

⁴²³ Genosko, Gary, op. cit., p. 165.

La vettorializzazione è vista infatti come ultima e performante evoluzione di un percorso atto alla ricerca di contenuti (*information retrieval*), per come è descritta spesso mascherando la sua natura contingente, parziale, politica. L'ideologia della pura forma è, in altre parole, l'ideologia con cui si traveste la corporate proprietaria del modello. Slogan democratico-liberali come: «Il modello non sa nulla che non sia in voi!» sono falsi nella misura in cui il modello *detiene già una forma di conoscenza* espressa secondo *policy* aziendali e inscritta nei suoi pesi, precipitato di una materia formata sotto l'aspetto dell'espressione e del contenuto.

Dopo aver specificato cosa sia la forma del contenuto e dell'espressione per Hjelmslev, è bene cercare di declinare questi concetti nel contesto degli LLMs. La forma del contenuto non è il dataset semplicemente (esso si potrebbe intendere piuttosto come materia formata in una sostanza), ma la *geometria* delle relazioni, appresa dal modello attraverso il *training* e il costante dialogo con gli utenti. Il termine forma infatti indaga proprio la modalità con cui la conoscenza si distribuisce da un punto di vista posizionale. Ad esempio il termine “aborto” si può trovare vicino a quello di “bioetica” o “scelta”, ma molto lontano da quello di “omicidio”, laddove un modello linguistico che dispone di un altro *fine-tuning*, riflettendo un altro contesto culturale, può assimilare i termini in maniera completamente diversa, dando vita così a una politica della *prossimità* tra modelli diversi, a seconda della disposizione lessicale. Sul piano dell'espressione, la forma è invece la *grammatica statistica* con cui i token si articolano in frasi, paragrafi e testi; si ritrova qui non solo l'interfaccia grafica delle *response*, ma anche i *template* o i *boilerplates*⁴²⁴. Anche nelle disposizioni e nelle risposte su determinati temi si riflette l'orientamento politico delle aziende; rispondere a un certo tema con formule simili a «Lei ha sollevato un punto importante» piuttosto che andare al sodo è il risultato di una particolare scelta di *fine-tuning* e non di una struttura formale del linguaggio.

L'aspetto formale conduce perciò una sovracodifica che apporta degli effetti reali: la persuasività di un LLM per come si è trattata nei capitoli precedenti, indirizzando il consumatore o l'elettore a determinate scelte ne fa da esempio; essa è mediata dalla forma del contenuto e dell'espressione. Allora Weatherby coglie un aspetto rilevante per la consistenza linguistica dei modelli, ritenendo che siano strumenti di una *nuova retorica*

⁴²⁴ Immagine da <https://arxiv.org/abs/2304.01481v1> p. 16.

Un LLM è un mondo retorico capovolto in cui le regole della composizione sono [...] un riflesso standardizzato di quella Babele che la "cultura" era dopo la retorica, ovvero la vasta estensione della produzione linguistica umana in concreto. Quando il linguaggio viene modellato, questa cultura retorica ne costituisce il volto. La macchina linguistica ci presenta una retorica nuova, una retorica computazionale.⁴²⁵

Ma sottovaluta il fatto che la retorica sia costruita dentro la macchina, facendo della “Babele culturale” l’oggetto di una politica corporativa.

Un aspetto che ora va esaminato è il posizionamento delle semiotiche a-significanti rispetto alla sovracodifica formalmente operata dalle macchine. Seguendo il ragionamento di Guattari gli LLMs dovrebbero esibire una componente significante e a-significante, assieme a una a-semiotica. *Dentro* il modello si possono trovare alcuni elementi a-significanti: le emergenze inattese, i *jailbreaks* possibili, degli scarti di significato tra prompt e risposta, spesso definiti come *hallucinations* oppure scoperte computazionali che gli ingegneri non sanno spiegare. Vi sono delle zone grige in cui il regime semiotico è messo in crisi, in connessione con i flussi materiali (i pesi numerici), ma la zona si ritrova compresa, *catturata*, ai due estremi degli input e degli output da una semiologia di significazione. L’utente parla una lingua naturale e riceve una risposta in una lingua naturale. La zona a-significante è perciò confinata nel mezzo, nei *layers* nascosti, inglobati però dentro il regime significante. La semiotica si conferma mista, ma la componente a-significante viene relegata in posizione subalterna e controllata da quella significante.

Lo stesso Guattari prevedeva questa dinamica, affermando che «*le formazioni di potere significanti, per mantenere le proprie posizioni, sembrano costrette a sottomettersi a una sorta di permanente escalation di adattamento e recupero dei macchinismi a-significanti*»⁴²⁶: così le aziende corporative catturano le occorrenze a-significanti interne e le restituiscono in forma significante di messaggi. Si ricorderà allora che di semiologie significanti si possono distinguere due grandi forme: le *semiologie simboliche* e le *semiologie di significazione*. Mentre le prime consentono a diversi regimi semiotici di proliferare indipendentemente, le seconde costituiscono un’inesausta attività di omologazione tra sistemi semantici differenti, riducendo una molteplicità di segni a una sua rigorosa, “unidimensionale”⁴²⁷ codificazione, in genere

⁴²⁵ I *boilerplates* sono delle disposizioni preimpostate per certe tipologie di risposte. Funzionano integrando a uno schema di risposta possibili variazioni su alcuni temi.

⁴²⁶ Weatherby, Leif, *Language Machines*, p. 207.

⁴²⁷ Genosko, Gary, *Felix Guattari: An Abberant Introduction*, , Continuum: London-New York, 2002, p.

linguistica, in questo caso vettoriale-linguistica. La sovracodifica avviene proprio nella traduzione secondo questo ulteriore regime. Tutto ciò che viene immesso come simbolico esce come significante. La molteplicità simbolica viene assorbita in un unico codice, secondo una traiettoria precisa di significazione.

La produttività delle sintesi connettive è divisa fra e distribuita su due piani nelle semiologie significanti (che hanno distorto i piani hjelmsleviani nell'immagine del segno linguistico) sotto la funzione referenziale del segno: un reale denotato e un'immagine rappresentazionale.⁴²⁸

Anche in questo caso il discorso sulle semiotiche miste sembra calzare con i principi computazionali di un LLM, poiché una doppia articolazione viene a realizzarsi sotto l'immagine del segno linguistico, producendo un effetto di referenza simulata (la *response* sembra parlare di qualcosa) tra i due piani: quello dell'espressione articola il testo prodotto, mentre quello del contenuto lo produce tramite la *heat map*.

A questo proposito, la tendenza all'omologazione delle pratiche umane può essere vista come effetto dell'impossibilità che la macchina si ancori al mondo, ovvero che intrattenga qualche tipo di legame causale con esso. Come si è visto sopra⁴²⁹, una delle critiche più frequenti ai LLMs è quella della loro mancanza di *grounding* referenziale. I modelli infatti non hanno corpo, non sono radicati nel mondo a cui rimandano le loro *response*. Weatherby intende la referenzialità come problematica in virtù della natura stessa della conoscenza: «La conoscenza è per sua natura fragile. La fragilità non è un difetto delle macchine, ma una caratteristica della conoscenza»⁴³⁰. Il punto è corretto, ma non basta, poiché il vero problema è la sostituzione dell'ancoraggio pragmatico delle pratiche umane con un ancoraggio statistico tramite i pesi del modello. Allora proprio perché il modello è ancorato ai propri pesi interni le sue risposte sono consistenti e possono essere applicate per diversi temi e situazioni. La varietà semiotica viene a ridursi alla sola comparazione vettoriale dei pesi, guadagnando consistenza interna.

In conclusione, le analisi critiche attraverso il modello delle semiotiche miste

170.

⁴²⁸ Tecnicamente, per come si è visto, lo spazio vettoriale conta migliaia di dimensioni. Il termine viene usato metaforicamente per richiamare una dimensione che sia onnicomprensiva, totalizzante e senza alcuna visa di fuga. L'assonanza terminologica vuole evocare il celebre libro di Herbert Marcuse, cfr. Marcuse, Herbert, *L'uomo a una dimensione*, Einaudi: Torino, 1968.

⁴²⁹ Genosko, Gary, op. cit., p. 168.

⁴³⁰ Cfr. parte 1, cap. 2.1.

di Guattari e la linguistica di Hjelmslev sono riuscite a dimostrare la natura politica e sociale della semantica vettoriale, ovvero il valore di significazione in relazione al suo posizionamento nello spazio. In altre parole, la vettorializzazione non può essere intesa come pura forma, ma già come una sostanzializzazione. Si è visto che ciò comporta già delle conseguenze, sia sul piano pratico che epistemologico. ogni discorso che tratti lo spazio vettoriale come un presupposto naturale e neutro, ritenendo necessario addestrare ogni modello linguistico sullo stesso campione di dati, dimentica che i *corpus* sono già una sostanza formata, da concepire a partire da determinate scelte politiche, concretizzate in uno specifico *fine-tuning*. Sul piano pratico ogni intervento che si voglia dire serio e pertinente sulla natura degli LLMs deve tenere in considerazione cosa entra a far parte del *corpus* di addestramento, quindi sulla calibrazione del *fine-tuning*. Altrimenti, qualsiasi discorso, come quello di Weatherby che ritiene che il modello non sia politico, che non abbia intenzioni, ma che rifletta solo sui dati, finisce, suo malgrado, ad essere consono all'ideologia della corporate. Stando ai termini del discorso di Peirce, non bisogna lasciarsi ingannare dalla proliferazione unilaterale di una delle tre tipologie di rimandi: indice, icona e simbolo.

La vera lezione di Peirce non è tanto la predilezione di una delle tre forme sulle altre, ma la loro distinzione e compostibilità. Anzi, sembrerebbe esserci un'indicazione genetica circa la loro natura tripartita. Si può leggere l'icona come la forma più primordiale perché immediatamente constatabile, grazie alla sua somiglianza con il referente; l'indice cattura invece delle relazioni causali con il referente che solo in un secondo momento possono venire recise e arbitrariamente poste dal simbolo. I tre modi non si escludono, poiché ogni segno combina le tre dimensioni. Ad esempio, una mappa è già un'icona (la geografia del territorio è rappresentata figurativamente), un indice (poiché viene prodotta a partire dai rilievi sul terreno) e un simbolo (in virtù delle convenzioni cartografiche). La vettorializzazione, invece, produce l'apparenza di un rinvio al referente poiché amputa indice e icona. La forma di mistificazione che si produce è quindi totalmente simbolica. Infatti, gli ultimi modelli che si dice lavorino su immagini, video, suoni, non risultano maggiormente ancorati rispetto ai precedenti poiché i pixel di un video sono trattati come simboli da inserire nello spazio vettoriale. La macchina rimane fondamentalmente digitale, retta su operazioni tra vettori-simboli che escludono il rimando analogico a indice e icona. La linguistica di Hjelmslev si è rilevata un aiuto indispensabile perché non riduce la lingua a un sistema di significanti omogeneo, ma ne descrive la pluralità di sostanze formate; le stesse che secondo

Deleuze e Guattari ne fanno «*la sola adatta alla natura d'insieme dei flussi capitalistici e schizofrenici*»⁴³¹.

⁴³¹ Weatherby, Leif, *Language Machines*, Minnesota University Press: Minneapolis, 2025, p. 130.

CONCLUSIONE

Il rapporto tra *token* e vettore, centrale nella riflessione qui proposta, vuole inserirsi nel paradigma dell'analogia Linguaggio-Macchina che nell'attuale dibattito sugli LLMs appare ancora sotto-esplorato⁴³². Se ci si riferisce al capolavoro *Le parole e le cose*⁴³³, Foucault infatti indica tre grandi paradigmi che costituirebbero il pensiero della modernità: Lavoro, Linguaggio e Organismo. Ebbene, la Macchina appare come un tassello mancante al discorso, intimamente connesso, oggi più che mai, al Lavoro. Nel caso dell'AI i due paradigmi di Macchina e Linguaggio sono raramente tenuti assieme, se non nell'ottica neostrutturalista, ma il loro connubio diventa sempre più rilevante a fronte di domande su come la struttura generativa del linguaggio possa venire automatizzata.

L'automatizzazione del processo linguistico è stata declinata in termini statistici: prendendo a riferimento il calcolo bayesiano e la sua ricorsività, la visione ergodica del linguaggio di Shannon con le catene di Markov si sono delucidati alcuni strumenti concettuali per stimarne le conseguenze. Il risultato è quello di un "distribuzionalismo" di stampo humeano: un'espressione ripresa da Weatherby per connotare uno sviluppo linguistico continuo, orizzontale sull'esclusiva base della successione empirica delle occorrenze. Il linguaggio non si contraddistingue per un nocciolo essenziale, una gerarchia di categorie; in questo la lezione strutturalista di Saussure viene riattualizzata da Weatherby giocando sulla relativa apposizione delle parole sulle cose: la parola non può prescindere dalle sue attuali relazioni, quindi dal concatenamento d'enunciazione in cui viene inserito. Ogni concatenamento d'enunciazione a cui ci si riferisce è già denotato in senso materiale, tramite la (non)⁴³⁴ relazione della macchina astratta. Anche Kant si interrogava per la dottrina dello schematismo sulla natura di un rapporto che unisse delle realtà totalmente altre (intuizione e concetto)⁴³⁵. Trascendentali sono allora quelle leggi

⁴³² Deleuze, Gilles, Guattari, Félix, *Anti-Edipo*, Einaudi: Torino, 1972, p. 243.

⁴³³ Pasquinelli, Matteo, *Machine, Organism, Language: a comparative epistemology of AI models*, in «*AI and Society*», 1–13, 2026.

⁴³⁴ Foucault, Michel, *Le parole e le cose*, BUR: Milano, 1978.

⁴³⁵ Deleuze nelle sue lezioni sul sapere in Foucault si interroga sulla natura diagrammatica del metodo archeologico foucaultiano. Per rappresentarsi il rapporto tra parole e cose, tra vedere e sentire, lo configura come un *non-rapporto*, ovvero una relazione tra elementi totalmente eterogenei, in ultima istanza non mediabili. In altre opere parla di *insistenza* delle parole sulle cose: «Ma, se c'è un primato, è perché le due forme eterogenee del sapere si costituiscono per integrazione, e, in condizioni proprie soltanto alle forze, fanno il loro ingresso in un rapporto indiretto, al di sopra del loro interstizio o del loro «non-rapporto». Così il rapporto indiretto tra le due forme del sapere non implica alcuna forma comune, e nemmeno una corrispondenza, ma soltanto quell'elemento in formale delle forze che le tocca entrambe. Il *diagrammatismo* di Foucault, cioè la presentazione di puri rapporti di forze o di emissione di pure

che regolano l'empiria intuibile. L'ordine, però, sembra più un riflesso dovuto a un effetto retroattivo piuttosto che un'arcana natura insita nella realtà. Chomsky con il suo concetto di ricorsività ha saputo certo nominare il carattere di esaustione del linguaggio, la sua possibile reiterazione e indefinito dispiegamento che ne fanno uno strumento aperto. Nessuna grammaticalità, nessuno "schematismo chomskyano" non può non venire iscritto entro la stessa ricorsione che accusa nel fluire del linguaggio.

Alla legiferazione kantiana controbatte l'indefinito procedere della ricerca empirica. La sua interrogazione non si chiede che consistenza abbia il linguaggio, ma ne vaglia le attuazioni. Da qui la lezione strutturalista sull'ordine simbolico: il segno è svincolato dalla cosa come il significato lo è dal significante. La barra tra i due indica un abisso incolmabile, quindi l'impossibilità di fondare la relazione. La relazione resta aperta: la presenza può essere giustificata non da un'ulteriore presenza, ma da un'assenza. Come mette in luce Eco⁴³⁶, l'assenza viene configurata nei termini ontologici di una mancanza, rilevata da Lacan quale cifra della finitudine umana; il discorso ontologico però si traduce sul piano semiotico in un'apertura indefinita di rimandi contingenti; è questa la questione che interessa davvero nel presente testo. L'intero discorso, infatti, s'innesta proprio a partire dal presupposto della *barra* del segno; non potrebbe essere altrimenti. A Weatherby bisogna riconoscere l'intuizione di aver trovato una maniera per giustificare l'iterazione, la ricorsività proprio non di una facoltà umana, ma di una nozione più generale di "macchina". L'epoca attuale, quella della *riproducibilità tecnica del linguaggio*, erode un'ulteriore frontiera che si è sempre creduta propria dell'uomo, ovvero quella del linguaggio stesso. Reazionario ed elusivo, rispetto alle questioni cogenti che gli LLMs muovono, appare allora il ragionamento *umanistico* di voler scovare un'ulteriore frontiera che caratterizzi l'animo propriamente umano. La nozione di *segno* in questo senso presenta due aspetti cruciali per proseguire una linea di pensiero matura, alla luce delle recenti acquisizioni tecniche dei modelli generativi:

- La sua natura differenziale esibisce un tratto adattivo, malleabile, ovvero una flessibilità che ne rende possibile il rimescolamento con altri regni semantici. Il caso degli LLMs è icastico in virtù della messa in rapporto tra *token* e vettori, altro rispetto alla tradizionale "immagine acustica" della parola significante e la definizione significativa dell'idea.

singularità, è quindi analogo allo schematismo kantiano: assicura la relazione da cui il sapere dipende tra le due forme della spontaneità e della ricettività». Cfr. Deleuze, Gilles, *Foucault*, Orthotes: Napoli, 2018, pp. 98-99.

⁴³⁶ Kant, Immanuel, *Critica della ragion pura*, Laterza: Roma-Bari, 1975.

- Il segno riflette un comportamento “oltre-umano” [*inhumanism*]⁴³⁷. L’aspetto anti-umanistico, sostenuto in questa sede dal ragionamento dello strutturalismo serve da presupposto al discorso, in maniera da dimettere ogni linguistica personalistica o soggettivistica, il cui fulcro viene ad appoggiarsi sulla soggettività dell’enunciante. Seppur distante dallo strutturalismo, il paradigma triadico di Peirce permette di descrivere l’andamento *oltre-umano* del discorso. Si pensi infatti che ogni segno e interpretante sono interscambiabile nei loro ruoli, ovvero che ogni pensiero umano può nascere o venire condensato in una sua “controparte” informatica.

L’eco della lezione strutturalista sull’arbitrarietà del segno si riverbera anche sul piano storico dell’evoluzione delle tecnologie: gli LLMs inaugurano la stagione del *vector grounding problem*, ovvero della questione sull’ancoraggio della codificazione digitale al mondo. Come si è visto a più riprese però, il rapporto LLM-mondo sembra essere reciso, almeno in maniera diretta. La macchina svolge un “autoriferimento”⁴³⁸ del linguaggio a sé stesso; la posizione si è infatti indicata attraverso la funzione poetica di Jakobson. Insomma, il segno è svincolato dal referente, facendone a meno e l’operazione linguisticamente la si è indicata come funzione poetica. Il dibattito però rimane aperto e non è ancora superato nella misura in cui, nel corso della trattazione si è definita la relazione dei modelli con l’uomo come *parassitaria*. Un LLM non ha bisogno *direttamente* di un referente per poter comporre un messaggio, appoggiandosi sulla funzione poetico-ricorsiva del linguaggio; però, i prodotti semantici di un LLM non riecheggiano forse l’andamento del discorso di alcuni esseri (umani) coinvolti con l’ambiente attraverso l’intenzionalità? Quindi sì, i modelli non hanno bisogno di un referente direttamente, ma rimangono, in ultima istanza ancora indirettamente fondati?

Questi ragionamenti, seppur utili al dibattito, esibiscono l’ardita pretesa di inquadrare la novità di un fenomeno in continua evoluzione con statiche categorie secolari, come quella dell’intenzionalità. L’attenta analisi delle modifiche empiriche, legate ai modelli in uso non basta: bisogna elaborare anche dei concetti che riflettano questa novità. Un esempio è fornito ancora da Weatherby nel riapplicare in veste innovativa il concetto di ideologia, rendendolo consono a una scansione empirica degli *output*. Far coincidere l’ideologia con l’esibizione di alcune tracce, più comuni e *standard*

⁴³⁷ Eco, Umberto, *La struttura assente*, Bompiani: Milano, 1968, p. 30.

⁴³⁸ Reza Negarestani parla di *inhumanism* come di un ordine significativo che non prende più a riferimento come sorgente l’intenzionalità di una persona, quindi il suo “pensiero”. Il testo sintetico produce senso al di fuori di ogni intenzionalità diretta verso un referente esterno. Cfr. <https://www.e-flux.com/journal/52/59920/the-labor-of-the-inhuman-part-i-human>.

rispetto ad altre, sembrerebbe comprovare il fatto che l'ideologia a cui si è sempre pensato, con la Scuola di Francoforte o Lacan, come istanza trascendentale regolatrice, rimanda uno strumento inefficace nella sua fuggibilità, rimanga, cioè, un residuo ineffabile. Al contrario, Weatherby non arriva alla conclusione che togliendo gli LLMs, si squarcia ogni velo ideologico, ma che l'ideologia funziona intimamente con l'utilizzo che si fa di oggetti di consumo culturali. Una relazione, quella tra trascendentale ed empirico, non priva di problemi, almeno per come la configura Foucault ne *Le parole e le cose*, non a caso libro di riferimento nella terza parte della tesi per chiare cosa sia l'ideologia. La dicotomia kantiana apre a un'epoca di riflessione sdoppiata tra questi due piani, facendo dell'uomo un allotropo empirico-trascendentale. Branche come l'antropologia e la psicanalisi possono apportare nuove riflessioni, sulla base di questa dicotomia, ma la "ferita aperta", lo scarto tra i due, rimane insanabile.

Forse, bisognerebbe dimettere la nozione di trascendentale dall'esclusivo ambito umano, connettendola direttamente con il precipuo "sostrato" empirico che la sostiene aprendo a una vicendevole interazione reale-simbolico, superando il rango subalterno in cui il reale è sprofondato con Lacan e la supremazia accordata al simbolico. Il simbolico, come registra Deleuze⁴³⁹ non è solo un registro a parte, ma addirittura ciò che governa quello dell'immaginario e del reale. Eppure, nella quarta parte si sono dimostrate diverse occorrenze con cui l'oscura e reale configurazione della *heat map*⁴⁴⁰ predispone comportamenti che si possono definire a tutti gli effetti "simbolici": comprare questo anziché quello, votare questo partito anziché l'altro, vestire in un determinato modo, leggere fenomeni storici secondo una certa prospettiva, creduta "oggettiva". La forma "abita" la materia, il contenuto che modello, per questo il modello metafisico ed epistemologico dell'empirismo trascendentale supera *l'impasse* della modernità secondo Foucault. Da qui si aprono due percorsi per ulteriori ricerche.

Da un lato si tratta di recuperare il concetto di entropia per Shannon, mostrandone non solo le radici termodinamiche (Boltzmann), ma studiandone le implicazioni nel caso di un vasto serbatoio di informazioni, ovvero saggiandone la torsione quando il canale non veicola un messaggio puntuale, bensì attinge a un serbatoio probabilistico sterminato. Infatti, l'informazione coincide con la sua autoinformazione, ovvero con il logaritmo della probabilità che quell'informazione si verifichi; in altre parole, quanto più un

⁴³⁹ Propriamente non si autoriferisce, in quanto non possiede alcun identità stabile e fissa, non potendo codificare alcun "sé".

⁴⁴⁰ Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo?*, SE: Milano, 2008, p. 15.

messaggio è inatteso, tanto più alto è il suo contenuto informativo. Al contrario, è la *ridondanza* a rimanere da sfondo, come ciò che si è già detto. La generatività stocastica si risolve invece nel percorso più battuto all'interno della distribuzione della *heat map*, privilegiando la ridondanza. La circolazione accelerata dell'informazione, condensata in miliardi di *prompt* scritti quotidianamente, ispessisce i solchi probabilistici, sedimentando le regolarità distribuzionali lungo l'asse sintagmatico. La novità, intesa come imprevedibilità, lascia il posto alla "rilevanza ridondante": l'apparente ossimoro sta ad indicare che ad essere pertinente è la modalità di discorso e gli argomenti che maggiormente vengono intrapresi, il già-detto.

Dall'altro lato si tratta di tornare all'empirismo trascendentale, un modello di pensiero concepito da Deleuze negli anni '60 e doveroso di un aggiornamento. Così l'indagine viene descritta dall'autore:

L'empirismo diventa veramente trascendentale, e l'estetica una disciplina apodittica, quando cogliamo direttamente nel sensibile ciò che può essere soltanto sentito, l'essere stesso del sensibile: la differenza, la differenza di potenziale, la differenza d'intensità come ragione del diverso qualitativo.⁴⁴¹

La mossa consiste in una vera e propria revoca dei fondamenti della modernità, per come Foucault ne parla; il trascendentale è empirico nella misura in cui non può venire concepito a priori, ma è necessario imbastire una ricerca contingente sull'ambiente in cui si vive, per sondare come forme ("soggetto" o "oggetto" ad esempio per la modernità) vengano a costituirsi. Un esempio di ciò si ha nelle pagine di *Differenza e ripetizione* in cui si indaga la formazione delle specie animali oppure nella genesi dinamica di *Logica del senso*, atta a rintracciare i fondamenti pre-significanti e pre-edipici del linguaggio. È un'idea che, mutata, ritorna anche per il progetto delle semiotiche miste che qui si è voluto, in conclusione, richiamare. L'aspetto cruciale di questa indagine è la necessità a richiamarsi agli oggetti che costellano l'esperienza, al precipuo ambiente che abitano:

⁴⁴¹ A questo punto qualcuno potrebbe obiettare che la *heat map* si riduce a una formula algoritmica e che ciò che veramente fa effetto sono i contenuti immagazzinati nello spazio latente, ovvero le relazioni stocastiche tra *tokens*. Così la questione si riduce a un aspetto di pura forma, come per dire che la soluzione risiederebbe nella mera modifica, ritaglio dei testi degli LLMs, ancora una volta distinguendo tra forma e contenuto, aprendo la strada a interpretazioni, come quella di Weatherby, che ritengono la questione LLM come puramente *formale*. È ancora possibile questa distinzione? Qui si è dimostrato di no. Un esempio, tratto dalla recente storia degli LLMs, può essere visto nell'iniziale critica degli *stochastic parrots* con cui molti studiosi hanno liquidato la questione. Nel giro di qualche anno, se non di qualche mese, la (forma della) loro produzione di messaggi si è complessificata a tal punto da renderli indistinguibili e ragionare in termini di forma o contenuto è inadatto. Forma [*morphé*] tiene assieme entrambi gli aspetti.

non tanto le condizioni di possibilità kantiane, ma le condizioni *reali*, specifiche dell'esperienza. Le condizioni, costituendo la genesi dell'esperienza stessa, sono perciò stesso mutevoli e plastiche. Da qui il progetto mai definitivo, sempre soggetto a ulteriori ricerche, a cui l'empirismo trascendentale costringe. Lo studio degli LLMs, il funzionamento della loro architettura *transformer*, luogo genetico dell'emersione del discorso non può non rientrare in questo progetto. La sensibilità presa ad oggetto dall'empirismo trascendentale non bisogna confonderla con la sensazione, rischiando di sconfinare in un ingenuo vitalismo. Ciò che si percepisce è già innervato da flussi e direzioni operative, è già sovracodificato perché rimescolato dai vettori, per come si è inteso nel paragrafo 4.3.2. L'intenzione soverchiante ogni impianto strutturale di Deleuze è chiara: configurare un campo preindividuale e impersonale in cui tutte le forme si disfano e ne viene mostrata la genesi. Bisogna allora concentrare gli studi in ottica di un empirismo trascendentale che non premetta il *sentendum* al *cogitum*, le intensità agli affetti, ma che li veda irrimediabilmente mescolati. Quello che interessa è mostrare l'emersione del significato da ciò che ancora ne sfugge, come quest'operazione possa venire automatizzata e in autonomia rispetto alla vigilanza dell'uomo.

D'altronde la stessa contaminazione dei paradigmi di Macchina e Linguaggio è volta a mostrare l'intimo nocciolo che lega senso e nonsenso, parole e *token*. Parlare significa esperire o ripetere un processo di calcolo? Il linguaggio ha una matrice esclusivamente biologica o anche informatica, nella misura in cui può venire riprodotto per mezzo di vettori? Se fosse questo il caso, paragrafi come il 2.2.1 sarebbero bisognosi di approfondimento, cercando di sondare la computabilità del linguaggio per mezzo dei programmi.

Aver illuminato il funzionamento dell'architettura *transformer* ha messo in luce l'apporto fondamentale della specificità del *corpus*, prevalentemente innestato su dati personali, assieme alle scelte corporativistiche di fine-tuning, due aspetti "politici" inscindibili da una riflessione attorno all'automatizzazione del linguaggio. A innervare politicamente il testo non sono state solo considerazioni di carattere ingegneristico sull'architettura dei modelli, ma anche i concetti di assemblaggio e diagramma. Deleuze e Guattari, oltre alla questione delle semiotiche miste, forniscono un appoggio teorico attorno alla questione della materialità del linguaggio. La giustificazione ultima dello strutturalismo è l'assenza, l'arbitrarietà del segno, assioma che non può bastare da presupposto a un dibattito attorno alla codificazione massiva di dati che interessano le vite degli utenti, le loro prospettive e la loro enciclopedia. Il ruolo di Deleuze all'interno

di questo lavoro non vuole riflettere una romantica attestazione del primato dell'analogico sul digitale, anzi vuole favorire una più corretta conoscenza di quest'ultimo. Come mostra anche Brian Massumi, se si può parlare di un "primato" dell'analogico sul digitale è solo in riferimento alla sua ineludibilità, ovvero al fatto che non può esservi digitale senza un analogico sostrato a cui appoggiarsi. *«La superiorità dell'analogico del digitale a cui si allude nel titolo non contraddice quest'ultima considerazione di vederli associati assieme. Si riferisce al fatto che il loro processo di cooperazione – integrazioni trasformative, trasduzioni – sono loro stesse operazioni analogiche»*⁴⁴². Ogni ripiegamento, ogni trasformazione della struttura non è dato che dall'attuale configurazione di questa, quindi da un centro immanente di forze che la abita. Quindi, Deleuze richiama alla relazione duale tra analogico e digitale, già-sempre in atto qualora ci si riferisca anche a un modello il cui meccanismo rimane espresso entro uno spazio latente puramente digitale; il concatenamento è parzialmente analogico in virtù dell'eterogeneità degli aspetti che richiama. Il rapporto tra analogico e digitale, tra senso e nonsenso viene esplicitato dallo studio genetico dell'empirismo trascendentale. La domanda sulla genesi di una struttura significante si fa ancora più cogente laddove si parla di automatizzazione del linguaggio. Concetto cardine per esplorare questo rapporto rimane quindi il *diagramma*, nel suo ripartire forma dell'espressione e del contenuto in maniera indissolubile. L'aspetto più interessante di un simile concetto è come tenga assieme la dimensione astratta del concetto, senza sfociare nell'idealismo e quella materiale senza sfociare nel realismo più ingenuo. Espressione e contenuto sono legate vicendevolmente, si presuppongono a vicenda e

Certo, un LLM non può essere capace di creare, nel senso dell'evento, come si è dimostrato in 3.1.1, ma la questione della generatività stocastica lascia ancora inesplorata la capacità di questi modelli a rendere sempre più "umani" i simulacri, ricreando le altre funzioni del linguaggio indicate da Jakobson. Le analisi portate avanti da Weatherby presuppongono la funzione poetica per dimostrare la ricorsività propria del linguaggio. Che tipo di rapporto intercorre tra funzione poetica e le altre? Doverosi sono degli approfondimenti a proposito, ovvero per cercare di capire se l'ancoraggio LLM-mondo sia una questione puramente risolvibile attraverso l'ipotesi ergodica del linguaggio e la sua reiterazione poetica oppure no. Nell'ultima parte, infatti, si è accennato a questo aspetto proprio per introdurre le semiotiche miste: l'operazione di Weatherby dispone unilateralmente la funzione poetica come fondativa della ricorsività linguistica,

⁴⁴² Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997, p. 180.

tralasciando le altre funzioni.

Parlare di Linguaggio non prescindendo dalla Macchina significa allora esibirne il concreto ancoraggio al concatenamento di effetti che dispone la Macchina; nel caso degli LLMs le pratiche giuridiche per le *policy* sulla *privacy*, il loro utilizzo d'intrattenimento anche nei social network, il sapere delle biblioteche che viene integrato, oppure contratti di lavoro precari di Mechanical Turk e di Sama e protocolli di moderazione delle imprese, favorendo un incontro pluridisciplinare per dipanare la complessità strutturale del dibattito. Soprattutto, parlare di Linguaggio, per come gran parte della quarta parte ha insistito, significa attestarsi su un piano virtuale e preindividuale, quindi a-significante. La concezione di un linguaggio unicamente simbolico e soggettivamente posseduto dall'enunciatore non può bastare; l'insufficienza si registra a livello di "asservimento macchinico", termine con cui Lazzarato⁴⁴³, riprendendo Deleuze e Guattari, si rivolge allo smantellamento della soggettività, o meglio al suo stesso processo di produzione secondo concatenamenti macchinici e flussi di desiderio. Come si è visto in 4.2.1, solo passando a livello *molecolare* si riesce a cogliere il flusso di intensità che scinde il soggetto in tanti *dividui*, ovvero molteplici declinazioni dell'Io, in tante versioni che possono venire ordinate proprio grazie alle strategie di calcolo dei vettori; la vettorializzazione incarna un ulteriore esempio di come la soggettività individuale non possa presentarsi in veste di preambolo del discorso, ma solo come suo effetto ultimo.

⁴⁴³ Massumi, Brian, *Parables of the virtual*, Duke University Press: London, 2002, p. 143.

BIBLIOGRAFIA

- Adorno, Theodor, *Contributo alla teoria dell'ideologia*, in *Scritti sociologici I*, Einaudi: Torino, 1976.
- Adorno, Theodor, Horkheimer, Max, *Dialettica dell'illuminismo*, Einaudi: Torino, 1966.
- Alammar, Jay, Grootendorst, Maarten, *Hands-On Large Language Models*, O' Reilly, 2024.
- Althusser, Louis, *Lo Stato e i suoi apparati*, Editori Riuniti: Roma 1977.
- Althusser, Louis, *Per Marx*, Mimesis: Milano, 2006.
- Althusser, Louis, *et al.*, *Leggere il Capitale*, Mimesis: Milano, 2006.
- Anders, Günther, *L'uomo è antiquato I. Considerazione sull'anima nell'epoca della seconda rivoluzione industriale*, Bollati Boringhieri: Torino, 2007.
- Bai, Hui *et al.*, *Artificial Intelligence Can Persuade Humans on Political Issues*, OSF Preprints, 2023.
- Balibar, Etienne, *Tu l'as dit!*, in *Per Althusser*, Manifestolibri, Roma 1991.
- Barthes, Ronald, *La retorica dell'immagine*, in *L'ovvio e l'ottuso*, Einaudi: Torino, 2001.
- Barthes, Roland, *Lezione*, Einaudi: Torino, 2001.
- Baudrillard, Jean, *Lo scambio simbolico e la morte*, Feltrinelli: Milano, 2022.
- Bellucci, Francesco, *Eco and Peirce on Abduction*, in «European Journal of Pragmatism and American Philosophy», X-1, 2018.
- Berrigo, Billy, *OpenAI Used Kenyan Workers on Less Than \$2 Per Hour*, in «Time», Gen. 18, 2023.
- Berto, Francesco, *Tutti pazzi per Gödel!*, Laterza: Roma-Bari, 2009.
- Borges, Luis, J., *Finzioni*, Adelphi: Milano, 2003.
- Brand, James *et al.*, *Using GPT for Market Research*, Harvard Business School Working Paper 23-062, 2023.
- Ceruti, Mauro, Bellusci, Francesco, *Abitare la complessità*, Mimesis: Milano, 2020.
- Chartonsky, Grégory, Citton, Yves, *The Fourth Memory*, in «Transit», 15-1, 2025.
- Chomsky, Noam, *Syntactic Structures*, De Gruyter Mouton: Berlino, 2002.
- Damerow, Peter, "Abstraction and Representation", in *Abstraction and Representation: Essays on the Cultural Evolution of Thinking*, Springer: Berlin, 2013.
- De Ruvo, Giuseppe, *Rimettere il fantasma nella macchina. Una critica deleuziana dell'ontologia digitale e della governabilità algoritmica*, La Deleuziana n. 14, 2022.
- Deleuze, Gilles, Guattari, Félix, *Anti-Edipo. Capitalismo e Schizofrenia*, Einaudi: Torino,

1975.

- Deleuze, Gilles, Guattari, Félix, *Che cos'è la filosofia?*, Einaudi: Torino, 2002.
- Deleuze, Gilles, *Da cosa si riconosce lo strutturalismo?*, SE: Milano, 2008.
- Deleuze, Gilles, *Differenza e ripetizione*, Cortina: Milano, 1997.
- Deleuze, Gilles, *Foucault*, Orthotes: Napoli, 2018.
- Deleuze, Gilles, *Francis Bacon. La Logica della sensazione*, Quodlibet: Macerata, 2023.
- Deleuze, Gilles, *La piega*, Einaudi: Torino, 2004.
- Deleuze, Gilles, *Logica del senso*, Feltrinelli: Milano, 1975.
- Deleuze, Gilles, *Marcel Proust e i segni*, Einaudi: Torino, 1967.
- Deleuze, Gilles, Guattari, Félix, *Mille piani. Capitalismo e schizofrenia*, trad. it. di G. Passerone, Castelvecchi, Roma, 2017.
- Deleuze, Gilles, *Pourparler*, Quodlibet: Macerata, 1999.
- Deleuze, Gilles, *Sulla pittura*, Einaudi: Torino, 2024.
- Descartes, René, *Meditazioni metafisiche*, Laterza: Roma-Bari, 1997.
- Dosse, Françoise, *Gilles Deleuze, Félix Guattari: une biographie croisée*, Editions La Découverte: Paris, 2009.
- Dosse, Françoise, *History of structuralism*, Minnesota University press: Minneapolis, 1997.
- Dreyfus, Hubert, Dreyfus, Stuart, *Making a Mind versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint*, in «Daedalus», 1, 1988.
- Eco, Umberto, *La struttura assente*, Bompiani: Milano, 1968.
- Erlich, Viktor, *Il formalismo russo*, Bompiani: Milano, 1966.
- Floridi, Luciano, *La quarta rivoluzione*, Cortina: Milano, 2017.
- Foucault, Michel, *La nascita della clinica*, Einaudi: Torino, 1998.
- Foucault, Michel, *Le parole e le cose*, BUR: Milano, 1978.
- Foucault, Michel, “Truth and Power” (1977), in *Power/Knowledge: Selected Interviews and Other Writings*, Colin Gordon: New York, 1980.
- Gastaldi, Juan, Luis, *Why Can Computers Understand Natural Language?: The Structuralist Image of Language Behind Word Embeddings*, in «Philosophy and Technology», 34(1).
- Genosko, Gary, *Felix Guattari: An Abberant Introduction*, , Continuum: London-New York, 2002.
- Geoghegan, Bernhard, D., *Code: from the theory of information to french theory*, Duke Press University: Durham, 2023.

- Gray, Mary L., Siddharth, Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, Houghton Mifflin Harcourt: Boston, 2019.
- Greenberg, Clement, *Art and culture*, Beacon press: Boston, 1971.
- Guattari, Félix, *Caosmosi*, Orthotes: Napoli, 2024.
- Guattari, Félix, *Inconscio macchinico*, Orthotes: Napoli, 2022.
- Guattari, Félix, *La rivoluzione molecolare. La nuova lotta di classe*, PGreco: Milano, 2017.
- Guattari, Félix, *Machine and structure*, in *Psychanalysis and transversality*, Semiotext(e): Los Angeles, 2015.
- Harnad, Stevan, *The symbol grounding problem* in «Physica D: Nonlinear Phenomena», 42 (1-3), 1990.
- Harris, Zellig, *Structural linguistics*, University of Chicago Press: Chicago, 1960.
- Hjelmslev, Louis, *Prolegomena to a Theory of Language*, University of Wisconsin Press: Madison, 1969.
- Hjelmslev, Louis, *The basic structure of language*, in *Essais linguistiques II*. Nordisk Sprog- og Kulturforlag, 1973.
- Holland, Eugene W., *Deleuze and Guattari's Anti-Oedipus*, Routledge: Londra, 1999.
- Holland, Eugene, W., *Deleuze and Guattari's A Thousand Plateaus*, Bloomsbury, London-New York, 2013.
- Jakobson, Roman, *Closing statements: linguistics and poetics*, in Sebeok, Thomas A., *Style in Language*. Cambridge, MA: MIT Press, 1966.
- Jakobson, Roman, *Parti e totalità nel linguaggio* in *Qualità e quantità e altre categorie della scienza*, Bollati Boringhieri: Torino, 1971.
- Jakobson, Roman, *Structure of Language and Its Mathematical Aspects* (Vol. 12). American Mathematical Society, 1961.
- Kant, Immanuel, *Critica della facoltà di Giudizio*, Laterza: Roma-Bari, 1907.
- Kant, Immanuel, *Critica della ragion pura*, Laterza: Roma-Bari, 1975.
- Kockelman, Paul, *Last Words*, Prickly Paradigmatic Press: Chicago, 2024.
- Lafontaine, Céline, *The cybernetic matrix of French Theory*, in «Theory, Culture and Society», 24, 2007.
- Lapoujade, David, *Aberrant Movements, the philosophy of Gilles Deleuze*, Semiotext(e): London, 2017.
- Lazzarato, Maurizio, *Segni e macchine*, Ombre corte: Verona, 2019.
- Leclerc, Jean-Jacques, *Deleuze and language*, Palgrave Macmillan: Londra, 2002.

- Liu, Lydia, *The Freudian Robot: digital media and the future of the unconscious*, University of Chicago Press: Chicago, 2010.
- Lorusso, Silvio, *DeepDreaming Willy Wonka: AI Weird as the New Kitsch*, Hidden Layers, 2024.
- Manning, Christopher, *Human language understanding and reasoning*, in «Daedalus», 2022, 151 (2).
- Marcuse, Herbert, *L'uomo a una dimensione*, Einaudi: Torino, 1968.
- Marx, Karl, *Grundrisse*, Einaudi: Torino, 1976.
- Massumi, Brian, *Parables of the virtual*, Duke University Press: London, 2002.
- Mikolov, Tomas et al., *Linguistic Regularities in Continuous Space Word Representations*, in «NAACL HLT», 2013, pp. 746-751.
- Monod, Jacques, *Il caso e la necessità*, Mondadori: Milano, 2022.
- Morton, Timothy, *Iperoggetti*, Nero: Roma, 2018.
- Nietzsche, Friedrich, *La genealogia della morale*, Adelphi: Milano, 1992.
- Odgen, Richard, *The meaning of meaning*, Dead Author Society, 2023.
- Ortigue, Edmond, *Le discours et le symbole*, Aubier: Parigi, 1962.
- Ouyang et al., *Training Language Models to Follow Instructions with Human Feedback*, in «Advances in Neural Information Processing Systems», 35, 2022.
- Pasquinelli, Matteo, *The Eye of the Master. A Social History of Artificial Intelligence*, Verso: London, 2023.
- Perrigo, Billy, *OpenAI Used Kenyan Workes on Less Than 2\$ Per Hour to Make ChatGPT Less Toxic*, in «Time», 18 gennaio 2023.
- Quaranta, Domenico, *Digital Kitsch: Art and Kitsch in the Informational Milieu*, in *The Changing Meaning of Kitsch: From Rejection to Acceptance*, Springer Verlag, 2023.
- Ricoeur, Paul, *Freud and Philosophy: An Essay on Interpretation*, Yale University Press: New Haven, CT, 1970.
- Rieder, Bernhard, *Engines of Order. A Mechanology of Algorithmic Techniques*, Amsterdam University Press, Amsterdam 2020.
- Robcis, Camille, *Disalienation*, University of Chicago Press: Chicago, 2021.
- Ruyer, Raymond, *La genesi delle forme viventi*, Bompiani: Milano, 1966.
- Salton, Gerard et al., *A Vector Space Model for Automatic Indexing*, in «Communication of the ACM», vol. 18.
- Salvi, Francesco et al., *On the Conversational Persuasiveness of Large Language Models. A Randomized Controlled Trial*, in «Nature Human Behaviour», 8, 2024.

- Saussure, Ferdinand de, *Corso di linguistica generale*, trad. it. di T. De Mauro, Laterza: Roma-Bari, 1967.
- Saussure, Ferdinand de, *Scritti inediti di linguistica generale*, Laterza: Roma- Bari, 2005.
- Schütze, Hinrich, *Dimension of meaning*, Proceedings of Supercomputing '92, Center for the Study of Language and Information, Standford, 1992.
- Serres, Michel, *Il parassita*, Mimesis: Milano, 2022.
- Shannon, Claude, Weaver, Nobert, *la teoria matematica delle comunicazioni*, ETAS Kompas: Milano, 1971.
- Simondon, Gilbert, *Del modo di esistenza degli oggetti tecnici*, Orthotes: Napoli, 2021.
- Simondon, Gilbert, *L'individuazione alla luce delle nozioni di forma e di informazione*, Mimesis: Milano, 2020.
- Stewart, Ian, *Le 17 equazioni che hanno cambiato il mondo*, Einaudi: Torino, 2017.
- Stiegler, Bernard, *La società automatica I. L'avvenire del lavoro*, Meltemi: Milano, 2019.
- Szepanski, Achim, *In the delirium of simulacrum*, Becoming press: Berlino, 2025.
- Vaccari, Franco, *Fotografia e inconscio tecnologico*, Einaudi: Torino, 2011.
- Vaswani, Ashish *et al.*, *Attention is all you need*, in «Advances in Neural Information Processing Systems 30 (NIPS 2017).
- Weatherby, Leif, *Language Machines*, Minnesota University Press: Minneapolis, 2025.
- Zuboff, Susan, *Il capitalismo della sorveglianza*, IUSS press: Roma, 2025.

SITOGRAFIA

- Askill, Amanda, *A General Language Assistant as a Laboratory for Alignment*, <https://arxiv.org/abs/2112.00861>.
- Bender, Emily, Gebru, Timnit, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", <https://dl.acm.org/doi/10.1145/3442188.3445922>.
- Leech, Gavin, *Questionable practices in machine learning*, <https://arxiv.org/abs/2407.12220>.
- Negarestani, Reza, <https://www.e-flux.com/journal/52/59920/the-labor-of-the-inhuman-part-i-human>.
- Schaeffer, Rylan *et al.*, *Are emergent Abilities of Large Language Models a Mirage*, <https://arxiv.org/abs/2304.15004>.
- Ullman, Tomer, *Large Language Models fail on Trivial Alterations to Theory of Mind Tasks*, <https://arxiv.org/abs/2302.08399>.

Dichiarazione del ricorso a strumenti di Intelligenza Artificiale (IA)

Il sottoscritto Ettore Visentin, matricola numero 908815, iscritto presso Ca' Foscari al corso di Laurea magistrale in Scienze Filosofiche, autore della tesi di laurea Strutture in trasformazione. Un'analisi semiotica dei modelli generativi, relatore Matteo Pasquinelli, consapevole delle sanzioni previste dall'art. 76 del Testo unico, D.P.R. 28/12/2000 n.445, e della decadenza dei benefici prevista dall'art. 75 del medesimo Testo unico in caso di dichiarazioni false o mendaci, sotto la propria personale responsabilità ai sensi degli artt. 46 e 47 del D.P.R. 28/12/2000 n. 445,

DICHIARA

di NON aver fatto ricorso a tecnologie di intelligenza artificiale generativa per supportare vari aspetti della ricerca e della scrittura della sua tesi di laurea.

